An out-of-sample analysis of investment guarantees for equity-linked products: Lessons from the financial crisis of the late-2000s

Maciej Augustyniak and Mathieu Boudreault

Abstract

In this paper, we analyze the risk underlying investment guarantees using 78 different econometric models, namely GARCH, regime-switching, mixtures, and combinations of these approaches. This extensive set of models is compared with returns observed during the financial crisis in an out-of-sample analysis, bringing a new perspective to the study of equity-linked insurance. We find that despite the very good fit of recent models, too few of them are capable of consistently generating low returns over long periods of time, which were in fact observed empirically during the financial crisis. Moreover, tail risk measures vary significantly across models and this emphasizes the importance of model risk. Since most insurance companies are now focusing on dynamically hedging their investment guarantees, we also investigate the robustness of the Black-Scholes delta hedging strategy. We find that hedging errors can be very large among the top fitting models implying that model risk must be taken into consideration when hedging investment guarantees.

Keywords: *investment guarantees, out-of-sample, actuarial approach, delta hedging, model risk*

1 Introduction

Investment guarantees are very popular features in life insurance policies because in addition to paying a death benefit, these policies are tied to the return of an underlying asset or an actively managed portfolio. Thus, the policy also acts as an investment because the investor's capital is credited a minimum return. In exchange for this protection, the policyholder pays a higher premium, reflecting the market risk assumed by the insurance company. Because the investment guarantee is essentially a non-standard long-term put option, it is very difficult for insurance companies to completely match this liability with a similar put option on the market. Thus, insurers need econometric models to forecast the potential loss on this guarantee.

The Canadian Institute of Actuaries (CIA) (through the Task Force on Segregated Fund Investment Guarantees), the American Academy of Actuaries (AAA) (through the RBC C3 Phase II report) and Hardy (2001, 2003) have strongly recommended the use of stochastic (econometric) models for reserving the loss on investment guarantees. The industry, backed by professional organizations, has been mainly managing these products using the traditional actuarial approach. This consists in projecting the loss on investment guarantees using multiple scenarios of the underlying asset, and reserving a sufficient amount based upon tail risk measures (say the Value-at-Risk (VaR) or the Conditional Tail Expectation (CTE)). However, the recent financial crisis has forced many banks and insurers to revise their risk management policies and many life insurance companies are now dynamically hedging their investment guarantees.

In this paper, we conduct an extensive analysis of basic and advanced univariate (single asset) econometric models for the purpose of managing investment guarantees (for a study of multivariate (multiple assets) models, refer to Boudreault and Panneton, 2009). There are mainly three reasons why we decided to make this investigation. First, the univariate econometric literature has evolved significantly since Hardy (2001, 2002, 2003), Wong and

Chan (2005) and Hardy et al. (2006). Second, the year 2008 has seen very large investment banks (almost) default, leading to the worst recession since the Great Depression. Many insurers suffered important losses on their segregated funds or variable annuities because of this financial crisis. As a consequence, insurance companies are now focusing more and more on dynamic hedging. However, and this is our third motivation for this analysis, the empirical literature regarding hedging investment guarantees is scarce.

The purposes of our paper are three-fold. First, we would like to complete the empirical analysis of Hardy et al. (2006) by analyzing the fit of 78 models (likelihood-based criteria, quantile-quantile plots, normality and heteroskedasticity tests). These models differ in their volatility dynamics (GARCH-type, regime-switching (RS) or combinations of these) and the use of different error distributions. Rather than recommending a specific model, these models form the basis of a thorough robustness analysis. Second, we analyze the capability of these models to consistently generate low returns over long periods of time. This enables us to assess if reserving approaches, based upon traditional actuarial techniques, would have been appropriate to cover losses of various investment guarantees that matured during the financial crisis. Third, we apply the Black-Scholes delta hedging strategy with scenarios generated from these 78 models to analyze the distribution of hedging errors. Results are compared with similar contracts that could have been issued in the past to check the validity of the outcomes.

Even if it is possible to find a model with an excellent fit, this model will always be a simplification of the true market dynamics, which are unknown. The fact that 78 models are analyzed helps understand an important risk that is often overlooked: model risk, i.e., the uncertainty regarding the appropriateness of a chosen model with respect to the true dynamics of the underlying process. Within a subset of models that fit the market data very well, we find significant variations across models, i.e., important model risk. This is true for tail risk measures of the loss on investment guarantees as well as for hedging errors.

This paper is structured as follows. Section 2 briefly discusses the various models that were fitted and the results of numerous statistical tests that were conducted on these models. Section 3 focuses on the capability of the models to generate low returns over long periods of time. In Section 4, we analyze the robustness of the Black-Scholes delta hedging strategy and Section 5 concludes. Appendix B shows our results for all of the 78 models since the body of the paper focuses on a subset of these models.

2 Models and their adequacy

This section summarizes the models used in our analysis along with the results of numerous statistical tests that assess the quality of the fit of these models. The data that we consider is the set of monthly log-returns on the S&P500 total return index between February 1956 and December 2010. This dataset comprises the financial crisis of the late-2000s. Note that financial markets showed an initial sign of weakness in July 2007 when investors lost confidence in mortgage-backed securities but the debacle started in September 2008 with the bankruptcy of Lehman Brothers. Financial markets generally reached their lows in the beginning of March 2009.

2.1 Overview of models

In this section, we briefly present and describe the different classes of models that are considered in our paper. These models are widely used in the econometric literature to model data and aim to replicate the broadly accepted stylized facts of financial data. These include fat tails and negative skewness of the returns' distribution, jumps in the volatility, volatility clustering and the leverage effect which suggests that past returns and future volatilities are negatively correlated (see Campbell et al., 1997; Tsay, 2010).

2.1.1 GARCH models

The GARCH family of models allows for the variance in the returns to be time-varying. Given past information, the current variance depends on previous innovations and variances, and is thus conditionally deterministic. There exists many different variations of GARCH models: (1) the standard GARCH model of Bollerslev (1986); (2) the Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) of Glosten et al. (1993); (3) the asymmetric power ARCH (APARCH) of Ding et al. (1993); (4) the exponential GARCH (EGARCH) of Nelson (1991). The GJR-GARCH model is an extension of the standard GARCH model as it allows the variance equation to respond asymmetrically to shocks to take into account the leverage effect. The APARCH model further generalizes that model by allowing the volatility exponent in the variance equation to be different than two. The EGARCH model is similar to the GJR-GARCH model in the sense that it allows asymmetric shocks but it parametrizes the logarithm of the variance instead of the variance. For a survey of GARCH models, refer to Bollerslev (2008) or Tsay (2010).

2.1.2 RS models

Models in the RS family involve a process that switches between two or more distributions according to a Markov chain. This Markov chain is usually unobserved and is often given an economic interpretation such as representing the states of an economy (e.g., recession and expansion). For instance, a process that switches between two normal distributions with different means and variances is a RS model if the switching is governed by a Markov chain. In that example, given past observed information, the conditional variance is stochastic since the current state of the economy is unknown. The term RS is not the only term used to describe such a model in the academic literature. Alternative terms include hidden Markov model (HMM), hidden Markov process, Markov-dependent mixture and Markov-switching model. Hamilton (1989) is generally cited as the one to have introduced and popularized RS models in the economic and econometric literature although some of his ideas were already present in Goldfeld and Quandt (1973). RS models became popular in actuarial science since Hardy (2001) used them in the context of investment guarantees. The RS model of Hardy (2001) has been shown to provide a very good fit to the left tail of the returns' distribution in addition to feature a closed-form solution for the computation of the CTE reserve.

2.1.3 RS-GARCH models

It is not difficult to combine a GARCH model with a RS framework to build a RS-GARCH model. One way to justify such a combination is given by Lamoureux and Lastrapes (1990) who show that the high persistence observed in the variance of financial returns can be due to time-varying GARCH parameters. Estimating parameters of a RS-GARCH model is a challenging task because at each time point, the conditional variance depends on the entire history of regimes. This path dependence problem renders exact computation of the likelihood infeasible in practice (see Hamilton and Susmel, 1994; Gray, 1996, for discussions). To circumvent this problem, many authors have proposed simplifications, either in the model or in its estimation. In this paper, we consider three popular methods in the econometric literature to estimate RS-GARCH models: Gray (1996), Klaassen (2002) and Haas et al. (2004). Refer to Appendix A for a description of these approaches. We note that Gray's RS-GARCH model was fitted in Hardy et al. (2006).

2.1.4 Distribution of the error term

In all of the previous models, the error distribution is usually normally distributed but other assumptions are possible as well. In this work, the following distributions were used for the error term: (1) the normal distribution (abbreviated NORM); (2) the Student's t-distribution (abbreviated STD); (3) the generalized error distribution (abbreviated GED) as defined by Nelson (1991); (4) the normal inverse Gaussian distribution (abbreviated NIG). Skewed versions of these distributions are also considered where skewness is introduced by the method of Fernandez and Steel (1998). These distributions and their skewed counterparts are available from the R software (R Development Core Team, 2012) as part of the fGarch package (Wuertz and Chalabi, 2012) and the fBasics package (Wuertz, 2010).

We note that it is common to combine GARCH models with heavy tailed error distributions. For example, when Nelson (1991) introduced the EGARCH model, he proposed to use the GED distribution. In addition, Bollerslev (1987) supports the use of the GARCH model with a Student's t-distribution based on its fit to five different monthly stock price indices for the U.S. economy. Moreover, the distribution of the error term in RS models is not restricted to the normal law and estimation of the parameters is not more complicated when distributions other than the normal are considered (see Hamilton, 2008). For example, Elliott and Miao (2009) advocate the use of a RS model with a Student's t-distribution for the error term.

2.1.5 Models used

All of the models used in this work were estimated by maximum likelihood with the R software. The GARCH models enumerated in Section 2.1.1 were considered[1] with the different error distributions of Section 2.1.4. These models are available in the rgarch package (Ghalanos, 2011). To refer to these models in our tables, we use the model name followed by the error distribution in parentheses. We add an S to the abbreviated name of the error distribution to denote the skewed version of the error distribution. For example, the EGARCH model with a skewed Student's t-distribution is denoted by EGARCH (SSTD).

We considered both RS and mixture (abbreviated MIX) models. Mixture models can be seen as RS models where the transition probability matrix has identical rows. These models were combined with all of the error distributions[2] of Section 2.1.4. We refer to these models

[1]We restricted ourselves to GARCH(1,1) representations, i.e., we allowed only one lag of the shock and variance terms in the GARCH equation.

[2]We let the kurtosis parameter switch across regimes but the skewness parameter was held constant in all regimes.

in our tables in a manner analogous to GARCH models. For example, we denote the RS model with a normal distribution by RS (NORM). Note that the RS (NORM) model in this paper is the same model as the RS lognormal (RSLN) model of Hardy (2001). Mixture models were used to determine whether regime persistence is important in the RS models. In all but one of the RS models that we fitted, we used two regimes, the exception being the RS3 (NORM) which includes three normal regimes. We note that independent models with the distributions of Section 2.1.4 were also considered.

Moreover, we used many models in the RS-GARCH class but we limited the error distribution in these models to the normal and the skewed normal. The models of Gray (1996), Klaassen (2002) and Haas et al. (2004) discussed in Appendix A were fitted as described in the original papers. We also exploited Haas (2010)'s generalization of the model of Haas et al. (2004) which involves an APARCH structure in each regime and a skewed normal distribution for the error term. Similarly as in that article, we did not estimate the power coefficient of the APARCH equations but we rather set it to one or two. For example, that model with a power coefficient of 1 is denoted by RS-APARCH-Haas (SNORM; pow=1) in our tables. We also extended Klaassen's RS-GARCH model in an equivalent way (see Reher and Wilfling, 2011, for a similar model) and considered many variants of Klaassen's model such as a RS-EGARCH. Whenever we do not specify which approach is used to estimate a RS-GARCH model, it is assumed that it is Klaassen's approach that is employed. We also considered a restricted version of Klaassen's model which is a MIX-GARCH model with coefficients α_{s_t} in equation (2) (see Appendix A) set to zero. With a skewed normal distribution for the error term, this model is denoted by MIX-GARCH (SNORM; $\alpha = 0$).

Moreover, we want to note that almost all of the RS-GARCH models that we fitted only allow the constant term in the GARCH equations to switch across regimes[3]. When this is not the case, we added the mention *full* in parentheses to indicate that all parameters are

[3]Haas (2010) finds that RS-GARCH models with only the constant term switching in the GARCH equations are preferred according to fit criteria.

allowed to switch across regimes. This restriction was adopted due to parsimony concerns; all but one of the models considered have 10 parameters or less.

Finally, a total of 78 models are considered in this article. Regarding the desired features of an econometric model, it should be noted that the very large majority of these models have time-varying volatility and jumps in the volatility process are present in RS models. For example, Chernov et al. (2003) fitted an extensive set of continuous-time financial models to the Dow Jones Industrial Average index and concluded that “abrupt changes in the volatility are an essential ingredient of a successful model.” Classic continuous-time stochastic volatility and jump-diffusion models were not considered in our article because once discretized, they behave very similarly to the previous models (see Nelson, 1990; Duan, 1997; Duan et al., 2006) with an added estimation complexity.

2.2 Overview of tests

The capability of the models to fit the data and to replicate patterns of time-varying volatility was checked with various criteria and statistical tests. First, the global fit among models was compared with log-likelihood values, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). More details on these criteria can be found in Klugman et al. (2008).

The Ljung-Box test (see Shumway and Stoffer, 2011) with 10 lags was applied to residuals and to squared residuals to check if there is autocorrelation in the residuals and if there is still some heteroskedasticity that has not been accounted for in the model. Only results for squared residuals are shown in our tables since, for all models, the test never rejects the null hypothesis of no autocorrelation in the residuals (based on a 5% significance level). We also used the ARCH Lagrange multiplier test (see Tsay, 2010, p. 114) with 10 lags as an additional heteroskedasticity test. Finally, the normality of the residuals was assessed with the Shapiro-Wilk and Jarque-Bera tests.

The tests mentioned in the previous paragraph rely on the assumption that residuals are normally distributed. In many instances in our models, the error distribution is not normal and these tests cannot be directly applied on the residuals. Moreover, the indicator or weighted residuals, which are often used in the context of RS models, are not normally distributed even when the error distribution is normal (see Freeland et al., 2009). Freeland et al. (2009) propose a way to obtain residuals in RS models which are normally distributed but the drawback of their approach is that we do not obtain a single set of residuals.

To compute residuals which are normally distributed using the same method for all models, we use the approach presented in Haas (2010) which is based on the Rosenblatt transform (Rosenblatt, 1952). To this end, we calculate the quantities $u_t = \widehat{F}(y_t \mid y_{t-1}, y_{t-2}, \ldots)$ for each t, where $\widehat{F}(y_t \mid y_{t-1}, y_{t-2}, \ldots)$ denotes the conditional cumulative distribution function (CDF) of the observation at time t, y_t, under the estimated model. We then transform these values using the inverse of the standard normal CDF and this resulting batch of residuals should behave as a sequence of i.i.d standard normal variates if the model is well specified.

2.3 Results

In this section, we analyze the quality of the fit of the models. Tables 1 and 2 present our results for a subset of the models that we considered. These are either important benchmarks or models that fit the data well. Results for all of the 78 models are given in Tables 8 and 9 of Appendix B.

A first observation that we can draw from Table 1 is that the inclusion of the financial crisis did not affect the relative ranking of benchmark models; the RS (NORM) still has a better fit than the GARCH (NORM) and many other simpler models. This confirms the results of Hardy (2001, 2003), even 10 years later. However, more recent econometric models that are also parsimonious do better with respect to the BIC. In Hardy et al. (2006), Gray's RS-GARCH had a better global fit, and it is still the case here. Klassen's RS-GARCH does even better. APARCH and EGARCH models with skewed error distributions are among the top models in terms of the BIC but it is the RS-EGARCH (SNORM) that achieves the highest score. The gain in fit over the RS (NORM) is approximately 20 points, which is roughly equivalent to the gain in fit of the RS (NORM) when compared to the NORM model. This can provide an idea of how important that improvement is. Moreover, when APARCH models are combined with RS, it is generally the mixture version of these models that is preferred. This entails that the role of RS is mainly to provide a possibility for the volatility to jump and that persistence in volatility may be better explained by GARCH-type dynamics than solely by regime persistence. We now take a look at the residual analysis for these models.

Table 1: Fit summary

Model	Params	Log-Lik	AIC	BIC
RS-EGARCH (SNORM)	10	1210.2	1200.2	1177.7
EGARCH (SSTD)	7	1198.1	1191.1	1175.4
MIX-APARCH-Haas (SNORM; pow=1)	8	1200.0	1192.0	1174.1
MIX-GARCH (SNORM; $\alpha = 0$)	7	1194.9	1187.9	1172.2
APARCH (SNIG)	8	1195.9	1187.9	1169.9
MIX-APARCH (SNORM; pow=1)	9	1198.3	1189.3	1169.1
MIX-GARCH (NORM)	7	1191.7	1184.7	1169.0
RS-GARCH-Klaassen (NORM)	8	1192.3	1184.3	1166.3
RS-APARCH (NORM; pow=2)	9	1192.8	1183.8	1163.6
RS-GARCH-Gray (NORM)	8	1188.4	1180.4	1162.5
RS (SGED)	9	1184.8	1175.8	1155.6
RS (NORM)	6	1174.7	1168.7	1155.2
SNIG	4	1164.2	1160.2	1151.2
GARCH (NORM)	4	1163.6	1159.6	1150.6
MIX (NORM)	5	1161.8	1156.8	1145.6
NORM	2	1138.7	1136.7	1132.2

Table 2: Residual analysis

Model	ARCH-LM	Ljung-Box	Jarque-Bera	Shapiro-Wilk
RS-EGARCH (SNORM)	0.469	0.431	0.046	0.022
EGARCH (SSTD)	0.046	0.035	0.980	0.445
MIX-APARCH-Haas (SNORM; pow=1)	0.027	0.020	0.607	0.300
MIX-GARCH (SNORM; $\alpha = 0$)	0.183	0.135	0.836	0.084
APARCH (SNIG)	0.083	0.074	0.909	0.295
MIX-APARCH (SNORM; pow=1)	0.134	0.095	0.671	0.154
MIX-GARCH (NORM)	0.236	0.180	0.518	0.015
RS-GARCH-Klaassen (NORM)	0.224	0.164	0.355	0.015
RS-APARCH (NORM; pow=2)	0.272	0.196	0.194	0.009
RS-GARCH-Gray (NORM)	0.247	0.178	0.142	0.020
RS (SGED)	0.062	0.053	0.566	0.541
RS (NORM)	0.028	0.014	0.186	0.072
SNIG	0.000	0.000	0.981	0.985
GARCH (NORM)	0.864	0.843	0.000	0.000
MIX (NORM)	0.000	0.000	0.437	0.422
NORM	0.000	0.000	0.000	0.000

Table 2 contains the p-values of the statistical tests which were discussed in Section 2.2. Results from that table indicate that there is no model that performs best overall since the best models (with respect to the BIC) do not necessarily pass all heteroskedasticity and normality tests. Nevertheless, there are a few models, such as the MIX-GARCH (SNORM; $\alpha = 0$) and the MIX-APARCH (SNORM; pow=1), that have a good global fit and that do pass these tests at a 5% significance level. The fact that there is not a single model or a class of models that dominates other candidates in all aspects will have to be taken into consideration when evaluating different risks.

The log-likelihood values as well as their penalized versions (AIC and BIC), normality and heteroskedasticity tests only measure how a model globally fits the data. This can be very interesting if one wants to draw inferences on the dynamics of a financial asset or market. When issuing equity-linked insurance, the left tail of the returns' distribution is most important because of the asymmetry in the payoff. Moreover, given that these insurance products typically have long maturities (three to 20 years), models need to be able to replicate stock market crashes that are long-lasting. This will be the focus of the next section.

3 Left tail analysis

Over the course of financial history, there have been many stock market crashes that took between six months to five years to fully recover. At the time of this writing (January 2012), the S&P500 still has not fully recovered from its low of March 2009. Models need to have the capability to generate accumulation factors with a sufficiently fat left tail over long periods of time for an investment guarantee to mature in-the-money. In this section, we intend to measure this element by comparing the cost of the investment guarantee generated by our various models to the actual cost incurred during the financial crisis.

3.1 Assumptions

We assume that an insurer writes an investment guarantee on an asset or portfolio that tracks the S&P500 total return index. The contract is issued T years (where $T = 3, 5, 7$ or 10) prior to the maturity date of February 28, 2009. We selected this maturity date because it corresponds to the lowest end-of-month index value during the financial crisis. The initial investment and the guarantee at maturity are both 100$ and we assume that the policyholder will not die or lapse its contract until the end. Fees, expressed as a percentage of the fund value, are deducted monthly from the fund. They comprise a 0.5% annual management expense ratio (MER) and an additional fee reserved to fund the guarantee that depends on the investment horizon T. The total fees that were used for maturities of $T = 3, 5, 7$ and 10 are 5.0%, 3.5%, 2.5% and 2.0%, respectively[4].

In order to set up an appropriate reserve, the insurance company uses data available from February 1956 to February 2009 $- T$ to estimate its model. It can then stochastically project the cost of the guarantee at maturity, calculate a risk measure and discount it to obtain the reserve amount needed at inception of the contract. This way of managing risk within investment guarantees is commonly called the actuarial approach. In our analysis, the discounting part is not necessary because we will be comparing the cost of the guarantee at maturity. The risk measures that we employed are the 99% VaR and the 95% CTE.

The interest in this out-of-sample analysis is to check whether the reserve forecasted by the models is large enough to meet the insurer's obligations in February 2009. It is important to note that our objective is not to point out the appropriateness of the reserve or the risk measure that should be used by actuaries. It is for actuaries and financial analysts to decide of the severity of the financial crisis and to argue whether it is a one in 20 event or a one in 200 event. Rather, we want to investigate if simple and complex models are capable of generating low returns over long-term periods.

3.2 Results

Based on the assumptions presented in the previous section, we calculated the 99% VaR and the 95% CTE of the guarantee cost based on 400,000 simulated paths for maturities of $T = 3, 5, 7$ and 10 for a total of 62 models[5]. Tables 3 and 4 show our results for the same subset of models that was previously analyzed in Section 2.3. Results for all models can be found in Tables 10 and 11 of Appendix B. The first line of each panel gives the

[4]The fee reserved to fund the guarantee was established based on the approach of Hardy (2003, p. 142). The method consists in equating the arbitrage-free valuation of the guarantee income to that of the value of the underlying put option sold under the Black-Merton-Scholes framework. For that purpose, we used a risk-free interest rate of 3% and assumed a volatility of the underlying asset corresponding to 1.1 times the empirical in-sample volatility. The factor of 1.1 is arbitrary and reflects a small margin of conservatism with respect to pricing.

[5]A small portion of the 78 models that were fitted in Section 2 was discarded in our simulation analyses since they did not provide a good fit to the data and they were not important benchmarks.

out-of-sample provision that would have been necessary in February 2009 to break even given the observed market performance. For example, over the three-year period ending in February 2009, the S&P500 total return index had a cumulative return of -38.8% so that the cost of the guarantee at maturity was 47.5% of the initial fund value (under assumptions of Section 3.1). Over the 10-year period that included the burst of the internet bubble, the observed cumulative return was -29.5% and the corresponding out-of-sample cost was 42.4%.

Table 3: Guarantee costs–3 and 5-year contracts

Model	3-Year		5-Year	
	95% CTE	99% VaR	95% CTE	99% VaR
Out-Of-Sample	47.5		40.6	
RS-GARCH-Gray (NORM)	40.3	46.7	41.1	48.8
RS (NORM)	36.8	43.3	37.4	45.4
RS-GARCH-Klaassen (NORM)	36.8	42.6	36.7	44.1
MIX-GARCH (SNORM; $\alpha = 0$)	36.7	42.9	38.4	46.1
MIX-GARCH (NORM)	36.2	42.0	36.1	43.3
MIX-APARCH-Haas (SNORM; pow=1)	35.9	42.8	35.6	44.2
RS-APARCH (NORM; pow=2)	35.7	42.0	35.1	42.7
MIX-APARCH (SNORM; pow=1)	35.3	42.3	34.8	43.6
APARCH (SNIG)	34.6	41.6	32.4	41.1
EGARCH (SSTD)	34.1	41.1	33.1	42.0
RS-EGARCH (SNORM)	34.0	41.2	36.8	46.0
SNIG	30.9	35.9	28.7	35.4
MIX (NORM)	30.9	35.8	29.0	35.7
NORM	30.0	34.7	27.8	34.2
RS (SGED)	27.8	33.6	31.3	38.5
GARCH (NORM)	27.1	32.2	28.1	35.2

First, we focus our analysis on short-term investment horizons (i.e., three and five-year periods). For a contract of three years, all models fail to generate a 95% CTE or 99% VaR that is high enough to meet the requirements of the financial crisis. However, the gap between the risk measures and the out-of-sample result is not that large. For a contract of five years, most models fail at the 95% CTE level but not at the 99% VaR level. There is some similarity between results of three and five-year contracts. For instance, the RS (NORM) model is one of the most conservative, indicating an important potential to generate negative returns over three and five-year periods. This had to be expected since in many scenarios, the latent Markov chain stays in its high volatility–low return regime for extended periods of time. More recent and sophisticated models also do comparably well over these periods. Moreover, the range of the 95% CTE across models with a good fit varies between 34% and 40% for a three-year period and between 32% and 41% for a five-year period. This is relatively narrow when compared to periods of seven and 10 years.

We now switch our focus to longer-term horizons of seven and 10 years. For a seven-year period, there are some models in the RS-GARCH class that do generate a 95% CTE that is close to the out-of-sample value of 36.4%. However, for a 10-year period this is not the case; most models generate a 95% CTE that is half or below half of the out-of-sample value of 42.4%. In other words, we would have needed to put aside twice as much money to meet the capital requirements of a 10-year contract issued in February 1999. During that period, the S&P500 index experienced two important crashes: the internet bubble and the financial crisis of the late-2000s. Moreover, the range of possible reserves for models providing a good fit to the data is much larger than over shorter-term periods. For example, the 95% CTE of the RS-EGARCH (SNORM) for a 10-year period is double that of the RS (NORM); for periods of three and five years the relative difference was less than 10%. This highlights the growing uncertainty of future returns and accumulation factors which implies that long-term investment risk is difficult to evaluate with accuracy. Finally, we note that recent econometric models do generally better than benchmark models in providing conservative figures but they still fail to generate very low returns over long-term periods.

Table 4: Guarantee costs–7 and 10-year contracts

Model	7-Year		10-Year	
	95% CTE	99% VaR	95% CTE	99% VaR
Out-Of-Sample	36.4		42.4	
MIX-GARCH (SNORM; $\alpha = 0$)	36.2	45.4	24.5	37.2
RS-EGARCH (SNORM)	36.1	47.3	20.3	35.1
RS-GARCH-Gray (NORM)	34.5	44.1	19.3	32.5
MIX-APARCH-Haas (SNORM; pow=1)	34.2	44.7	20.8	34.5
MIX-APARCH (SNORM; pow=1)	33.3	44.3	18.8	33.1
RS-GARCH-Klaassen (NORM)	32.5	41.5	19.0	31.9
MIX-GARCH (NORM)	32.2	41.0	18.9	31.3
RS-APARCH (NORM; pow=2)	32.0	41.4	18.5	31.1
EGARCH (SSTD)	30.8	41.9	17.8	32.2
APARCH (SNIG)	30.6	41.5	18.6	33.3
RS (NORM)	27.6	37.5	10.2	20.8
GARCH (NORM)	23.9	33.0	11.4	22.4
RS (SGED)	20.9	29.6	5.9	12.0
MIX (NORM)	20.1	28.9	5.6	10.7
SNIG	19.8	28.5	5.3	10.3
NORM	18.9	27.3	4.9	9.3

It is worth pointing out that the risk measures generated by the RS (NORM) model over a 10-year period are much less conservative than those of more complex models that also provide an adequate fit to the data. A parameter that has a large influence on the capability of the RS (NORM) model to generate low returns over an extended period of time is p_{22} which represents the probability of staying in the high volatility–low return regime. For the period going from February 1956 to February 1999, we estimated this parameter at 61% for the S&P500 index[6]. To illustrate the importance of that parameter, we repeated our simulations by modifying it to 80% and leaving all other parameters unchanged. With that change, the RS (NORM) model generates a 95% CTE of 47.4%, which is almost a five-fold increase with respect to the original 95% CTE of 10.2%. It is now even sufficient to cover

[6]This result is in line with the estimated parameters given in Hardy (2001) who considered a similar estimation period.

the out-of-sample value of 42.4%. We must stress that we are not suggesting to calibrate the value of p_{22} to better match the out-of-sample CTE. Rather, we want to point out the importance that parameter has on the left tail of the RS model, i.e., the calculation of the CTE is strongly influenced by its value. It is important to be aware of this element since Hardy (2001) showed that p_{22} is the parameter which is estimated with the most uncertainty in the RS model.

The previous example illustrates that to evaluate long-term investment risk, it is more prudent to analyze the results given by various models instead of focusing on the output of a single model. We showed that many models having a good global fit can generate very different risk measures when we considered a period of 10 years. Hence, it is important to be aware that model risk is important in the context of investment guarantees and that it is difficult to accurately estimate the distribution of the guarantee cost over long periods of time.

4 Dynamic hedging

4.1 Background

The traditional actuarial approach to managing risks is to set up a reserve at the inception of the contract to meet future obligations with a high probability. This reserve is usually updated as time passes by and as new information becomes available. However, no matter how well the reserve is managed, the insurance company always assumes the underlying risk of the contract. The reserve's role is to absorb the risk rather than to eliminate it.

Advances in financial theory due mainly to the work of Fischer Black, Robert Merton and Myron Scholes showed that it is theoretically possible to exactly replicate some financial risks (derivatives) with tradable assets under a certain set of assumptions. This is known as hedging. A growing number of insurance companies have now established hedging strategies with the objective of mitigating the financial risk on their equity-linked insurance products.

In the Black-Merton-Scholes (BMS) framework, one can replicate the payoff of an investment guarantee, which is a essentially a put option, by trading in the underlying stock and the risk-free asset. For that replication to be perfect, rather stringent assumptions are required: the market model should follow a geometric Brownian motion (GBM), there should be no market frictions (no transaction costs and no constraints on trading) and trading in continuous time should be possible so that the replicating portfolio can be continuously rebalanced. Although market frictions are usually small for large investment banks and insurance companies, trading in continuous time is not possible, and, as shown in Section 2.3, the market model is far from being a GBM. Consequently, dynamically hedging the long-term investment guarantee using the basic hedging strategy of the BMS framework will not perfectly replicate the payoff of the guarantee. Thus, hedging under imperfect conditions entails a risk.

During the time of the contract, the company will incur gains or losses each time it will rebalance its hedging portfolio because the required investment in the updated portfolio will generally be different from the current portfolio value. These gains and losses are a stream of cash inflows and outflows which are called hedging errors. Hedging errors come mainly from two sources: discretization error and model error. Discretization error results from the fact that rebalancing cannot be done in continuous time. Model error stems from the fact that the true market model is different from a GBM. It is not possible to know at contract inception whether hedging errors will lead to a gain or to a profit. Hence, the insurance company must set up a reserve to protect itself from that uncertainty.

4.2 Objectives and assumptions

In this section, we aim to investigate to what extent model risk matters when a long-term investment guarantee is dynamically hedged based on the BMS framework. More precisely, we will restrict ourselves to the standard Black-Scholes (BS) delta hedge (see Hardy, 2003). This analysis will enable us to evaluate the robustness of that hedging strategy with respect to its underlying assumption that the market model is a GBM.

For each of the models previously estimated[7], we generated 400,000 market scenarios under the real-world probability measure based on a monthly frequency and investment horizons of three and 10 years. We then applied the BS delta hedge dynamically through time (at a monthly frequency) on each of these simulated paths. This way, we were able to calculate the present value of hedging errors (PVHE) for each simulation. We discounted the hedging errors with a 3% interest rate which corresponds roughly to the average 1-Month Treasury Constant Maturity rate for periods of three and 10 years prior to the financial crisis. We remain again in an out-of-sample context and assume that the investment guarantee matures in February 2009.

First, we may look at the 95% CTE of the PVHE generated by each model and analyze whether model specific dynamics influence that risk measure. If the 95% CTE is highly variable across models, then this implies that the BS delta hedge is not robust with respect to the GBM assumption, i.e., there is high model risk. Second, we can compare the 95% CTE generated by our models to the out-of-sample values observed during the financial crisis to check whether models underestimate hedging risk or not.

We will now state the assumptions used in the BS delta hedge. The risk-free rate was set to 3% for the whole length of the contract. The constant volatility parameter was estimated by the in-sample volatility at the inception of the contract. Whenever transactions costs were included, they were set to represent 0.2% of the change in the market value of the stock position used for hedging.

[7]A small portion of the 78 models that were fitted in Section 2 was discarded in our simulation analyses since they did not provide a good fit to the data and they were not important benchmarks. Hence, 62 models are considered here.

4.3 Results

Tables 5 and 6 show the standard deviation (abbreviated StDev) and the 95% CTE of the PVHE for an investment guarantee with a maturity of three and 10 years, respectively, for the same subset of models that were previously analyzed in Sections 2.3 and 3.2. Results for all models can be found in Tables 12 and 13 of Appendix B. Each panel also includes the mean volatility of the simulated paths. The mean volatility of the NORM model corresponds to the volatility input used in the BS delta hedge (i.e., the in-sample empirical volatility). Moreover, the *incl.* and *excl.* column titles denote values with and without transaction costs, respectively. We first focus on a maturity of three years.

Table 5: PVHE–3-year contract

Model	Mean Vol. of Simul.	PV of Hedging Errors			
		StDev		95% CTE	
		Excl.	Incl.	Excl.	Incl.
MIX-APARCH (SNORM; pow=1)	0.137	2.18	2.24	5.94	6.45
MIX-GARCH (SNORM; $\alpha = 0$)	0.138	2.33	2.40	5.83	6.38
RS-APARCH (NORM; pow=2)	0.138	2.31	2.37	5.75	6.30
EGARCH (SSTD)	0.134	2.16	2.22	5.62	6.15
MIX-GARCH (NORM)	0.140	2.24	2.30	5.61	6.16
MIX-APARCH-Haas (SNORM; pow=1)	0.140	2.09	2.15	5.53	6.07
RS-GARCH-Klaassen (NORM)	0.140	2.17	2.23	5.38	5.92
APARCH (SNIG)	0.135	2.06	2.12	5.24	5.77
RS-EGARCH (SNORM)	0.131	2.08	2.14	5.07	5.61
RS-GARCH-Gray (NORM)	0.140	1.97	2.01	4.96	5.47
RS (NORM)	0.142	1.85	1.89	4.94	5.46
GARCH (NORM)	0.130	1.98	2.02	4.43	4.96
SNIG	0.144	1.57	1.60	4.12	4.61
MIX (NORM)	0.145	1.56	1.59	4.03	4.52
RS (SGED)	0.126	2.04	2.09	3.74	4.26
NORM	0.145	1.29	1.30	3.07	3.57

Table 6: PVHE–10-year contract

Model	Mean Vol. of Simul.	PV of Hedging Errors			
		StDev		95% CTE	
		Excl.	Incl.	Excl.	Incl.
EGARCH (SSTD)	0.148	2.17	2.29	7.54	8.22
APARCH (SNIG)	0.148	2.16	2.29	7.54	8.24
RS-EGARCH (SNORM)	0.149	2.07	2.21	7.14	7.87
MIX-APARCH (SNORM; pow=1)	0.148	2.04	2.17	7.06	7.78
MIX-GARCH (SNORM; $\alpha = 0$)	0.151	1.97	2.11	6.63	7.38
MIX-APARCH-Haas (SNORM; pow=1)	0.151	1.89	2.02	6.55	7.27
GARCH (NORM)	0.154	1.79	1.87	6.03	6.68
RS-APARCH (NORM; pow=2)	0.147	1.69	1.80	5.45	6.15
MIX-GARCH (NORM)	0.147	1.62	1.73	5.19	5.90
RS-GARCH-Klaassen (NORM)	0.147	1.61	1.72	5.13	5.84
RS-GARCH-Gray (NORM)	0.143	1.40	1.49	4.28	4.93
RS (NORM)	0.144	1.18	1.26	3.65	4.24
MIX (NORM)	0.144	0.94	0.99	2.64	3.18
RS (SGED)	0.144	0.95	0.99	2.63	3.15
SNIG	0.143	0.91	0.97	2.59	3.14
NORM	0.144	0.70	0.72	1.72	2.23

A first observation that we can make from Table 5 is that for most models the mean volatility of the simulated paths is generally below that of the in-sample volatility of 14.5%. This occurs because models are generally in a low volatility state at the start of the projection in February 2006. In such a situation, when the mean volatility of the simulated paths is lower than the volatility used for hedging, the distribution of hedging errors is shifted towards the left (see Hardy, 2003, p. 152). Another remark that can be made is that the NORM model generates the smallest standard deviation and 95% CTE of the PVHE. This is of course expected as the BS delta hedge should perform best under its own assumption of a GBM. When this assumption is not valid, we see that model risk can more than double the 95% CTE of the PVHE. That risk measure varies between 3.0 and 7.5 (assuming no transaction costs) when all models are considered.

Moreover, the 95% CTE of the PVHE can be decomposed into two parts: one part accounts for the discretization error and the remaining part for model error. The discretization error corresponds to roughly 3.0 since it is the value associated with the NORM model. The excess over that figure represents model error. Hence, we note here that model error can account for more than half of the total hedging error so that it must be taken into account. To better illustrate the influence of model risk, we must look at its impact in relative terms. The PVHE is not the only cost that the insurer has to bear in its hedging program; rather, it is the uncertain part of its total cost. Its total cost also comprises the initial purchase of the replicating portfolio whose price is known *a priori* and corresponds to that of the underlying put option in the equity-linked product. Under the assumptions of Section 4.2, the initial hedge cost is 12.0 in this example. Therefore, under many sophisticated models the hedging errors can represent more than half of that initial cost which is substantial.

Lastly, we may look at the impact of transaction costs on the PVHE. At first, it may be reasonable to suppose that the proportion of transaction costs contained in the 95% CTE of the PVHE is roughly constant across models, however this is not what we observe. For all models, transaction costs in the 95% CTE of the PVHE are approximately constant in absolute terms (0.5) but not in relative terms. Moreover, the standard deviation of the PVHE does not increase by much when transaction costs are included. This implies that the main cause of the high 95% CTE observed in some models is not more variability in the movement of the stock position used in the hedge (since this would entail higher transaction costs) but rather the direction of that movement which is more one-sided (hedging errors tend to cancel each other out much less frequently than under the NORM model). We now repeat our example with a maturity of 10 years.

First, we must stress that it is difficult to directly compare the values in Tables 5 and 6; the models are estimated based on a different sample, they do not necessarily start from the same volatility state and the present value factor gives less weight to the hedging errors that occur later. However, conclusions that can be reached for a 10-year maturity go in the same direction as those that were made previously for a three-year maturity. For instance, the range of the 95% CTE of the PVHE across models is large (it is now even much wider than for a three-year maturity) and this entails that model risk is important. While that risk measure is under a value of 8.0 for most models, it can reach a figure as high as 15.1 (see Table 13). The discretization error given by the NORM model is 1.7 in this example so that model error now represents the vast majority of the total hedging error. For example, the model error component in the 95% CTE of the PVHE of the EGARCH (SSTD) model corresponds to 5.8 which is almost 80% of the total error. Therefore, model risk has an even larger influence for a maturity of 10 years. Morever, if we put that same model error component in perspective to the initial hedge cost (which is 10.5), we deduce that model error alone can amount to more than half of that initial cost. Finally, concerning the impact of transaction costs, we also observe that increases in the 95% CTE of the PVHE from one model to the other are not accompanied by proportional increases in transaction costs.

The results and discussions presented in this section illustrate that different volatility dynamics can lead to very different reserves for the PVHE. Moreover, we showed that model risk is very important to consider, especially for a maturity of 10 years. For that maturity, model risk represents the majority of the uncertainty related to the PVHE and can significantly increase hedge costs. Since the 95% CTE of the PVHE was very unstable across models (it was in the range of 1.7 to 15.1) for a maturity of 10 years, we must seriously question the robustness of the BS delta hedge with respect to the GBM assumption since its effectiveness was highly dependent on the underlying market model.

4.4 Bootstrap

Given the growing importance of dynamic hedging in the insurance industry, we further investigate the behavior of the PVHE using bootstrap. More precisely, we now wish to examine whether the PVHE reserves forecasted by our models are consistent with those that can be obtained with bootstrapped data. Bootstrap is a technique that consists in creating samples by drawing with replacement from the original sample (this is commonly called resampling). Because asset returns are serially correlated in their squares, it is preferable to resample blocks of data to take that dependence into account. For this exercise, we employed all available data (i.e., February 1956 to December 2010) and generated 400,000 samples with blocks of different sizes. Table 7 shows the 95% CTE of the PVHE using bootstrapped data (with and without transaction costs, denoted respectively by *incl.* and *excl.* in the column titles) for horizons of three and 10 years.

Table 7: PVHE based on bootstrapped data

Bootstrap Blocks	3-Year 95% CTE		10-Year 95% CTE	
	Excl.	Incl.	Excl.	Incl.
Blocks of 12 (Annual)	5.20	5.69	4.53	5.18
Blocks of 6 (Semi-Annual)	5.29	5.77	4.25	4.91
Blocks of 4	5.17	5.65	4.04	4.71
Blocks of 3 (Quarter)	5.08	5.56	3.90	4.55
Independent (Monthly)	4.80	5.27	3.38	4.01

A first observation that can be made is that the 95% CTE of the PVHE generally increases with the block size which demonstrates that an independent bootstrap would underestimate the distribution of hedging errors. Moreover, for a three-year maturity, we obtain a reserve of 5.3 (without transaction costs) when the data is bootstrapped with semi-annual blocks. This is consistent with results obtained in Table 5 for models that provide a good fit to the data. However, for a 10-year maturity this statement does not hold because the reserve based on bootstrapped data is in the low range of those generated by most models. This may suggest that some models may overestimate the risk of hedging, but jumping to such a conclusion may not be very cautious since models that do generate high PVHE also provide a good fit to the data. We also need to stress that the data is based upon one sample path of the true market model so that bootstrap results may not give a complete picture of future possible dynamics.

4.5 Out-of-sample

In this final example, we investigate the out-of-sample PVHE using observed data from the S&P500 total return index. The objective here is to determine whether there were moments in history where the BS delta hedge was not effective and generated high PVHE for contracts of three and 10 years. For example, suppose that a contract is issued in March 1998 and matures three years later in March 2001. Applying the BS delta hedge to the S&P data for that period yields a PVHE of 4.4 (assuming no transaction costs) under the same assumptions used in the previous exercises (see Section 4.2). Figure 1 presents the out-of-sample PVHE (assuming no transaction costs) for three-year contracts maturing from January 1985 to December 2010.

The vertical dotted line in Figure 1 is simply a marker for June 2008. Hence, the points located to the right of that line represent the PVHE for contracts maturing when the market crashed significantly during the financial crisis. A first observation that can be made is that the PVHE for contracts maturing in the late 1980s reached over 6.0. For contracts

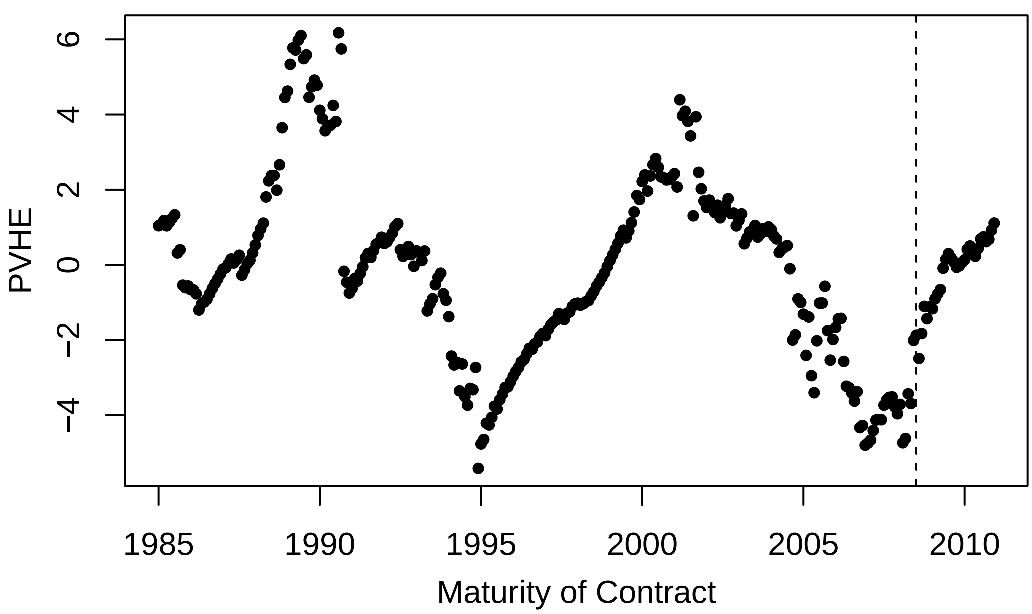

Figure 1: Empirical PVHE through time for a 3-year contract

maturing right after the internet bubble, the PVHE could go as high as 4.4. Therefore, there are periods in the data where the observed PVHE reach levels corresponding to the 95% CTE of the PVHE obtained under many models with a good fit to the data (see Table 5). Nevertheless, we note that the BS delta hedge did not do so badly for products maturing during the recent financial crisis as the PVHE were always under 2.0.

We now repeat that same exercise for 10-year contracts maturing from January 1985 to December 2010. Figure 2 shows our results. Barring one exception[8] the PVHE in Figure 2 are always under 3.0. This value is much lower than the 95% CTE obtained under many models and bootstrapped data. It is important to stress that there are not too many disjoint 10-year periods in the data. The fact that we did not observe many 10-year periods that generate high PVHE does not imply that this will be the case in the future. The risk of having much higher hedging errors exists as shown in Section 4.3.

[8]The PVHE for a 10-year contract maturing in October 2008 is 5.9.

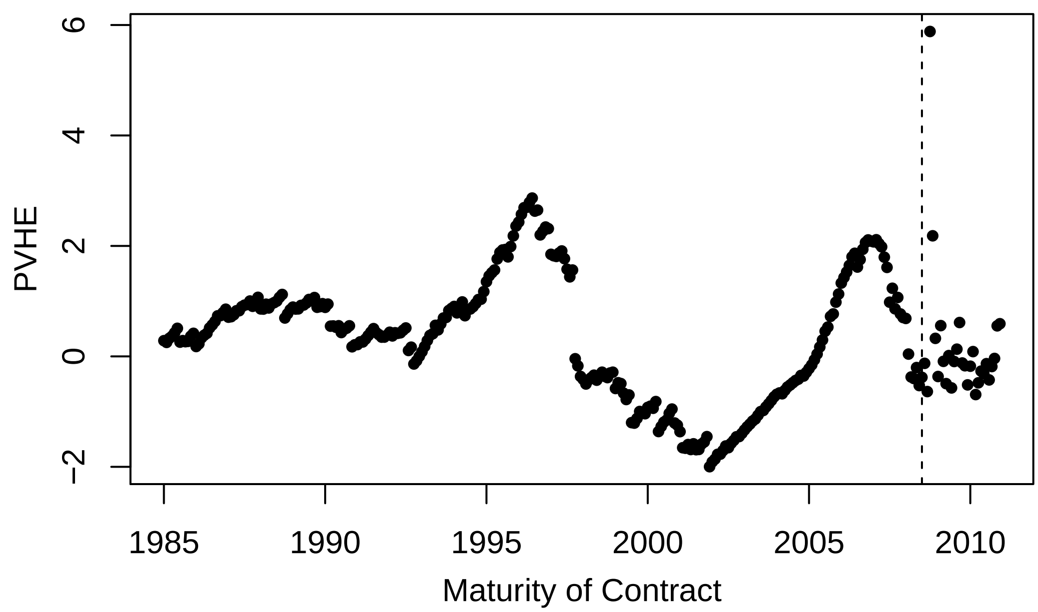

Figure 2: Empirical PVHE through time for a 10-year contract

5 Conclusion

In this paper, we analyzed the risk underlying investment guarantees using a very large set of financial econometric models. We found that despite the excellent fit of many models, too few of them are capable of generating low returns over long periods of time (say five to 10 years), i.e., the type of cumulative returns observed during the financial crisis of the late-2000s. This is a crucial element for insurance companies that issue equity-linked insurance. We generated scenarios under each of these models to check the robustness of the Black-Scholes delta hedging strategy. We found that hedging losses can be significant implying that large reserves for hedging errors are required. We also showed that results can be very variable across models and this highlights that model risk is important to take into consideration when managing the risk underlying investment guarantees.

How can we make the Black-Scholes delta hedging strategy more robust, i.e., less sensitive to model risk? One may rebalance weekly, daily or even several times a day. However, it is well documented in the finance literature that daily or high frequency data has much fatter tails than monthly data (see Campbell et al., 1997; Tsay, 2010), meaning even more deviations from the GBM. Thus, it is not clear how effective rebalancing more frequently can really be. Moreover, one may use other Greeks in the replicating portfolio, such as gamma, vega and rho, which are respectively the sensitivity of the derivative's price with respect to large price movements of the underlying asset, its volatility and the risk-free rate. Using these Greeks may help reduce hedging errors, but their effectiveness will depend on the true market model. Finally, one may turn to the financial engineering literature and use the true replicating portfolio of some of these 78 models that fit the data well. In addition to being more challenging to implement due to the incompleteness of the underlying market, their reliance on a risk premium parameter or process makes it difficult to predict as to how robust this approach will be. Those three ways of improving the delta hedging strategy will be investigated in an upcoming paper. We conclude by quoting Rantala (2006) who mentions that “in the face of model risk, rather than to base decisions on a single selected ‘best’ model, the modeller can base his inference on an entire set of models by using model averaging.”

Acknowledgments

The authors acknowledge the financial support provided by the Natural Sciences and Engineering Research Council of Canada and the Society of Actuaries.

References

Bauwens, L., Preminger, A., and Rombouts, J. V. K. (2010). Theory and inference for a Markov switching GARCH model. *The Econometrics Journal*, 13(2):218–244.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3):542–547.

Bollerslev, T. (2008). Glossary to ARCH (GARCH). CREATES Research Papers 2008-49, School of Economics and Management, University of Aarhus.

Boudreault, M. and Panneton, C.-M. (2009). Multivariate models of equity returns for investment guarantees valuation. *North American Actuarial Journal*, 13(1):36–53.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, New Jersey, second printing, with corrections edition.

Chernov, M., Gallant, A. R., Ghysels, E., and Tauchen, G. (2003). Alternative models for stock price dynamics. *Journal of Econometrics*, 116(1-2):225–257.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106.

Duan, J.-C. (1997). Augmented GARCH(p, q) process and its diffusion limit. *Journal of Econometrics*, 79(1):97–127.

Duan, J.-C., Ritchken, P., and Sun, Z. (2006). Approximating GARCH-jump models, jump-diffusion processes, and option pricing. *Mathematical Finance*, 16(1):21–52.

Elliott, R. J. and Miao, H. (2009). VaR and expected shortfall: a non-normal regime switching framework. *Quantitative Finance*, 9(6):747–755.

Fernandez, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.

Freeland, R. K., Till, M. C., and Hardy, M. R. (2009). Residual based assessment of regime switching models. Working paper, University of Waterloo.

Ghalanos, A. (2011). *rgarch: Flexible GARCH modelling in R.* R package version 1.92.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.

Goldfeld, S. M. and Quandt, R. E. (1973). A Markov model for switching regressions. *Journal of Econometrics*, 1(1):3–15.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62.

Haas, M. (2010). Skew-normal mixture and Markov-switching GARCH processes. *Studies in Nonlinear Dynamics & Econometrics*, 14(4):Article 1.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, 2(4):493–530.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Hamilton, J. D. (2008). Regime switching models. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke.

Hamilton, J. D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1-2):307–333.

Hardy, M. R. (2001). A regime-switching model of long-term stock returns. *North American Actuarial Journal*, 5(2):41–53.

Hardy, M. R. (2002). Bayesian risk management for equity-linked insurance. *Scandinavian Actuarial Journal*, 2002(3):185–211.

Hardy, M. R. (2003). *Investment guarantees: Modeling and risk management for equity-linked life insurance*. John Wiley & Sons, New Jersey.

Hardy, M. R., Freeland, R. K., and Till, M. C. (2006). Validation of long-term equity return models for equity-linked guarantees. *North American Actuarial Journal*, 10(4):28–47.

Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, 27:363–394.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2008). *Loss models.* Wiley Series in Probability and Statistics. John Wiley & Sons Inc., New Jersey, third edition. From data to decisions.

Lamoureux, C. G. and Lastrapes, W. D. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics*, 8(2):225–234.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH$(1,1)$ model. *Econometric Theory*, 6(3):318–334.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rantala, J. (2006). On joint and separate history of probability, statistics and actuarial science. In *Festschrift for Tarmo Pukkila on his 60th birthday*, pages 261–284. Dep. Math. Stat. Philos. Univ. Tampere, Tampere.

Reher, G. and Wilfling, B. (2011). Markov-switching GARCH models in finance: a unifying framework with an application to the German stock market. CQE Working Papers 1711, Center for Quantitative Economics (CQE), University of Muenster.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23:470–472.

Shumway, R. H. and Stoffer, D. S. (2011). *Time series analysis and its applications*. Springer Texts in Statistics. Springer, New York, third edition. With R examples.

Tsay, R. S. (2010). *Analysis of financial time series*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., New Jersey, third edition.

Wong, A. C. S. and Chan, W.-S. (2005). Mixture Gaussian time series modeling of long-term market returns. *North American Actuarial Journal*, 9(4):83–94.

Wuertz, D. (2010). *fBasics: Rmetrics—Markets and Basic Statistics*. R package version 2160.85. With Rmetrics core team members.

Wuertz, D. and Chalabi, Y. (2012). *fGarch: Rmetrics—Autoregressive Conditional Heteroskedastic Modelling.* R package version 2150.81. With contribution from Michal Miklovic, Chris Boudt, Pierre Chausse and others.

A RS-GARCH models

In this appendix, we present some approaches available in the econometric literature to estimate RS-GARCH models. To illustrate these, consider the following RS-GARCH model:

$$y_t = \mu_{s_t} + \sigma_t(s_t)\eta_t, \tag{1}$$

$$\sigma_t^2(s_t) = \omega_{s_t} + \alpha_{s_t}\epsilon_{t-1}^2 + \beta_{s_t}h_{t-1}, \tag{2}$$

where $\{y_t\}_{t=1}^T$ and $\{s_t\}_{t=1}^T$ represent, respectively, the returns to be modeled and the RS process. The symbol μ_{s_t} represents the mean in regime s_t, $\sigma_t^2(s_t)$ is the conditional variance at time t in regime s_t, η_t is an independent and identically distributed innovation with zero mean and unit variance, h_{t-1} and ϵ_{t-1} are, respectively, the variance and the shock terms used to update the GARCH equation at time t. We assume that there are N regimes and the parameters of the model are $\{\mu_i, \omega_i, \alpha_i, \beta_i, \kappa_i, \xi\}_{i=1}^N$, where κ_i denotes the kurtosis parameter of the innovation (if applicable) in regime i and ξ represents the skewness parameter of the innovation (if applicable) which is constant across regimes. We now present the approaches used by Gray (1996), Klaassen (2002) and Haas et al. (2004) to estimate the RS-GARCH model.

1. Gray (1996) is among the first to propose a way to estimate a RS-GARCH model. He circumvents path dependence by defining h_{t-1} and ϵ_{t-1} in equation (2) as $\mathrm{Var}(y_{t-1} \mid \mathcal{F}_{t-2})$ and $y_{t-1} - \mathrm{E}(\mu_{s_{t-1}} \mid \mathcal{F}_{t-2})$, respectively, where $\mathcal{F}_{t-2} = (y_1, \ldots, y_{t-2})$. This definition avoids the dependence of the conditional variance $\sigma_t(s_t)$ on the entire history of regimes since the

quantities updating the GARCH equation only depend on the observed data. Therefore, the trick to avoid path dependence here is to collapse conditional variances at each time period into a single value before passing it on.

2. Klaassen (2002) improves on Gray's model by using a broader conditioning set to collapse quantities. For instance, he defines h_{t-1} and ϵ_{t-1} in equation (2) as $\mathrm{E}(\sigma^2_{t-1}(s_{t-1}) \mid \mathcal{F}_{t-1}, s_t)$ and $y_{t-1} - \mathrm{E}(\mu_{s_{t-1}} \mid \mathcal{F}_{t-1}, s_t)$, respectively.

3. Haas et al. (2004) formulate the RS-GARCH model in a manner that avoids path dependence by hypothesizing a number of GARCH processes evolving independently from each other. At each time period, one of these GARCH processes is picked by the RS process to determine the variance of the observation. This set-up is not path dependent because on the one hand, shocks are defined with respect to a common mean (i.e., the mean is not allowed to switch across regimes) and, on the other hand, the conditional variance in each regime is updated with the previous conditional variance in that same regime (and not necessarily with the conditional variance of the previous observation) so that knowledge of the previous regime is not needed. In the model of Haas et al. (2004), equations (1) and (2) are modified to:

$$y_t = \mu + \sigma_t(s_t)\eta_t,$$
$$\sigma^2_t(i) = \omega_i + \alpha_i\epsilon^2_{t-1} + \beta_i\sigma^2_{t-1}(i), \quad i = 1, \ldots, N.$$

   This model has been generalized in Haas (2010) by incorporating an APARCH structure in each regime.

Finally, we note that Bauwens et al. (2010) are able to estimate the RS-GARCH model without resorting to the above modifications through the use of Bayesian Markov chain Monte Carlo techniques.

B Results for all models

Table 8: Fit summary

Model	Params	Log-Lik	AIC	BIC
RS-EGARCH (SNORM)	10	1210.2	1200.2	1177.7
EGARCH (SSTD)	7	1198.1	1191.1	1175.4
EGARCH (SNIG)	7	1196.8	1189.8	1174.1
MIX-APARCH-Haas (SNORM; pow=1)	8	1200.0	1192.0	1174.1
MIX-GARCH (NORM; $\alpha = 0$)	6	1192.2	1186.2	1172.7
RS-APARCH-Haas (SNORM; pow=1)	9	1201.7	1192.7	1172.4
MIX-GARCH (SNORM; $\alpha = 0$)	7	1194.9	1187.9	1172.2
EGARCH (SGED)	7	1194.9	1187.9	1172.2
MIX-GJR-GARCH-Scale (SNORM)	9	1201.3	1192.3	1172.1
GJR-GARCH (SSTD)	7	1194.8	1187.8	1172.1
MIX-APARCH-Haas (SNORM; pow=2)	8	1197.3	1189.3	1171.3
APARCH (SSTD)	8	1197.2	1189.2	1171.3
MIX-EGARCH (SNORM)	9	1200.2	1191.2	1171.0
GJR-GARCH (SNIG)	7	1193.5	1186.5	1170.8
RS-GJR-GARCH-Scale (SNORM)	10	1203.1	1193.1	1170.7
GARCH (SSTD)	6	1189.9	1183.9	1170.4
APARCH (SNIG)	8	1195.9	1187.9	1169.9
GARCH (SNIG)	6	1189.0	1183.0	1169.5
RS-APARCH-Haas (SNORM; pow=2)	9	1198.5	1189.5	1169.3
MIX-APARCH (SNORM; pow=1)	9	1198.3	1189.3	1169.1
MIX-GARCH (NORM)	7	1191.7	1184.7	1169.0
MIX-APARCH (SNORM; pow=2)	9	1197.9	1188.9	1168.7
GJR-GARCH (SGED)	7	1191.4	1184.4	1168.6
EGARCH (STD)	6	1187.7	1181.7	1168.2
APARCH (SGED)	8	1194.0	1186.0	1168.0
GARCH (SGED)	6	1187.4	1181.4	1167.9
MIX-APARCH (NORM; pow=1)	8	1193.8	1185.8	1167.8
MIX-APARCH-Haas (NORM; pow=1)	7	1190.4	1183.4	1167.7
MIX-EGARCH (NORM)	8	1193.6	1185.6	1167.6
MIX-APARCH (NORM; pow=2)	8	1192.7	1184.7	1166.8
RS-GARCH-Klaassen (NORM)	8	1192.3	1184.3	1166.3
RS-APARCH (SNORM; pow=1)	10	1198.5	1188.5	1166.0
RS-APARCH (SNORM; pow=2)	10	1198.0	1188.0	1165.5
MIX-GJR-GARCH-Scale (NORM)	8	1191.4	1183.4	1165.4
GJR-GARCH (STD)	6	1184.6	1178.6	1165.1
RS-APARCH (NORM; pow=1)	9	1193.9	1184.9	1164.7
RS-EGARCH (NORM)	9	1193.8	1184.8	1164.6
RS-APARCH-Haas (NORM; pow=1)	8	1190.4	1182.4	1164.4
APARCH (STD)	7	1187.0	1180.0	1164.3
MIX-APARCH-Haas (NORM; pow=2)	7	1186.8	1179.8	1164.1
RS-APARCH (NORM; pow=2)	9	1192.8	1183.8	1163.6
MIX-GARCH (NORM; full)	9	1192.3	1183.3	1163.1
RS-GJR-GARCH-Scale (NORM)	9	1191.8	1182.8	1162.6
GARCH (STD)	5	1178.8	1173.8	1162.6
RS-GARCH-Gray (NORM)	8	1188.4	1180.4	1162.5
EGARCH (GED)	6	1181.7	1175.7	1162.2
RS-APARCH-Haas (NORM; pow=2)	8	1186.8	1178.8	1160.9
RS (SNORM)	7	1183.2	1176.2	1160.5
RS-GARCH-Klaassen (NORM; full)	10	1192.4	1182.4	1160.0

Continued on next page...

Table 8: Fit summary

Model	Params	Log-Lik	AIC	BIC
GJR-GARCH (GED)	6	1178.1	1172.1	1158.7
APARCH (GED)	7	1181.2	1174.2	1158.4
GARCH (GED)	5	1174.6	1169.6	1158.4
RS (SSTD)	9	1186.7	1177.7	1157.5
RS-GARCH-Gray (NORM; full)	10	1189.6	1179.6	1157.1
RS (SNIG)	9	1185.6	1176.6	1156.4
EGARCH (NORM)	5	1171.9	1166.9	1155.7
RS-GARCH-Haas (NORM)	7	1178.3	1171.3	1155.6
RS (SGED)	9	1184.8	1175.8	1155.6
RS (NORM)	6	1174.7	1168.7	1155.2
APARCH (NORM)	6	1171.9	1165.9	1152.4
RS (STD)	8	1177.7	1169.7	1151.7
SSTD	4	1164.3	1160.3	1151.3
SNIG	4	1164.2	1160.2	1151.2
GARCH (NORM)	4	1163.6	1159.6	1150.6
GJR-GARCH (NORM)	5	1166.4	1161.4	1150.2
RS-GARCH-Haas (NORM; full)	9	1178.9	1169.9	1149.7
RS (GED)	8	1175.2	1167.2	1149.2
SGED	4	1162.1	1158.1	1149.2
STD	3	1157.8	1154.8	1148.1
RS3 (NORM)	12	1185.4	1173.4	1146.4
MIX (NORM)	5	1161.8	1156.8	1145.6
GED	3	1154.7	1151.7	1145.0
MIX (STD)	7	1164.5	1157.5	1141.8
MIX (SGED)	8	1167.0	1159.0	1141.1
MIX (GED)	7	1163.5	1156.5	1140.8
MIX (SNIG)	8	1166.6	1158.6	1140.7
MIX (SSTD)	8	1164.6	1156.6	1138.6
NORM	2	1138.7	1136.7	1132.2

Table 9: Residual analysis

Model	ARCH-LM	Ljung-Box	Jarque-Bera	Shapiro-Wilk
RS-EGARCH (SNORM)	0.469	0.431	0.046	0.022
EGARCH (SSTD)	0.046	0.035	0.980	0.445
EGARCH (SNIG)	0.060	0.046	0.906	0.236
MIX-APARCH-Haas (SNORM; pow=1)	0.027	0.020	0.607	0.300
MIX-GARCH (NORM; $\alpha = 0$)	0.335	0.287	0.303	0.011
RS-APARCH-Haas (SNORM; pow=1)	0.035	0.027	0.657	0.310
MIX-GARCH (SNORM; $\alpha = 0$)	0.183	0.135	0.836	0.084
EGARCH (SGED)	0.117	0.093	0.124	0.036
MIX-GJR-GARCH-Scale (SNORM)	0.042	0.030	0.868	0.232
GJR-GARCH (SSTD)	0.064	0.043	0.984	0.240
MIX-APARCH-Haas (SNORM; pow=2)	0.022	0.013	0.574	0.379
APARCH (SSTD)	0.068	0.060	0.981	0.529
MIX-EGARCH (SNORM)	0.052	0.041	0.722	0.398
GJR-GARCH (SNIG)	0.090	0.062	0.917	0.119
RS-GJR-GARCH-Scale (SNORM)	0.052	0.039	0.875	0.271
GARCH (SSTD)	0.225	0.165	0.983	0.550
APARCH (SNIG)	0.083	0.074	0.909	0.295
GARCH (SNIG)	0.258	0.193	0.994	0.376
RS-APARCH-Haas (SNORM; pow=2)	0.021	0.013	0.612	0.397

Continued on next page...

Table 9: Residual analysis

Model	ARCH-LM	Ljung-Box	Jarque-Bera	Shapiro-Wilk
MIX-APARCH (SNORM; pow=1)	0.134	0.095	0.671	0.154
MIX-GARCH (NORM)	0.236	0.180	0.518	0.015
MIX-APARCH (SNORM; pow=2)	0.133	0.077	0.623	0.117
GJR-GARCH (SGED)	0.169	0.124	0.128	0.022
EGARCH (STD)	0.356	0.342	0.000	0.000
APARCH (SGED)	0.157	0.141	0.133	0.053
GARCH (SGED)	0.321	0.250	0.384	0.166
MIX-APARCH (NORM; pow=1)	0.279	0.220	0.150	0.006
MIX-APARCH-Haas (NORM; pow=1)	0.253	0.246	0.002	0.000
MIX-EGARCH (NORM)	0.134	0.110	0.139	0.010
MIX-APARCH (NORM; pow=2)	0.289	0.211	0.194	0.008
RS-GARCH-Klaassen (NORM)	0.224	0.164	0.355	0.015
RS-APARCH (SNORM; pow=1)	0.144	0.102	0.733	0.165
RS-APARCH (SNORM; pow=2)	0.137	0.080	0.644	0.120
MIX-GJR-GARCH-Scale (NORM)	0.207	0.186	0.050	0.006
GJR-GARCH (STD)	0.378	0.342	0.001	0.000
RS-APARCH (NORM; pow=1)	0.308	0.246	0.156	0.006
RS-EGARCH (NORM)	0.155	0.125	0.155	0.011
RS-APARCH-Haas (NORM; pow=1)	0.256	0.248	0.002	0.000
APARCH (STD)	0.431	0.436	0.001	0.000
MIX-APARCH-Haas (NORM; pow=2)	0.277	0.247	0.002	0.000
RS-APARCH (NORM; pow=2)	0.272	0.196	0.194	0.009
MIX-GARCH (NORM; full)	0.315	0.260	0.340	0.011
RS-GJR-GARCH-Scale (NORM)	0.142	0.122	0.062	0.008
GARCH (STD)	0.563	0.541	0.001	0.000
RS-GARCH-Gray (NORM)	0.247	0.178	0.142	0.020
EGARCH (GED)	0.672	0.647	0.000	0.000
RS-APARCH-Haas (NORM; pow=2)	0.280	0.249	0.002	0.000
RS (SNORM)	0.083	0.073	0.014	0.090
RS-GARCH-Klaassen (NORM; full)	0.277	0.218	0.276	0.011
GJR-GARCH (GED)	0.712	0.673	0.000	0.000
APARCH (GED)	0.741	0.739	0.000	0.000
GARCH (GED)	0.729	0.705	0.000	0.000
RS (SSTD)	0.053	0.046	0.991	0.854
RS-GARCH-Gray (NORM; full)	0.539	0.499	0.029	0.003
RS (SNIG)	0.040	0.030	0.971	0.595
EGARCH (NORM)	0.886	0.868	0.000	0.000
RS-GARCH-Haas (NORM)	0.476	0.460	0.001	0.000
RS (SGED)	0.062	0.053	0.566	0.541
RS (NORM)	0.028	0.014	0.186	0.072
APARCH (NORM)	0.923	0.919	0.000	0.000
RS (STD)	0.032	0.020	0.001	0.000
SSTD	0.000	0.000	0.992	0.992
SNIG	0.000	0.000	0.981	0.985
GARCH (NORM)	0.864	0.843	0.000	0.000
GJR-GARCH (NORM)	0.883	0.862	0.000	0.000
RS-GARCH-Haas (NORM; full)	0.508	0.488	0.000	0.000
RS (GED)	0.034	0.017	0.220	0.069
SGED	0.000	0.000	0.390	0.584
STD	0.000	0.000	0.008	0.006
RS3 (NORM)	0.002	0.001	0.291	0.184

Continued on next page...

Table 9: Residual analysis

Model	ARCH-LM	Ljung-Box	Jarque-Bera	Shapiro-Wilk
MIX (NORM)	0.000	0.000	0.437	0.422
GED	0.000	0.000	0.000	0.001
MIX (STD)	0.000	0.000	0.985	0.992
MIX (SGED)	0.000	0.000	0.000	0.000
MIX (GED)	0.000	0.000	0.879	0.850
MIX (SNIG)	0.000	0.000	0.990	0.995
MIX (SSTD)	0.000	0.000	0.989	0.990
NORM	0.000	0.000	0.000	0.000

Table 10: Guarantee costs–3 and 5-year contracts

Model	3-Year		5-Year	
	95% CTE	99% VaR	95% CTE	99% VaR
Out-Of-Sample	47.5		40.6	
RS-GARCH-Gray (NORM)	40.3	46.7	41.1	48.8
MIX-APARCH (SNORM; pow=2)	38.6	45.4	39.8	48.1
RS-APARCH-Haas (SNORM; pow=2)	38.3	45.2	38.9	47.3
RS-APARCH (SNORM; pow=2)	38.0	44.8	39.1	47.6
MIX-APARCH-Haas (SNORM; pow=2)	38.0	45.0	38.9	47.7
MIX (SGED)	37.2	43.1	37.3	44.5
RS (NORM)	36.8	43.3	37.4	45.4
RS-GARCH-Klaassen (NORM)	36.8	42.6	36.7	44.1
MIX-GARCH (SNORM; $\alpha = 0$)	36.7	42.9	38.4	46.1
GJR-GARCH (SNIG)	36.6	44.3	38.0	47.7
GJR-GARCH (SSTD)	36.5	44.2	37.3	47.1
GJR-GARCH (SGED)	36.4	44.0	38.9	48.4
MIX-GARCH (NORM)	36.2	42.0	36.1	43.3
MIX-APARCH (NORM; pow=2)	36.2	42.3	36.0	43.7
MIX-GARCH (NORM; full)	35.9	41.9	36.7	44.4
MIX-APARCH-Haas (SNORM; pow=1)	35.9	42.8	35.6	44.2
RS-APARCH (NORM; pow=2)	35.7	42.0	35.1	42.7
RS-APARCH-Haas (SNORM; pow=1)	35.7	42.5	35.7	44.3
MIX-GARCH (NORM; $\alpha = 0$)	35.5	41.5	36.3	43.9
MIX-EGARCH (SNORM)	35.5	42.4	35.0	43.6
RS-GARCH-Gray (NORM; full)	35.4	42.2	40.4	48.2
MIX-APARCH (SNORM; pow=1)	35.3	42.3	34.8	43.6
RS-APARCH (SNORM; pow=1)	35.2	42.0	35.1	43.9
RS-GARCH-Klaassen (NORM; full)	34.9	40.9	35.7	43.4
APARCH (SNIG)	34.6	41.6	32.4	41.1
RS-EGARCH (NORM)	34.6	41.1	32.6	40.7
EGARCH (SNIG)	34.5	41.5	33.9	42.7
EGARCH (SGED)	34.5	41.4	35.1	43.9
MIX-EGARCH (NORM)	34.4	40.9	32.4	40.8
MIX-GJR-GARCH-Scale (SNORM)	34.4	41.2	34.4	43.2
APARCH (SGED)	34.4	41.4	34.6	43.1
RS-APARCH (NORM; pow=1)	34.3	40.9	32.6	41.1
MIX-APARCH (NORM; pow=1)	34.2	40.8	32.5	41.0
APARCH (SSTD)	34.2	41.3	32.1	40.7
EGARCH (SSTD)	34.1	41.1	33.1	42.0
RS-EGARCH (SNORM)	34.0	41.2	36.8	46.0
MIX-GJR-GARCH-Scale (NORM)	32.6	39.0	30.7	38.8
APARCH (NORM)	32.4	38.6	31.9	39.1

Continued on next page. . .

Table 10: Guarantee costs–3 and 5-year contracts

Model	3-Year		5-Year	
	95% CTE	99% VaR	95% CTE	99% VaR
Out-Of-Sample	47.5		40.6	
EGARCH (NORM)	31.4	37.3	30.8	38.5
RS-APARCH-Haas (NORM; pow=1)	31.1	37.7	27.4	36.0
SSTD	30.9	36.0	28.7	35.3
MIX-APARCH-Haas (NORM; pow=1)	30.9	37.5	27.3	36.0
SNIG	30.9	35.9	28.7	35.4
MIX (NORM)	30.9	35.8	29.0	35.7
SGED	30.8	35.8	28.7	35.3
GJR-GARCH (NORM)	30.7	36.1	29.6	36.9
MIX-APARCH-Haas (NORM; pow=2)	30.5	37.2	26.9	35.5
RS-APARCH-Haas (NORM; pow=2)	30.5	37.2	27.2	35.9
MIX (SSTD)	30.4	35.4	28.2	35.0
NORM	30.0	34.7	27.8	34.2
GARCH (SSTD)	29.5	36.0	30.8	39.7
GARCH (SNIG)	29.4	35.9	31.1	39.7
GARCH (SGED)	29.1	35.3	31.1	39.8
RS (SGED)	27.8	33.6	31.3	38.5
RS (SNORM)	27.3	32.9	29.8	37.0
GARCH (NORM)	27.1	32.2	28.1	35.2
RS (SSTD)	26.9	33.0	31.3	38.6
RS3 (NORM)	26.7	32.5	32.5	39.9
STD	26.0	31.1	20.5	27.7
RS-GARCH-Haas (NORM)	25.9	31.8	21.6	29.5
GED	25.6	30.7	19.5	26.7
RS-GARCH-Haas (NORM; full)	24.2	29.8	29.3	37.8

Table 11: Guarantee costs–7 and 10-year contracts

Model	7-Year		10-Year	
	95% CTE	99% VaR	95% CTE	99% VaR
Out-Of-Sample	36.4		42.4	
RS-GARCH-Gray (NORM; full)	41.0	52.3	30.9	45.9
GJR-GARCH (SGED)	40.3	52.7	34.6	51.4
GJR-GARCH (SNIG)	38.6	51.0	30.6	47.6
MIX-APARCH (SNORM; pow=2)	37.9	48.0	27.7	41.2
RS-APARCH (SNORM; pow=2)	37.2	47.2	26.5	39.8
GJR-GARCH (SSTD)	36.5	48.7	27.8	44.5
RS-APARCH-Haas (SNORM; pow=2)	36.3	46.5	25.3	38.3
MIX-GARCH (SNORM; $\alpha = 0$)	36.2	45.4	24.5	37.2
MIX-APARCH-Haas (SNORM; pow=2)	36.2	46.3	25.0	38.1
RS-EGARCH (SNORM)	36.1	47.3	20.3	35.1
RS-GARCH-Klaassen (NORM; full)	35.3	45.1	22.3	36.2
MIX-GJR-GARCH-Scale (SNORM)	35.2	46.7	22.1	37.6
RS-GARCH-Gray (NORM)	34.5	44.1	19.3	32.5
APARCH (SGED)	34.3	45.2	21.4	37.0
MIX-APARCH-Haas (SNORM; pow=1)	34.2	44.7	20.8	34.5
MIX-GARCH (NORM; full)	34.1	43.6	21.2	34.5
RS-APARCH (SNORM; pow=1)	33.8	44.6	19.4	33.9
MIX-GARCH (NORM; $\alpha = 0$)	33.7	43.1	20.5	33.4
EGARCH (SGED)	33.5	44.2	21.9	36.5
MIX-EGARCH (SNORM)	33.5	44.0	18.8	33.2

Continued on next page...

Table 11: Guarantee costs–7 and 10-year contracts

Model	7-Year		10-Year	
	95% CTE	99% VaR	95% CTE	99% VaR
Out-Of-Sample	36.4		42.4	
RS-APARCH-Haas (SNORM; pow=1)	33.5	43.9	20.1	33.5
MIX-APARCH (SNORM; pow=1)	33.3	44.3	18.8	33.1
MIX-APARCH (NORM; pow=2)	32.8	42.2	19.2	32.0
RS-GARCH-Klaassen (NORM)	32.5	41.5	19.0	31.9
EGARCH (SNIG)	32.4	43.1	19.7	34.7
MIX-GARCH (NORM)	32.2	41.0	18.9	31.3
MIX (SGED)	32.1	41.2	18.5	31.0
RS-APARCH (NORM; pow=2)	32.0	41.4	18.5	31.1
EGARCH (SSTD)	30.8	41.9	17.8	32.2
APARCH (SSTD)	30.6	41.4	17.2	31.8
APARCH (SNIG)	30.6	41.5	18.6	33.3
RS-APARCH (NORM; pow=1)	29.8	40.4	14.4	27.6
MIX-EGARCH (NORM)	29.5	39.8	14.6	27.6
MIX-APARCH (NORM; pow=1)	29.5	39.9	14.5	27.7
RS-EGARCH (NORM)	29.4	39.5	14.6	27.6
GARCH (SGED)	28.6	39.7	18.2	33.4
MIX-GJR-GARCH-Scale (NORM)	28.1	38.5	12.1	24.2
GARCH (SNIG)	28.0	39.1	16.4	31.1
GJR-GARCH (NORM)	27.7	37.0	14.5	26.8
RS (NORM)	27.6	37.5	10.2	20.8
EGARCH (NORM)	27.2	36.9	13.1	25.6
GARCH (SSTD)	27.1	38.2	15.9	30.2
RS (SSTD)	25.4	34.4	9.8	20.5
APARCH (NORM)	24.9	35.0	4.2	6.0
RS (SNORM)	24.1	33.0	6.6	13.2
GARCH (NORM)	23.9	33.0	11.4	22.4
RS3 (NORM)	23.2	32.6	6.4	12.9
MIX (SSTD)	22.9	31.9	4.5	7.7
MIX-APARCH-Haas (NORM; pow=1)	22.0	33.2	7.1	13.6
RS-APARCH-Haas (NORM; pow=1)	21.8	32.9	7.1	14.1
RS (SGED)	20.9	29.6	5.9	12.0
MIX-APARCH-Haas (NORM; pow=2)	20.6	31.5	6.7	12.8
RS-APARCH-Haas (NORM; pow=2)	20.4	31.4	6.9	13.6
MIX (NORM)	20.1	28.9	5.6	10.7
SSTD	19.8	28.6	5.7	11.3
SNIG	19.8	28.5	5.3	10.3
SGED	19.5	28.2	5.2	10.0
NORM	18.9	27.3	4.9	9.3
RS-GARCH-Haas (NORM; full)	16.0	27.0	7.8	14.3
RS-GARCH-Haas (NORM)	12.3	22.1	2.9	0.0
STD	9.1	17.6	1.4	0.0
GED	8.5	17.0	1.0	0.0

Table 12: PVHE–3-year contract

Model	Mean Vol. of Simul.	PV of Hedging Errors			
		StDev		95% CTE	
		Excl.	Incl.	Excl.	Incl.
GJR-GARCH (SSTD)	0.140	2.66	2.71	7.50	8.04
GJR-GARCH (SNIG)	0.139	2.60	2.66	7.26	7.81

Continued on next page. . .

Table 12: PVHE–3-year contract

Model	Mean Vol. of Simul.	PV of Hedging Errors			
		StDev		95% CTE	
		Excl.	Incl.	Excl.	Incl.
RS-GARCH-Haas (NORM)	0.138	2.59	2.63	7.26	7.75
GJR-GARCH (SGED)	0.138	2.54	2.60	6.97	7.53
RS-APARCH (SNORM; pow=2)	0.140	2.59	2.66	6.84	7.40
MIX-APARCH (SNORM; pow=2)	0.140	2.58	2.65	6.82	7.38
MIX-APARCH-Haas (SNORM; pow=2)	0.146	2.39	2.45	6.80	7.35
MIX-APARCH-Haas (NORM; pow=2)	0.135	2.40	2.43	6.75	7.22
RS-APARCH-Haas (NORM; pow=2)	0.135	2.39	2.43	6.73	7.19
RS-APARCH-Haas (SNORM; pow=2)	0.146	2.35	2.41	6.66	7.20
GARCH (SSTD)	0.130	2.51	2.56	6.54	7.07
RS-GARCH-Haas (NORM; full)	0.132	2.51	2.55	6.47	6.97
MIX-APARCH-Haas (NORM; pow=1)	0.136	2.24	2.28	6.30	6.76
RS-GARCH-Gray (NORM; full)	0.129	2.46	2.51	6.27	6.81
RS-APARCH-Haas (NORM; pow=1)	0.136	2.23	2.26	6.25	6.71
GARCH (SNIG)	0.129	2.44	2.49	6.18	6.71
RS-GARCH-Klaassen (NORM; full)	0.137	2.39	2.46	6.18	6.73
MIX-EGARCH (SNORM)	0.137	2.23	2.28	6.03	6.54
RS-APARCH (SNORM; pow=1)	0.137	2.19	2.25	5.98	6.48
MIX-APARCH (SNORM; pow=1)	0.137	2.18	2.24	5.94	6.45
MIX-GARCH (SNORM; $\alpha = 0$)	0.138	2.33	2.40	5.83	6.38
RS-APARCH (NORM; pow=2)	0.138	2.31	2.37	5.75	6.30
MIX-APARCH (NORM; pow=1)	0.137	2.13	2.18	5.72	6.24
MIX-EGARCH (NORM)	0.137	2.15	2.20	5.72	6.23
MIX-APARCH (NORM; pow=2)	0.138	2.30	2.36	5.72	6.26
RS-APARCH (NORM; pow=1)	0.137	2.12	2.17	5.69	6.21
RS-EGARCH (NORM)	0.137	2.14	2.19	5.69	6.20
MIX-GARCH (NORM; full)	0.137	2.26	2.32	5.64	6.19
MIX-GJR-GARCH-Scale (SNORM)	0.134	2.18	2.23	5.64	6.16
GARCH (SGED)	0.128	2.30	2.35	5.63	6.17
EGARCH (SSTD)	0.134	2.16	2.22	5.62	6.15
MIX-GARCH (NORM)	0.140	2.24	2.30	5.61	6.16
MIX-APARCH-Haas (SNORM; pow=1)	0.140	2.09	2.15	5.53	6.07
MIX (SGED)	0.140	2.20	2.26	5.49	6.03
EGARCH (SNIG)	0.134	2.14	2.19	5.46	5.99
RS-GARCH-Klaassen (NORM)	0.140	2.17	2.23	5.38	5.92
APARCH (SSTD)	0.135	2.09	2.14	5.37	5.91
RS-APARCH-Haas (SNORM; pow=1)	0.139	2.05	2.11	5.35	5.89
EGARCH (SGED)	0.134	2.11	2.17	5.33	5.88
MIX-GARCH (NORM; $\alpha = 0$)	0.135	2.20	2.26	5.28	5.82
APARCH (SNIG)	0.135	2.06	2.12	5.24	5.77
APARCH (SGED)	0.135	2.06	2.12	5.21	5.76
APARCH (NORM)	0.145	1.84	1.90	5.13	5.69
RS-EGARCH (SNORM)	0.131	2.08	2.14	5.07	5.61
RS-GARCH-Gray (NORM)	0.140	1.97	2.01	4.96	5.47
RS (NORM)	0.142	1.85	1.89	4.94	5.46
MIX-GJR-GARCH-Scale (NORM)	0.134	1.97	2.01	4.92	5.43
STD	0.145	1.72	1.73	4.62	5.06
MIX (SSTD)	0.145	1.69	1.72	4.61	5.08
EGARCH (NORM)	0.135	1.88	1.93	4.58	5.13
SSTD	0.145	1.67	1.69	4.55	5.02

Continued on next page. . .

Table 12: PVHE–3-year contract

Model	Mean Vol. of Simul.	PV of Hedging Errors			
		StDev		95% CTE	
		Excl.	Incl.	Excl.	Incl.
GARCH (NORM)	0.130	1.98	2.02	4.43	4.96
GJR-GARCH (NORM)	0.132	1.89	1.93	4.37	4.90
SNIG	0.144	1.57	1.60	4.12	4.61
MIX (NORM)	0.145	1.56	1.59	4.03	4.52
RS (SSTD)	0.124	2.05	2.10	3.98	4.47
RS (SGED)	0.126	2.04	2.09	3.74	4.26
GED	0.144	1.50	1.52	3.74	4.19
RS (SNORM)	0.127	2.05	2.10	3.71	4.22
SGED	0.144	1.46	1.49	3.66	4.16
RS3 (NORM)	0.126	1.90	1.95	3.55	4.06
NORM	0.145	1.29	1.30	3.07	3.57

Table 13: PVHE–10-year contract

Model	Mean Vol. of Simul.	PV of Hedging Errors			
		StDev		95% CTE	
		Excl.	Incl.	Excl.	Incl.
GJR-GARCH (SGED)	0.167	4.06	4.21	15.07	15.87
RS-GARCH-Gray (NORM; full)	0.160	4.05	4.20	14.45	15.25
GJR-GARCH (SNIG)	0.162	3.74	3.87	13.72	14.46
GJR-GARCH (SSTD)	0.159	3.62	3.74	13.16	13.87
RS-GARCH-Haas (NORM; full)	0.159	3.72	3.79	12.29	12.87
GARCH (SGED)	0.158	3.13	3.25	11.33	12.06
GARCH (SSTD)	0.154	3.09	3.19	10.88	11.56
GARCH (SNIG)	0.155	3.02	3.14	10.81	11.51
MIX-GJR-GARCH-Scale (SNORM)	0.148	2.53	2.66	8.85	9.58
APARCH (SGED)	0.152	2.35	2.49	8.36	9.10
RS-GARCH-Klaassen (NORM; full)	0.150	2.36	2.50	8.14	8.90
EGARCH (SGED)	0.152	2.28	2.42	8.06	8.80
MIX-APARCH (SNORM; pow=2)	0.154	2.26	2.41	7.88	8.65
RS-APARCH (SNORM; pow=2)	0.154	2.25	2.39	7.85	8.62
MIX-APARCH-Haas (SNORM; pow=2)	0.155	2.17	2.31	7.72	8.46
EGARCH (SSTD)	0.148	2.17	2.29	7.54	8.22
APARCH (SNIG)	0.148	2.16	2.29	7.54	8.24
EGARCH (SNIG)	0.149	2.16	2.28	7.52	8.21
RS-APARCH-Haas (SNORM; pow=2)	0.154	2.12	2.26	7.51	8.25
APARCH (SSTD)	0.148	2.15	2.27	7.49	8.18
RS-EGARCH (SNORM)	0.149	2.07	2.21	7.14	7.87
MIX-EGARCH (SNORM)	0.148	2.06	2.19	7.13	7.84
RS-APARCH (SNORM; pow=1)	0.148	2.05	2.18	7.10	7.81
MIX-APARCH (SNORM; pow=1)	0.148	2.04	2.17	7.06	7.78
MIX-GARCH (SNORM; $\alpha = 0$)	0.151	1.97	2.11	6.63	7.38
MIX-APARCH-Haas (SNORM; pow=1)	0.151	1.89	2.02	6.55	7.27
RS-APARCH-Haas (SNORM; pow=1)	0.151	1.85	1.99	6.42	7.15
MIX-GARCH (NORM; full)	0.147	1.90	2.03	6.27	7.00
GARCH (NORM)	0.154	1.79	1.87	6.03	6.68
MIX-GARCH (NORM; $\alpha = 0$)	0.147	1.79	1.92	5.82	6.55
MIX-EGARCH (NORM)	0.145	1.73	1.84	5.80	6.47
MIX-APARCH (NORM; pow=1)	0.146	1.72	1.83	5.78	6.45
RS-APARCH (NORM; pow=1)	0.146	1.72	1.83	5.77	6.44

Continued on next page. . .

Table 13: PVHE–10-year contract

Model	Mean Vol. of Simul.	PV of Hedging Errors			
		StDev		95% CTE	
		Excl.	Incl.	Excl.	Incl.
RS-EGARCH (NORM)	0.145	1.72	1.83	5.77	6.44
MIX-APARCH-Haas (NORM; pow=2)	0.143	1.70	1.77	5.63	6.17
RS-APARCH-Haas (NORM; pow=2)	0.143	1.67	1.74	5.51	6.05
GJR-GARCH (NORM)	0.150	1.64	1.72	5.49	6.15
RS-APARCH (NORM; pow=2)	0.147	1.69	1.80	5.45	6.15
MIX-APARCH (NORM; pow=2)	0.147	1.67	1.79	5.40	6.11
RS-APARCH-Haas (NORM; pow=1)	0.144	1.59	1.67	5.29	5.84
MIX-APARCH-Haas (NORM; pow=1)	0.144	1.59	1.67	5.27	5.83
MIX (SGED)	0.148	1.63	1.74	5.26	5.97
EGARCH (NORM)	0.151	1.53	1.64	5.21	5.90
MIX-GARCH (NORM)	0.147	1.62	1.73	5.19	5.90
RS-GARCH-Klaassen (NORM)	0.147	1.61	1.72	5.13	5.84
MIX-GJR-GARCH-Scale (NORM)	0.141	1.55	1.64	4.99	5.62
RS-GARCH-Haas (NORM)	0.146	1.43	1.47	4.38	4.85
RS-GARCH-Gray (NORM)	0.143	1.40	1.49	4.28	4.93
RS (SSTD)	0.144	1.33	1.41	4.27	4.85
RS (NORM)	0.144	1.18	1.26	3.65	4.24
RS3 (NORM)	0.144	1.13	1.20	3.42	3.99
RS (SNORM)	0.144	1.18	1.24	3.41	3.95
SSTD	0.144	1.04	1.10	3.10	3.63
MIX (SSTD)	0.144	1.02	1.07	3.06	3.59
APARCH (NORM)	0.146	0.86	0.93	2.71	3.26
MIX (NORM)	0.144	0.94	0.99	2.64	3.18
RS (SGED)	0.144	0.95	0.99	2.63	3.15
SNIG	0.143	0.91	0.97	2.59	3.14
STD	0.143	0.92	0.94	2.56	2.94
SGED	0.143	0.82	0.88	2.24	2.79
GED	0.143	0.74	0.76	1.89	2.26
NORM	0.144	0.70	0.72	1.72	2.23