Actuarial Inference and Applications of Hidden Markov Models

by

Matthew Charles Till

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2011

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Matthew Charles Till

Abstract

Hidden Markov models have become a popular tool for modeling long-term investment guarantees. Many different variations of hidden Markov models have been proposed over the past decades for modeling indexes such as the S&P 500, and they capture the tail risk inherent in the market to varying degrees. However, goodness-of-fit testing, such as residual-based testing, for hidden Markov models is a relatively undeveloped area of research. This work focuses on hidden Markov model assessment, and develops a stochastic approach to deriving a residual set that is ideal for standard residual tests. This result allows hidden-state models to be tested for goodness-of-fit with the well developed testing strategies for single-state models.

This work also focuses on parameter uncertainty for the popular long-term equity hidden Markov models. There is a special focus on underlying states that represent lower returns and higher volatility in the market, as these states can have the largest impact on investment guarantee valuation. A Bayesian approach for the hidden Markov models is applied to address the issue of parameter uncertainty and the impact it can have on investment guarantee models.

Also in this thesis, the areas of portfolio optimization and portfolio replication under a hidden Markov model setting are further developed. Different strategies for optimization and portfolio hedging under hidden Markov models are presented and compared using real world data. The impact of parameter uncertainty, particularly with model parameters that are connected with higher market volatility, is once again a focus, and the effects of not taking parameter uncertainty into account when optimizing or hedging in a hidden Markov are demonstrated.

Acknowledgements

I would first like to express unbound thanks to my two supervisors, Mary Hardy and Keith Freeland. Mary has been supportive in an infinite number of ways throughout the years, from providing me with life-changing opportunities, to providing me with a place to stay and a meal to eat, to sharing thoughts and advice on any topic that arose, to giving me several needed pushes, and ultimately providing the needed non-push that all led to the successful completion of this degree. Keith has shared with me much of his wisdom and serenity over the years, of which I can only aspire to acquire. I am very lucky to have gotten to know them both.

I would like to give heartfelt thanks to Professors Paul Marriott, Yulia Gel, Joseph Kim, Steve Drekic and Dinghai Xu for their insightful comments and suggestions and for their participation in my examining committee, especially to Professor Marriott for help with the Bayesian analysis, Professor Gel for the letters of recommendation and Professor Kim for Latex help. I also thank Professor Wai-Sum Chan of the Chinese University of Hong Kong who graciously served as my external examiner and provided valuable insight and comments.

I offer much gratitude to the Department of Statistics and Actuarial Science for giving me this chance, and I especially would like to thank Mary Lou Dufton for the invaluable help she has provided for both my Masters and Ph. D. degrees. I also would like to acknowledge the research support I received from the department, from the University of Waterloo, from the Natural Sciences and Engineering Research Council of Canada, the Society of Actuaries and the Standard Life Assurance Company of Canada.

I also would like to thank Professor Phelim Boyle at Wilfred Laurier University for the letters of recommendation, for fostering an unmatched environment for thought-provoking conversation, and for putting up with me in his house from time to time.

I would like to express very sincere appreciation to my mom, dad, oma and opa, nana and grandad, sister Nicole and brother Travis for being such a part of who I am. I love you all.

Finally, to my wife Shelagh and to our three wonderful boys, Isaac, Patrick and Eric, I can only try to express the gratitude I feel. You are my greatest passion, and my greatest source of joy. I am and will be forever aware of with what I have been blessed. Let us explore the future together!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The notion that, in the long run, the stock market always goes up has been embedded into western civilization. The second half of the 20th century saw strong gains throughout, with the occasional significant drop, which would always be corrected over the short period following the drop.

This notion led many insurers and financial institutions in North America to include long-term guarantees for their clients investment portfolios, often times at no extra cost to the client. The assumption that the market would never go down over the long-term meant these institutions felt that the guarantees they were providing were effectively worth nothing. Imagine the surprise when the S&P 500 experienced an approximate 30% drop over the January 1999 to January 2009 period, or that the Nikkei 225 has been in a state of steady decline since 1989.

The original motivation for this work is the investment guarantee market. A simple example of an investment guarantee is the Guaranteed Minimum Maturity Benefit, which guarantees a policyholder a specified minimum amount their investment will mature for at the termination of the investment period. Investment guarantees usually fall into the category of undiversifiable risk, as the guarantees are made on the same equities. If one contract is in-the-money at maturity, they all are.

Investment guarantees are often long-term, lasting five, ten or twenty years. The Amer-

ican Academy of Actuaries proposed a stochastic methodology (AAA 2005) for reserving investment guarantee contracts, following the recommendations of the Canadian Institute of Actuaries (2001). This method involves modeling the index's monthly returns, and then using simulation to quantify the risk involved in the contract. Reserves can be held using the "actuarial approach", which requires the insurer hold enough capital to cover all losses that fall within the bottom $95\% - 99\%$ of the loss distribution, or using a dynamic hedging approach that creates a replicating portfolio that moves with the liability.

In the past, stock markets were modeled with simple, constant and tractable models. Black & Scholes (1973) in their pioneering paper on option pricing used a simple lognormal process for stock returns. In his ground-breaking paper in portfolio optimization, Markowitz (1952) modeled the variances of a group of stocks with constant parameters. Since these times, understanding of market movements and tendencies has evolved a great deal, and market decisions and strategies must evolve in the same way.

Because of the nature of the risk (low frequency, high severity), modeling the tails of the index correctly becomes quite paramount. In particular, the long-term left tail of the index is of paramount relevance when reserving using the actuarial approach, as it will determine with what level of uncertainty the chosen level of reserving will be adequate. The short term left and right tails will be more relevant when reserving such a contract using a dynamic hedge. For instance, if the contract is delta hedged, then it will be vulnerable to gamma risk of the underlying index, which is typically associated with the steeper market movements.

One key feature apparent in most stocks and indexes is the concept of volatility clustering. Markets experience periods where the price movements are relatively small, and experience other periods where price movements are larger. Adapting models to capture this volatility clustering has been the focus of much research. One idea that has enjoyed much success is the concept of underlying market 'states'. The main idea is simple: the market will go through calm periods with very steady index returns, and the market will go through much more volatile periods of uncertainty. When exactly the market changes states or how a long a market stays in particular state is unknown, and many models that employ underlying state processes categorize them as unknown.

However, educated guesses about the state of markets can be made. If the past week of

trading has seen some very sharp price movements, investors would be more likely to suggest that the market is in a more volatile state than if the past week of trading has been relatively relaxed. There is inherent information about the likelihood of the market experiencing a particular state in how its prices have been recently moving. Successful models will make use of this information to more accurately predict market movements.

One class of multiple state models that has been the focus of much research are regime-switching models, proposed by Hamilton (1989). Regime-switching models define the underlying state process to be a Markovian process. In accordance with all Markov chains, under a regime-switching model, the distribution of the next event's underlying state depends only on the current underlying state of the process (and does not depend directly on past underlying states). An attractive feature of the Markov process is that it allows for correlation in the data through the underlying state process, which can capture the volatility clustering mentioned above.

Since the CIA's task force on segregated fund investment guarantees in 2001, many equity models have been proposed. A detailed examination of many of these models is found in Hardy, Freeland and Till (2006). This paper concluded that the most promising models had an embedded regime-switching process, which was successful in accounting for periods of low volatility and high volatility over a long time horizon. Inference about regime-switching frameworks for stock returns and their applications will be the focus of this thesis.

The issue of the underlying state process being unobserved presents some non-trivial challenges. When testing a model's fit to the data, the common strategy is to test the residuals that result from the fit to see if they are properly distributed. However, in a regime-switching framework, the model parameters for any particular data point are unknown, as the underlying state is unknown (and the model parameters are dependent on the underlying state). The residual associated with that particular data point then will also be unknown. Moreover, for a regime-switching model with K regimes, and a time horizon of T, then there are K possible residuals for each data point, and correspondingly K^T possible residual sets for the entire data set.

The task of arriving at a suitable set of residuals to test is an undeveloped area of research. Freeland, Hardy & Till (2006) made use of the probabilities associated with each regime

or state of a model to arrive at a weighted average residual. This thesis will demonstrate how that method can generate an unsuitable set of residuals for standard residual tests in many cases, due to the nature of the weighted average. The magnitude of each residual is often effectively lost through the average, and the magnitudes of those residuals is an important factor in deciding whether or not a residual set is acceptable. Alternative methods of generating residuals from the possible residual sets are explored here. The objective of the research is to generate a set of residuals that maintain as much as possible the magnitude and sign of the actual residual, which, of course, is unknown because of the unknown residual process. The result arrived at for this problem is that a stochastic approach to deriving a residual set provides an ideal set for the standard residual tests. This result is very powerful, as it allows multi-state models to be tested to the same degree of effectiveness as single-state models.

One particular focus of regime-switching models that this thesis will explore is parameter uncertainty. While important for all statistical modeling, parameter uncertainty can have a larger impact for regime-switching models for financial data because of the nature of the underlying states in an equity setting. For most indexes modeled using a regime switching framework, there will typically be a high frequency regime with calm volatility, and in addition to possible other regimes, a low frequency regime with very high volatility. Because this high volatility regime is less often visited, parameter uncertainty will be more profound than it will for the low volatility regime. Yet the high volatility regime is the premier regime of interest for investment guarantee reserving. The large stock movements, both up and down, will typically be associated with this regime, and since the expected return of the high volatility regime will typically be negative under the model for most financial data, the long-term left tail of the model will also be defined by this regime. Summarily, the state of the market an investment guarantee analyst is most interested in is the state she knows the least about.

To address this issue, a Bayesian approach to estimation of the popular regime switching models in finance will be explored. Bayesian estimation involves treating the unknown model parameters as themselves random variables, first assigning a 'prior' distribution to them which effectively represents pre-existing understanding, and then using the observed data to adjust this distribution to a more likely one given the path the index has traveled. The Bayes

approach to financial regime-switching models has been developed in literature (Robert et al, 1999, Hardy, 2002 and Bauwens, Preminger and Rombouts, 2006, for example), and this thesis will employ and adapt those methods to the commonly used models in investment guarantee analysis.

Often, it is beneficial to model multiple indexes in a multivariate framework. For instance, if an investor were to create a fund that comprised of many indexes or stocks, then how those equities interacted with each other would need to be adequately captured. As was the case for single-variate models, volatility clustering is a prevalent theme among many of the world's main indexes today. Regime-switching can again be used to capture this volatility clustering in the multivariate case.

Another issue arises when trying to invest in an optimal portfolio strategy. In a single-state market, investment strategies have been developed for a very long time (see Black and Litterman(1992), Alexander, Coleman and Li(2004), etc). These methods are very effective at accomplishing their set task. However, when working in a multi-state framework, with the underlying state being unknown, one can handle the information given about a data set different ways.

Using the unconditional distribution of the market, deriving an optimal portfolio strategy would be as simple as a single state case. But, if one knew exactly which state the market was in during optimization, then the process could be made more efficient, as the information about the market would be more precise. This is not the case, as the underlying market state process is unobserved. However, for each data point there is embedded information about the likelihood of the market being in each of the underlying states. Taking advantage of this information would produce more efficient portfolios than the unconditional approach. There are different methods of using the described information about the underlying states, and some main ones will be explored in this thesis.

An extension of the optimized portfolio problem of investment guarantees is the hedging of an investment guarantee. Hedging derivative contracts has been well developed for single-state models, but hedging under a multiple state market has in general meant approximating the markets movement with a single state model. The error of this can become increasingly problematic if the markets different states become further and further distinct. Again, the

information about the underlying state can prove to be invaluable, as knowledge about the current market state will allow a hedger to more accurately assess the likeliness of a market's movements. Using this information, similar in fashion to what was done for model testing and portfolio optimization, would mean less hedging error and thus less hedging costs. This these will explore different methods of handling the underlying state process when hedging an investment guarantee contract, and show the impact of different methods using examples and real world data.

Chapter 2 will describe popular regime-switching models in a financial setting, fit them to a standard data set using frequentist estimation and discuss the implications of each for investment guarantee contracts. Chapter 3 will discuss validation of long-term equity models that have been estimated using a frequentist approach, and will introduce a method of opening up standard residual testing to regime switching models that was previously unavailable. Chapter 4 will approach estimation of the same models using a Bayesian methodology, and discuss the impact of parameter uncertainty on long-term equity distributions and compare with the results from Chapter 2. Chapter 5 will discuss portfolio optimization in a regime switching setting, and contribute to the existing literature by demonstrating the impact of parameter uncertainty and provide some insight for practitioners. Chapter 6 will discuss portfolio replication, also known as hedging, in a hidden Markov setting and demonstrate its implications for deep out-of-the-money options that resemble investment guarantees. Finally, Chapter 7 will be a discussion of future work.

Chapter 2

Long-Term Equity Return Models

2.1   Introduction

The equity models used to quantify the risks associated with long-term, deep out-of-the-money options, such as investment guarantees, must accurately model the tails the of the returns on equities over very long periods. Unfortunately, these tails are the part of the long-term equity distributions we know the least about. Modelling, say, 10-year returns directly is problematic, because we have very few non-overlapping 10-year segments of the time series available. A popular way to address this problem is to model a higher frequency (for example, annual, quarterly or monthly) time series, and then use this model to draw inference about the longer-term returns from the model. In addition to the usefulness of the model for the longer term returns, modelling the higher frequency process allows us to analyze path dependent cashflows, which will be important for assessing hedging effectiveness.

Many models for stock price returns have been developed since the 1950's. Early models, such as the geometric Brownian motion assumed by Black & Scholes (1973), assumed independence of stock price movements over different (non-overlapping) periods. Observed dependence in stock returns, either through the stock price itself or the volatility of its movements, was modeled through building autoregressive and moving average coefficients into the log-normal framework. This was later further refined by Engle (1982) and Bollerslev (1986)

through the popular class of ARCH/GARCH models. More recently, stochastic volatility (see AAA, 2005, for example) and hidden Markov models have been proposed to capture stock price movements over time.

This chapter will explore some single state models and hidden Markov models developed for long-term equity analysis. These models will be fitted using maximum likelihood estimation to US stock index data, and the long-term implications of each of the models, in particular for out-of-the-money options, will be explored. This chapter will set the stage for further chapters, where contributions to the hidden Markov model literature in finance will be made in the areas of model validation, portfolio optimization and portfolio replication.

2.2   Long-Term Equity Data

Stocks prices are often affected by many factors, including firm performance, corporate decisions, government policies and practices, consumer confidence, investor moods and systemic shifts in the overall economy. Stock prices are also affected by decisions by firms to pay out dividends. For the purposes of all equity models in this thesis, dividends will be assumed to be reinvested in the stock; stock levels will thereby represent the stock's full growth over time. Other factors that determine stock prices will not be included in the models explicitly.

Stock indices are, in general, non-stationary time series. Index levels tend to increase exponentially with time. A log-transformation, followed by a series differencing, is typically applied to stock prices before analysis; the resulting data are called log-returns.

2.2.1   The S&P 500

The Standard & Poor's 500 (S&P 500) is a weighted index of the prices of the 500 most actively traded stocks in the United States. It is one of the most widely followed indexes in finance, with a long history. The monthly index levels of the S&P 500, with dividends reinvested, from January 1950 to October 2010 are displayed in Figure 2.1(a). The data are taken from finance.yahoo.com, and the index levels for this data set are the levels at the
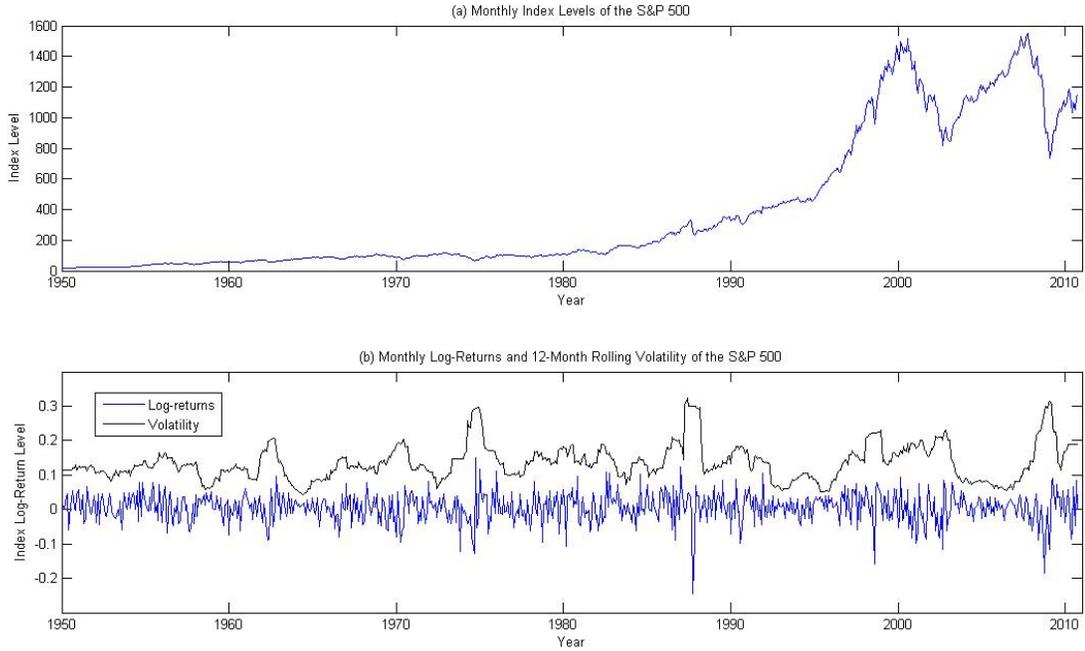
Figure 2.1: S&P 500 Monthly Index and Log-Return Levels

close of the first trading day of the month. The index is clearly increasing over time, and the magnitudes of its movements are also increasing over time. As described above, a log transformation was applied to the series, followed by it being once differenced. The series of S&P 500 log-returns, along with its twelve-month rolling volatility, is displayed in Figure 2.1(b).

Mathematically, let S_t denote the index level at the end of month t. Then the transformed series is

$$Y_t = \log\left(\frac{S_t}{S_{t-1}}\right)$$

There is no visible upward trend in the series of log-returns Y_t, nor in the volatility of the series. However, it is apparent that the volatility of the S&P 500 log-returns is not constant. The rolling-average annualized volatility of the S&P 500 exhibits significantly different levels over the past 50 years, ranging from 0.043 to 0.325. Moreover, the volatility

9

of the log-returns clusters in areas. Low volatility periods in the series are generally followed by other low volatility periods, such as the periods in the 1950's, the late 1970's to early 1980's, the mid 1990's and the mid 2000's. High volatility periods are also often followed by other high volatility returns, such as the mid-1970's, the late 1990's to early 2000's, and the most recently the late 2000's. Volatility clustering is a major driver of many of the more sophisticated models that are presented later in this chapter. Specifically for investment guarantees, failure to capture these volatility trends would be a significant drawback of any model, as, among other things, the adequacy of investment guarantee reserving depends on the ability of the reserving approach to capture the large movements in the index.

While the monthly log-returns of the S&P 500 comprise the time series to be modeled, it is important to re-emphasize that the long-term behaviour of the index is of equal concern. The monthly price movements are of concern specifically if one is dynamically hedging the liability of an option or guarantee (see Chapter 6) or if the contract in question has a 'ratchet' or other path-dependent clause attached to it. However, monthly data are also used out of necessity. There aren't enough non-overlapping 10-year periods of S&P data to effectively fit a 10-year model. While capturing the monthly price movements is important under this approach, one can obtain an idea from Figure 2.1 about how candidate models should behave in the long-term. Among other things, this means that periods of relative calm and periods of relatively uncertainty should be plausible under candidate models, and should both happen frequently enough to mimic the series displayed above.

Popular long-term equity models will be presented in the next section. Some background and reasoning for each of the models will be discussed, and all the candidate models will be fitted to the above data set using maximum likelihood estimation.

2.3 Simple Long-term Equity Models

In this section we provide a description of some simple models for financial data, fit them to the S&P 500 using maximum likelihood estimation, and discuss some of the limitations of each of the models. These models are generally unsatisfactory for investment guarantee modeling purposes. Much of the analysis in this section is presented in Hardy, Freeland & Till

(2006). It provides a foundation for the remainder of the chapter as the more complicated models presented later use the simpler models as building blocks.

All models described in this chapter assume a process of random innovations, $\epsilon_t, t = 1, \ldots, N$, where t is a time index and N is the length of the series of log-returns. For each of the models described, the random variables ϵ_t will always be assumed to be independent, identically distributed $N(0, 1)$ random variables. Other distributions are left as a subject for further research.

In this thesis, expressions such as 'the model is too thin tailed compared to the data' are often used. Let $F_X(x) = P[X \leq x]$ represent the cumulative distribution function for random variable X evaluated at x. For this specific discussion, let X_M represent the random variable of the total returns over any time period under the model, and let X_D represent the empirical random variable of the total returns over the same time period, using the observed S&P index values. Then our contention that the model is 'too thin in the left tail' relative to the data can be expressed as

$$F_{X_M}^{-1}(x) > F_{X_D}^{-1}(x) \quad \forall x \in (0, x_1)$$

where x_1 is a small value such as 0.01 or 0.05. That is, the x_1 quantile of the model is greater than the x_1 quantile of the data, for left tail values of x_1. The model has is 'too thin in the right tail' if

$$F_{X_M}^{-1}(x) < F_{X_D}^{-1}(x) \quad \forall x \in (x_2, 1)$$

where x_2 now takes on relatively larger values of 0.95 or 0.99.

2.3.1　The Independent Log-Normal Model

The independent log-normal (ILN) models the log-return series Y_t with an independent and identical Normal distribution. The process may be written as

$$Y_t = \mu + \sigma \cdot \epsilon_t$$

The ILN model is the discretely obeserved version of a geometric Brownian motion, which is one of the assumptions of the Black-Scholes option pricing formula (Black & Scholes, 1973). The model captures short term market returns reasonably well, and has the advantage of tractability, as well as consistency with the Black Scholes formula.

We find though that the ILN is generally not suitable for modeling index returns over longer terms, and particularly where tail risk is important. The model assumes a constant volatility parameter σ, and does not capture the volatility clustering described in Section 2.2. The fitted ILN models generally have tails that are too thin to capture the risk of long-term market drops. This will result in severe underestimation of the value of contracts that with payouts in such cases.

2.3.2　The GARCH Model

The Generalized Autoregressive Conditionally Heteroskedastic (GARCH) family of models, discussed extensively in Engle (1995), has been a popular choice for modeling indexes and stocks over longer terms, in discrete time. The ARCH family allows changing volatility, and therefore offers more flexibility than the ILN model. The ARCH family was first presented by Engle (1982) followed by the GARCH extension in Bollserslev (1986). The models are still popular today; Engle received the 2003 Nobel Memorial Prize in Economics, due to the invaluable contribution of the ARCH family to econometric modelling.

The GARCH model differs from the ILN by assuming a volatility process, $\sigma_t, t = 1, \ldots, N$. The volatility level at each time point depends on previous values of the volatility, and on the squared deviation of the process from its mean value.

The variance process of the GARCH(p, q) model is defined to be

$$\sigma^2(t) = \alpha_0 + \sum_{i=1}^{p} \alpha_i (Y_{t-i} - \mu)^2 + \sum_{j=1}^{q} \beta_j \sigma^2(t-j)$$

The parameters $\alpha_i, i = 0, \ldots, p$ and $\beta_j, j = 1, \ldots, q$ are generally assumed to be greater than zero, which avoids more complex constraints to ensure the variance process is always strictly positive.

The variance process is stochastic, unconditionally, but deterministic, conditional on the previous values of Y_t and σ_t^2. Given $\sigma(t)$, the process at each time point is similar to a standard ILN model:

$$Y_t | \sigma(t) = \mu + \sigma(t) \cdot \epsilon_t$$

This process does capture volatility clustering. With positive β parameters, relatively higher volatility levels for previous observations under the series will result in higher volatility levels for future observations. A single, randomly high value of the Y_t process will increase the variance in the next value of the series, and the β parameters will determine how significantly that higher variance impacts subsequent values. Similarly, lower volatilities will increase the probability that subsequent observations will have lower volatilities.

The GARCH(1,1) model is the most popularly used for equity return modelling (see, for example, the seminal text by Campbell, Lo and MacKinlay, 1997). The model has five parameters: μ, α_0, α_1, β_1 and the starting volatility $\sigma^2(0)$. The inclusion of $\sigma^2(0)$ in the likelihood can sometimes have a significant impact on the resulting values of the other parameters under maximum likelihood estimation, which can be undesirable given the parameter (usually) does not significantly impact long term projections of the series. Instead of treating the starting volatility as a free parameter, one could also set the starting volatility to the empirical average volatility of the data and have it remain fixed during model estimation.

The sum $\alpha_1 + \beta_1$ measures the persistency of the volatility process. In order for the process

μ_1	α_0	α_1	β_1	$\sigma^2(0)$
0.00700	0.00000	0.13135	0.32016	0.00114

Table 2.1: The GARCH(1,1) MLE Parameters Fitted to the S&P 500 Monthly Log-Return from Jan 1956 to Oct 2010

to be covariance-stationary we require

$$\alpha_1 + \beta_1 < 1.$$

For some monthly data, we find that fitted values of $\alpha_1 + \beta_1$ can become close to 1.0, but we do not find any cases where $\alpha_1 + \beta_1 > 1$. The issue of covariance stationarity for GARCH models for stocks is more of a problem for higher frequency data. See Diebold (1986), for example.

GARCH(1,1) and the S&P 500

The GARCH(1,1) model was fitted using maximum likelihood estimation to the S&P 500 monthly log-returns in Figure 2.1 using the Generalized Reduced Gradient Solver tool from Microsoft Excel. The MLE parameters of the model are listed in Table 2.1.

2.3.3 Issues with the GARCH model for long-term equity data

To illustrate the ineffectiveness of the estimated GARCH(1,1) process at capturing the tails of the S&P 500 index, a sample path of monthly log-returns the same length as the S&P 500 was simulated from the model. This sample path is compared to the monthly log-returns of the S&P 500 monthly log-returns in Figure 2.2. The objective is to illustrate that the S&P log-returns show more variability, particularly on the down-side. The GARCH(1,1) model does not generate crashes of the sort experienced in 1987 or 2008. This is not just a feature of this example path, it is a feature of the GARCH(1,1) model generally.

In October of 1987, the S&P 500 experienced a monthly fall of 21.76%. The probability of experiencing a crash at least as big as this under the GARCH(1,1) model fitted to the S&P
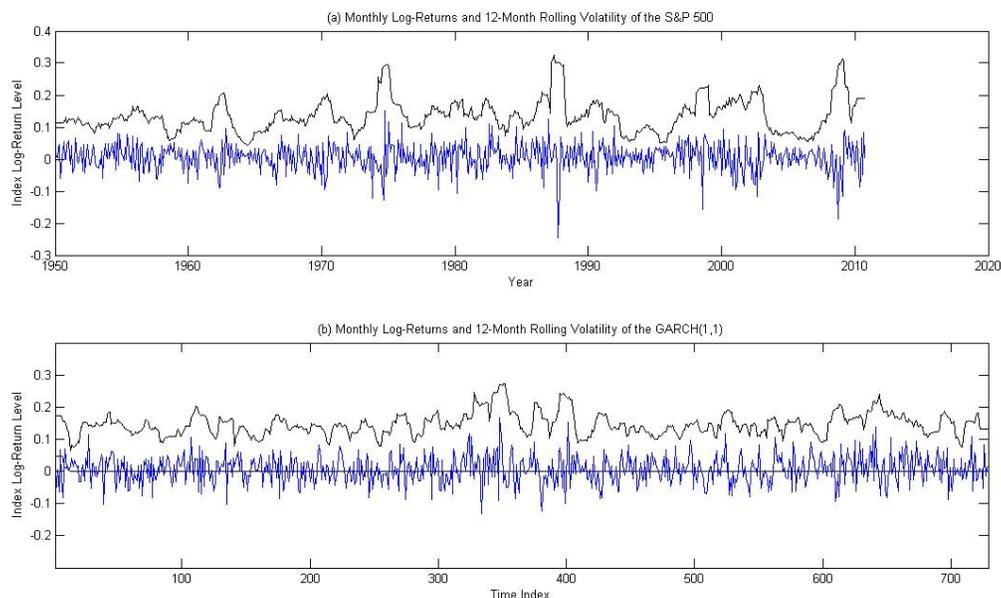
Figure 2.2: S&P 500 Monthly Log-Return Levels and a Sample Path of Log-Return Levels from the GARCH(1,1)

data is approximately 0.00001, which would mean one would see such a one month drop every 8,000 years. More pertinent to long-term performance, the S&P 500, from January 1999 to January 2009, experienced a ten year drop of 37.07%. The probability of experiencing at least as large a drop as this under the fitted GARCH(1,1) model is 0.002, which would mean one would see a ten year drop of such a magnitude around once every 5,000 years.

Several other GARCH(p, q) were explored, with values of p and q up to 3. The resulting series suffered the same issues displayed in Figure 2.2. The extra parameters, although sometimes significant, would simply replace existing parameters in terms of their effect. There is limited flexibility in the contribution of the additional GARCH parameters. For a more detailed comparison of GARCH models to other candidate models for the S&P 500, see Hardy, Freeland & Till (2006).

With the focus of this thesis on the tails of both the short and long-term distributions, the

need to move beyond GARCH modeling for the S&P 500 has been shown. Models need to capture the larger volatility movements of the process. This development led to the more sophisticated hidden Markov class of models detailed in Section 2.4.

2.4   Hidden Markov Models

Introduced in a series of papers by Baum in the 1960's, with initial applications used in the field of speech recognition, the class of models known as Hidden Markov, or regime-switching models also has the capability of capturing changing volatility and volatility clustering. The models are built around an underlying state process which is treated as unobserved. That means that in this framework we have two dependent stochastic processes; the original log-return process, and the unobserved, underlying process which we assume to be Markov. We refer to this underlying process as the state process.

Under these models, the original process is assumed to be in any of K specified states at each time point; at the next point, the process either stays in the same state or transitions to another. For each of the K states a process is defined. The distribution for the original process at each time point is dependent on which state the underlying state process lies in at that time.

Hamilton (1989) provided some intuition for the use of hidden Markov models for economic series. In the case of stock prices, we might conjecture that the market, or perhaps more correctly the traders, experience periods of relative calm or optimism, where normal everyday trading will take place. However, after some random trigger event, for example the collapse of a key financial institution, the market will shift to a mood of pessimism, leading to poor returns and higher volatility.

We denote the underlying state stochastic process by $\rho_t \in \{1, \ldots, K\}, t = 1, \ldots, N$, indicating the underlying state of the market for each time t. Since this state of the market is unobserved, these parameters are not estimated directly using typical maximum likelihood estimation.

As the name suggests, discrete hidden Markov models define the underlying state process as a discrete time Markov chain. Homogeneous Markov models have the property that the distribution of the underlying state at time $t + 1$ depends only on the state at time t, and moreover this relationship does not depend on t, which is a very strong modeling assumption.

The underlying state process is described by a set of transition probabilities. Define $p_{i,j}$ as the probability, conditional on the process being in state i at time t, that the state at time $t + 1$ will be j, independent of the value of t. More formally,

$$p_{i,j} = P[\rho_{t+1} = j | \rho_t = i], \qquad i, j \in \{1, \ldots, K\}, \qquad t \in \{1, 2, \ldots\}$$

These transition probabilities are parameters of the hidden Markov model. A model with K states will have $K \cdot (K - 1)$ transition probability parameters, meaning the number of parameters becomes very large very quickly as more states are added to the model. The transition probability parameters are typically combined into a transition probability matrix. Define the matrix P as

$$P = [p_{i,j}]_{K \times K}$$

Since the underlying regimes are not estimated as point parameters, but instead treated as random variables, it is often necessary to determine the probability that the process is in each state for each time point t. Without any information from the data, the underlying states all have the same distribution known as the stationary distribution.

Let $\pi_i, i \in \{1, \ldots, K\}$ represent the stationary probabilities that the process is in state i respectively (these also can be interpreted as the long-run proportion the series spends in state i). The π_i's are then the solution to

$$\pi P = \pi, \qquad \text{where} \qquad \pi = \begin{pmatrix} \pi_1 & \pi_2 & \ldots & \pi_K \end{pmatrix}$$

The distribution of the log-returns is dependent on the underlying state of the process. As this thesis deals with models with Gaussian innovations, the distribution of the log-returns

17

at each time t under a regime switching model can be written as

$$Y_t|\rho_t \sim N(\mu_{\rho_t,t}, \sigma^2_{\rho_t,t}), \qquad Y_i|\rho_i, Y_j|\rho_j \text{ independent for } i \neq j$$

The parameters $\mu_{\rho_t,t}$ and $\sigma^2_{\rho_t,t}$ can be constant parameters, or may be stochastic processes themselves. In following sections we present some of the popular choices.

Given the data observation set $\vec{y} = \{y_1, \ldots, y_t\}$, inference about the underlying state process at each time point t can be made. We denote time 0 as the state immediately before observation. We assume the probability function for ρ_0 follows the stationary distribution, π, and since it is only dependent on ρ_0, ρ_1 also follows the stationary distribution. At the next time point, we may condition on the observation, y_1. Let $p_i(1), i \in \{1, \ldots, K\}$ denote the probability that the underlying state at time 1 is state i, conditional on the log-return y_1, that is:

$$p_i(1) = P[\rho_1 = i|y_1], \qquad \forall i$$

For each $i \in \{1, ..., K\}$,

$$p_i(1) \propto \pi_i \cdot f(y_1|\rho_1 = i)$$

where $f(y|\rho)$ represents the probability density function of log-return y under the distribution associated with state ρ.

For each subsequent time point t, one can obtain a distribution of the underlying state at t conditional on all data observations up to and including the observation at time t. Let $p_i(t), i \in \{1, \ldots, K\}, t \in \{1, \ldots, N\}$ be the probability that the underlying state at time t is state i, conditional on the data observations $y_1, \ldots, y_t$. These probabilities are easily obtained recursively, using the relations

$$p_i(t) \propto \sum_{j=1}^{K} p_j(t-1) \cdot p_{j,i} \cdot f(y_t|\rho_t = i)$$

Conditioning on all data observations up to and including y_t instead of just y_t alone, adds information regarding where the underlying state path has likely traveled. The probabilities $p_i(t)$ are central to much of the decision making and analysis performed in the rest of this thesis.

The long-term equity models described below all maintain the hidden Markov structure described in this section. The difference between the models comes from assuming different distributions within the states (or regimes) for the original log-return process.

2.4.1 The Regime-Switching Log-Normal Model

Proposed for long-term equity modeling by Hardy (2001), the regime-switching log-Normal model (RSLN) uses Normal distributions for each state under the regime-switching framework. Mean and variance parameters are defined for each regime, meaning an RSLN model with K regimes, denoted RSLN-K, would have $K \cdot (K-1) + 2K = K \cdot (K+1)$ parameters: $\mu_1, \ldots, \mu_K$, $\sigma_1, \ldots, \sigma_K$ and the $K \cdot (K-1)$ transition probability parameters.

The model can be expressed as

$$Y_t | \rho_t = \mu_{\rho_t} + \sigma_{\rho_t} \cdot \epsilon_t, \qquad Y_i | \rho_i, Y_j | \rho_j \text{ independent for } i \neq j$$

where $\rho_t \in \{1, \ldots, K\}$, and conditional on the value of ρ_{t-1}, takes value k with probability $p_{\rho_{t-1}, k}$.

One of the attractive features of the RSLN model for long-term equity prices is its ability to capture volatility clustering through the regime-switching framework. Different variance parameters apply in each of the regimes, which allows the modeling of log-returns with K specified potential values for the monthly (say) volatility. If the persistence of each regime is relatively high, then the process may spend extended periods in one regime with a relatively low level of volatility, then transition to a different regime and spend extended periods with a relatively higher level of volatility.

In investment guarantee literature (CIA, 2001, Hardy, 2001 and Hardy, Freeland & Till,

Regime	μ	σ	Transition Parameters
One	0.01024	0.03384	$p_{1,2} = 0.0337$
Two	-0.01448	0.06486	$p_{2,1} = 0.1517$

Table 2.2: The RSLN-2 MLE Parameters Fitted to the S&P 500 Monthly Log-Return from Jan 1950 to Oct 2010

2006, for example), typically two regimes are used. Using maximum likelihood estimation for example, fitting stock index data generates one regime with a positive expected return and a relatively low variance, and a second regime with a negative expected return and a relatively high variance. Both regimes have a high persistency, with the probability of remaining in the high return / low volatility regime around 95%, and the probability of remaining in the low return / high volatility regime around 80%. This typical framework is consistent with the intuitive justification for financial regime-switching models described at the beginning of the section.

When adding a third regime to the framework, the third regime often acts as a transition regime between the two existing regimes, with a very low persistency. Entering this regime will usually only occur from one of the two original regimes, and the process exits to the other. The parameters of the original regimes will change slightly, compared with the 2-regime parameters but they maintain the general form, with a high mean low volatility regime and a low mean high volatility regime.

RSLN and the S&P 500

Both the RSLN-2 and RSLN-3 models were fitted using maximum likelihood estimation to the S&P 500 monthly log-returns from January 1950 to October 2010 using the same Generalized Reduced Gradient Solver algorithm from Microsoft Excel. The parameters for the RSLN-2 model are listed in Table 2.2, and the parameters for the RSLN-3 model are listed in Table 2.3.

The third regime acts as a transition regime between the low return regime and the high return regime. The expected time the process stays in this regime once it enters it is less than two months. The third regime appears to capture a short positive period of recovery

Regime	μ	σ	Transition Parameters
One	0.00876	0.03471	$p_{1,2} = 0.0234$, $p_{1,3} = 0.0000$
Two	-0.03598	0.06601	$p_{2,1} = 0.0000$, $p_{2,3} = 0.1956$
Three	0.05944	0.01945	$p_{3,1} = 0.6159$, $p_{3,2} = 0.0000$

Table 2.3: The RSLN-3 MLE Parameters Fitted to the S&P 500 Monthly Log-Return from Jan 1950 to Oct 2010

before the system returns to the more persistent long term regime. Adding the third regime has a significant effect on the regime two mean in this case; as some of the returns that are positive, coming towards the end of a period of uncertainty, are re-allocated to the third, intermediate regime, the remaining regime 2 returns are more negative, on average. The changes to the low return regime in the RSLN-3 model mean that a larger negative return when entering a period of high volatility, captured by regime 2, is that more likely, and larger positive returns are more likely coming out of the period of high volatility, now captured by regime 3.

However, with such little time spent in the third regime, questions about its statistical significance arise. Under the RSLN-3 model, for a series the same length as the S&P 500, the process only expects to spend about 24 months in the third regime, for an extra six model parameters (although three of them ended up being estimated as zero).

2.4.2   The Regime-Switching Draw-Down Model

The regime-switching draw-down model (RSDD), proposed by Panneton (2004), builds on the RSLN model by adding a form of mean reversion to the distributions within the regimes. The purpose of the extra framework is to encourage higher monthly returns after experiencing low returns. The model achieves this by first defining a stochastic process D_t, which can be described as the draw-down level, that tracks how far the total log-returns have fallen below the processes previous high. More formally,

$$D_t = \min(0, D_{t-1} + Y_t)$$

The model incorporates the draw-down level into the expected log-returns by applying a factor $\varphi_k, k \in \{1, \ldots, K\}$, which is specific to the processes underlying regime. The φ_k's are treated as model parameters.

The process can be expressed as

$$
\begin{aligned}
Y_t | \rho_t &= \mu_{\rho_t} + \varphi_{\rho_t} \cdot D_{t-1} + \sigma_{\rho_t} \cdot \epsilon_t \qquad \text{where } \epsilon_t \sim N(0,1) \\
\rho_t | \rho_{t-1} &= k \quad \text{w.p.} \quad p_{\rho_{t-1},k} \qquad k \in \{1, \ldots, K\}
\end{aligned}
$$

With an extra parameter per regime, relative to the RSLN model, the RSDD model has $K \cdot (K+2)$ parameters.

The φ_k parameters are generally negative, which means the regime-specific means $\mu_{\rho_t} + \varphi_{\rho_t} \cdot D_{t-1}$ will always be greater than or equal to the base mean μ_{ρ_t}. More specifically, the larger the process falls below its previous high, the larger the expected return of the next observation will be. The intuition for this framework from a financial standpoint is that markets will 'bounce back' after experiencing losses. While this addition can appear to be only a slight change relative to the RSLN model for the monthly observations, this type of framework has drastic implications for the long-term returns of the process. Strings of negative returns are less likely under the model, resulting in a significantly thinner long-term return left tails. Incidentally, this 'bounce-back' appears to be the feature captured in the third regime in Table 2.3.

The states for the an RSDD-2 model typically break down the same way as the RSLN-2: one regime has a positive return with a low variance, and one regime with a negative return and a high variance. Typically, however, when fitting the RSDD-2 and RSLN-2 to the same data set, the expected return for the second regime under the RSDD-2 model will be lower than the respective expected return for the RSLN-2 model to counterbalance the draw-down effect.

Regime	μ	φ	σ	Transition Parameters
One	0.00648	-0.04687	0.03435	$p_{1,2} = 0.0282$
Two	-0.04051	-0.05580	0.06446	$p_{2,1} = 0.1809$

Table 2.4: The RSDD-2 MLE Parameters Fitted to the S&P 500 Monthly Log-Return from Jan 1950 to Oct 2010

RSDD-2 and the S&P 500

The RSDD-2 was fitted by maximum likelihood estimation to the S&P 500 from January 1950 to October 2010, and the parameters are shown in Table 2.4. The two states are similar to the states from RSLN-2 model in function, but the state-two mean is almost three times that of the RSLN-2 model, counterbalancing the draw-down functionality of the model. The RSDD-2 model also spends more relative time in the low volatility state than does the RSLN-2 model. Interestingly, the φ parameters from the two states are quite close together, suggesting that the incorporated mean-reversion is irrespective of the state of the market.

2.4.3   The Regime-Switching GARCH model

As described by Gray (1996), the GARCH architecture can be defined within a regime-switching framework for a regime-switching GARCH (RSGARCH) model. This model is significantly more complicated than either the GARCH or the RSLN models. An extra complication arises from the fact that current data observations under the GARCH framework are dependent on previous volatilities, yet those volatilities under a regime switching setting are still random variables after conditioning on the data, due to the unknown nature of the underlying regimes.

Gray solved this issue through averaging the volatilities across the regimes. First, the regime-specific distributions are defined as for any regime-switching model, except now with a GARCH process within regimes:

$$Y_t|\rho_t = \mu_{\rho_t} + \sigma_{\rho_t,t}\epsilon_t$$
$$\sigma_{\rho_t,t}^2 = \alpha_{\rho_t,0} + \alpha_{\rho_t,1}e_{t-1}^2 + \beta_{\rho_t}\sigma_{t-1}^2$$

Regime	μ	α_0	α_1	β_1	Transition Parameters
One	0.00790	0.00000	0.00000	0.85279	$p_{1,2} = 0.0433$
Two	-0.06459	0.00335	0.00000	0.57147	$p_{2,1} = 1.0000$

Table 2.5: The Two-Regime RSGARCH MLE Parameters Fitted to the S&P 500 Monthly Log-Return from Jan 1956 to Oct 2010

Using the probabilities $p_i(t)$ as weights, where $p_i(t)$, as defined above is the probability that at time t the underlying state is i, conditional on the data observations up to and including the observation at time t, e_t^2 and σ_t^2 can be expressed as

$$
\begin{aligned}
e_t &= y_t - \{p_1(t)\mu_1 + (1-p_1(t))\mu_2\} \\
\sigma_t^2 &= p_1(t)(\mu_1^2 + \sigma_{1,t}^2) + (1-p_1(t))(\mu_2^2 + \sigma_{2,t}^2) - \{p_1(t)\mu_1 + (1-p_1(t))\mu_2\}^2
\end{aligned}
$$

In exchange for the added complexity, the RSGARCH model allows a wide range of possible values for the volatility process, as does the GARCH model, but, additionally, has the ability to incorporate the association of periods of increased volatility with poorer market returns through the regime-switching framework. However, the number of parameters for the RSGARCH grows more quickly than simpler regime-switching models and can become very difficult to fit models with as few as three regimes. Moreover, the number of parameter permutations available within the framework is also quite large, which can make model selection with the RSGARCH framework itself quite cumbersome.

RSGARCH and the S&P 500

For the S&P 500 from Jan 1950 to Oct 2010, a RSGARCH-2 model with a GARCH(1,1) framework within both of the regimes was fitted using maximum likelihood. This model has ten parameters: $\mu_1, \mu_2, \alpha_{1,0}, \alpha_{2,0}, \alpha_{1,1}, \alpha_{2,1}, \beta_{1,1}, \beta_{2,1}$ and the transition probabilities $p_{1,2}$ and $p_{2,1}$. Even this simple RSGARCH model was quite difficult to estimate using maximum likelihood estimation, with many local maxima close to the global maximum. The MLE parameters are listed in Table 2.5.

When coupled with a GARCH(1,1) framework within regimes, the regime pattern is quite different compared to the RSLN-2 model. The transition probability from regime two to regime one is 1, meaning that after transition to the second regime, the process immediately transitions back to the first. This process more resembles the regular GARCH(1,1) model, except once in a while the process changes to produce an extreme left tail observation (the mean of the log-returns in the second regime is -0.06459), and then immediately reverts to the GARCH(1,1) framework again. This might suggest that the regime-switching framework is perhaps not an ideal extension of the GARCH model, at least under maximum likelihood estimation, for this data set.

The MARCH model and the S&P 500

Wong & Chan (2005) presented a specific version of the RSGARCH model, which was known as the Mixture-ARCH (MARCH) process. The model had two major assumptions that differentiate it from the full RSGARCH model. The first was that all β parameters under the GARCH model were fixed to be zero, meaning the processes within regimes are ARCH not GARCH. The second assumption was that the regime-switching process was constrained to be a mixture process, meaning that the transition probabilities are now independent of the current state.

Formally, the mixture assumption requires

$$p_{1,k} = p_{2,k} = \cdots = p_{K,k}, \qquad \forall k$$

The mixture assumption means that volatility clustering can no longer be captured through the underlying state process, as the processes state at time t is now independent of the processes state at any other time. Now, only can volatility clustering be captured through the state-specific ARCH framework.

Hardy, Freeland & Till (2006) fitted the MARCH model to the S&P 500, and found a two-point mixture with one mixture component having an ARCH process with two or three lags and the other mixture component having no lags resulted in a good fit under the model's framework. This model was again fitted to the S&P 500 data from Jan 1950 to Oct 2010,

State	μ	α_0	α_1	α_2	α_3	State Probability
One	0.00953	0.00000	0.07143	0.05702	0.16358	0.9667
Two	-0.06593	0.00492	–	–	–	0.0333

Table 2.6: The Two-Point MARCH MLE Parameters Fitted to the S&P 500 Monthly Log-Return from Jan 1950 to Oct 2010

and the model parameters are listed in Table 2.6.

Under the ML estimated model, the probability of the underlying state of the process at any point in time being state 1 is 96.7%. As was the case for the RSGARCH model, this probability is considerably higher than the state one frequency for the RSLN and RSDD models. In a broader sense, however, the model parameters are quite similar to the other hidden Markov models: an often visited state having a positive expected, and a less often visited state with a negative expected return. The positive correlation seen in the volatility of the data is captured here through the three α parameters from state 1.

2.4.4 The Hidden Markov Models and the Tails of the S&P 500

The fitted single-state GARCH family of models were unsatisfactory in that the models did not satisfactorily capture either realistic volatility clustering, or the extreme values seen in the data. A first look into the adequacy of the class of hidden Markov models when modeled to the S&P 500 should look at these elements.

The S&P 500 index, and its observed volatility pattern is displayed in Figure 2.3, along with sample paths from each of the hidden Markov models fitted and their respective observed volatility patterns, similar to Figure 2.2 for the GARCH(1,1) model. The volatility patterns of the hidden Markov models resemble much more closely that of the S&P 500, as the larger jumps in observed volatility are seen for each of the models.

Also important the ability of the hidden Markov models to capture the extreme observations seen in the data. The probability of each of the hidden Markov models experiencing the market crash of October 1987 are displayed in Table 2.7, along with the probability of experiencing the crash under the single state GARCH model. The probability of experiencing
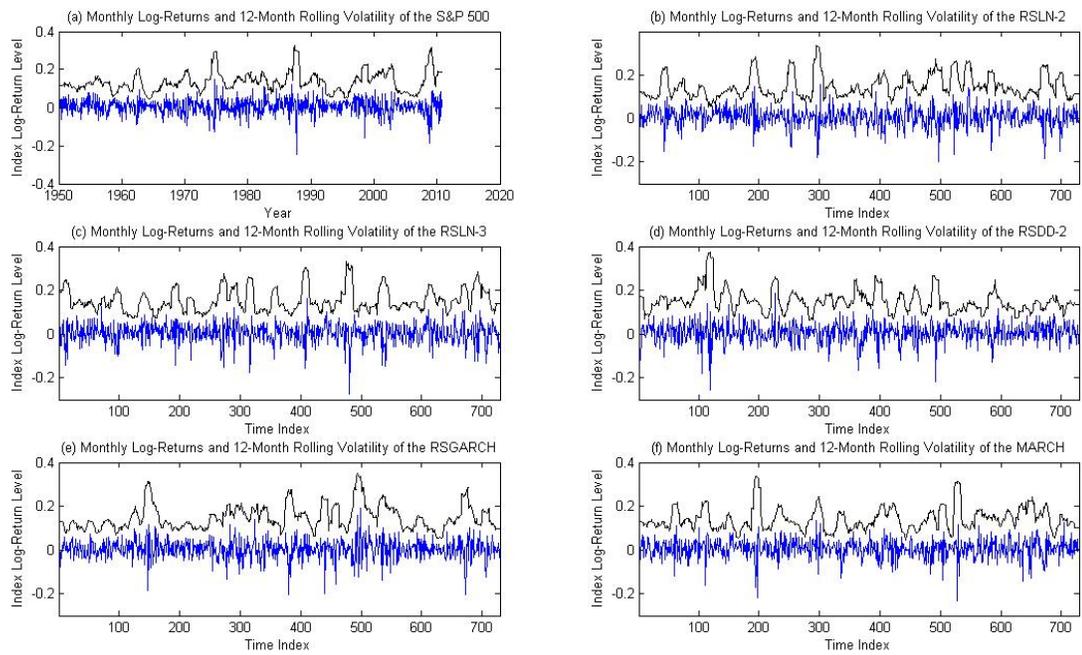
Figure 2.3: S&P 500 Monthly Log-Return Levels and Sample Paths of Log-Return Levels from the Hidden Markov Models

Model	Probability
RSLN-2	0.0043 %
RSLN-3	0.0101 %
RSDD-2	0.0089 %
RSGARCH	0.0090 %
MARCH	0.0185 %
GARCH(1,1)	<0.0001 %

Table 2.7: Probability of experiencing the S&P 500 October 1987 Crash under the Hidden Markov Models

the crash is over 100 times greater under the more complex hidden Markov models.

The Long-term Tail of the S&P 500

Of more importance for long-term equity modeling than the single log-returns are the longer-term implications of the candidate models. As noted earlier, over the 10-year period from January 1999 to January 2009, the S&P 500 experienced a drop of 37.07%. Since a drop of this magnitude would push a typical investment guarantee on the S&P 500 very deep into-the-money, stronger candidate models used for such a contract would better capture it such a drop.

The probabilities that a drop at least the size of the record drop in the S&P from January 1999 to January 2009 would occur in a 10-year period for each of the candidate hidden Markov models are listed in Table 2.8. The RSLN-2 and RSLN-3 models capture the period the best; under both models the probability of experiencing the 10-year crash is above 2.75%. The drop occurs with about 1.8% probability under the RS-GARCH model, 0.4% probability under the MARCH model, and with 0.2% probability under the RSDD-2 model. The probability of such a drop under the single state GARCH model was 0.2% as well, meaning the RSLN models especially offer a drastic improvement for capturing this period.

Model	Probability
RSLN-2	0.0276
RSLN-3	0.0302
RSDD-2	0.0016
RSGARCH	0.0182
MARCH	0.0042

Table 2.8: Probability of experiencing the 1999-2009 S&P Return under the fitted Hidden Markov Models

2.5 Long-term Equity Hidden Markov Model Selection

In this section, the recent history in hidden Markov model selection for long-term equity data will be discussed. First an illustration of differences of the models discussed and fitted in the previous section will be presented, providing the justification for a thorough model decision process. A summary of the model selection discussion in Hardy, Freeland & Till (2006) will follow. Finally, the need for hidden Markov goodness of fit tests will discussed, paving the way for the hidden Markov residual tests that are introduced in the next chapter.

2.5.1 The 10-year Outlook for Long-term equity Hidden Markov Models

The hidden Markov models discussed in the previous section all share the same framework: they all have one underlying state representative of a normal stable market with positive expected returns and relatively low volatility, and another underlying state representative of market instability, where observed volatilities are high, and expected returns are negative. The differences in the models are mostly in the intra-state distributions, such as the mean-reversion parameters of the RSDD model, or the additional autoregressive parameters in the RSGARCH and MARCH models. However, at least under maximum likelihood estimation, the differences in the model specifications can force stark differences in the parameters of the models, the most notable ones being the mean return within the high volatility state, and the transition probabilities.

These differences can have significant effects on the long-term implications of the models,
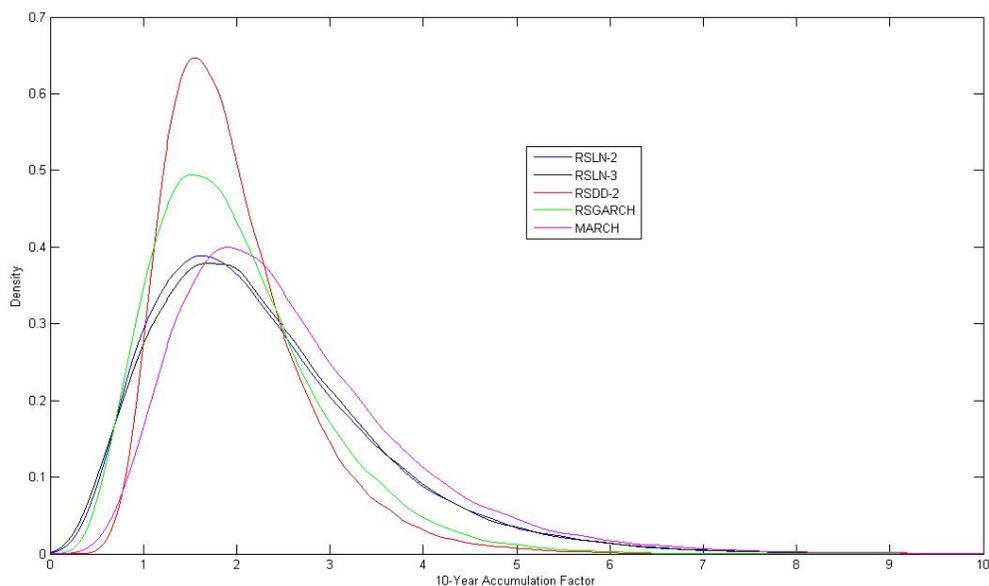
Figure 2.4: Density Functions of the 10-Year Accumulation Factors of the Popular Hidden Markov Models

and consequently can have significant effects on the valuation of derivative contracts of the equities being modeled. Figure 2.4 displays the probability density functions of the 10-year total returns under each of the hidden Markov models fitted. The returns represent the accumulation of a \$1 investment in the index for a 10-year period. The distributions of the accumulation factors, in particular in the left and right tails, are quite different under the different models. The RSLN-3 and RSLN-2 models have quite comparable left tails, the thickest of the candidate models. The RSGARCH left tail is slightly thinner, while the MARCH model left tail is noticeably thinner again, and the RSDD-2's left tail significantly thinner still. The MARCH, RSLN-3 and RSLN-2 models have relatively thicker right tails, while the RSDD-2 has the thinnest right and left tails.

The tails of the distributions are quite central to the valuation of equity linked investment guarantees. Following is a simple example of an investment guarantee contract, to show how the reserve level can be quite different under the different models. The effects of the differences in the tails of the models displayed in Figure 2.4 will be illustrated. The objective of the example is to show that the choice of model for investment guarantee analysis is a

non-trivial decision, and that a rigorous validation and selection process for the models is justified.

2.5.2   Guaranteed Minimum Accumulation Benefit Example

The example contract will be a 10-year Guaranteed Minimum Accumulation Benefit (GMAB) contract (see Hardy, 2003 for details). A single premium is assumed to be paid by the policyholder. the premium which is placed in a fund which is assumed to be invested in the S&P 500 with dividends reinvested. The policy benefit, paid either at the end of the month of death or at contract maturity, is the greater of the accumulated investor's fund value and a guarantee on the investor's fund. The guarantee is initially set to the single premium. If the policy is still in force at five years, then there is a one-time ratchet function which sets the guarantee to the accumulated fund value at five years, if the fund is higher than the initial guarantee level.

Formally, let F_t represent the policyholder's fund value at time t, in months, before the monthly expense charge has been deducted from the fund. Let G_t be the guarantee level at time t. Then G_t is defined as

$$
G_t \;=\; \begin{cases} F_0 & 0 \le t \le 5 \\ max(F_0, F_5) & 5 < t \le 10 \end{cases}
$$

To fund the guarantee, an expense charge at the rate of 3% per year is deducted from the fund value at the beginning of every month. This expense charge is treated independently of the estimated cost of the guarantee under the assumption that supply and demand market forces of GMAB contracts will determine the expense charge level.

The policyholder is assumed to be aged 50 and has mortality consistent with Appendix A of Hardy(2003). Lapses for the contract are assumed to be 8% per annum (a 'lapse' in this case is assumed to be the policyholder prematurely withdrawing the whole fund; an event under which no guarantee is paid regardless of the fund's level). The interest rate is assumed to

CTE-Level	RSLN-2 %	RSLN-3 %	RSDD-2 %	RSGARCH %	MARCH %
75%	18.05	17.90	12.48	16.73	9.89
90%	34.67	35.28	22.85	30.53	21.96
95%	44.29	45.75	28.79	38.52	30.09

Table 2.9: CTE's of the GMAB Present Value Liability (as a percentage of initial premium) under different Regime-Switching Models, for different levels of α

be a fixed 5% per year.

For this simple example, the guarantee will assumed to be reserved using the 'actuarial approach'. This approach uses Monte Carlo simulation to project the liabilities of the guarantee under the model (the 'real world' distribution as opposed to the risk neutral distribution). The reserve is then set to an appropriate risk measure of the discounted liabilities. The risk measure used here will be the Conditional Tail Expectation (CTE), at the $\alpha \cdot 100\%$ level, which is an average of the worst $100(1 - \alpha)\%$ of the fund performance simulations.

The cash flows for the policy were simulated, 100,000 times, for each of the popular regime-switching models, and reserve levels (as percentages of the initial premium investment F_0) for the $75\%, 90\%$ and 95% CTE of the GMAB liability are listed in Table 2.9. The reserves for the GMAB liability are quite different across the different models. At one end of the spectrum are the RSLN-2, RSLN-3 and RSGARCH models, and at the other end are the RSDD-2 and MARCH models. Over the different levels for the CTE-measure, an insurer would need 45-55% more capital under the RSLN-2 model compared to the RSDD-2 model.

This example demonstrates that model choice can have a very significant impact on economic capital. It is therefore important for analysts to perform a thorough model validation investigation to see which model is most consistent with the data. In the remainder of this chapter and in Chapters 3 and 4 we develop tools to improve model validation and selection for long term equity models, in particular with regime switching.

2.5.3   Long-term Equity Hidden Markov Model Selection

Most of the hidden Markov equity model literature uses likelihood-based model selection criteria after the models have been fitted to data. The aim of these criteria, such as the Akaike Information Criterion (Akaike, 1974) or the Schwarz-Bayes criterion (Schwarz, 1978), is to weigh the value of each the model's parameters in an effort to select a single model for use. Panneton (2002), for instance, based the argument for the RSDD model around its significantly higher log-likelihood, and success in the corresponding tests. We note that Hamilton and Susmel (1994) demonstrated that one should not use the likelihood ratio test to determine the number of states in a hidden Markov model.

Other non-likelihood based testing has also been developed. Hardy, Freeland & Till (2006) made use of a modified resampling technique for model inference. However, the results were somewhere between the two extremes of the RSGARCH and RSLN-2 models and the RSDD-2 model. The conclusion reached was that the RSLN-2 and RSGARCH models were the safer choices, but limited scientific conclusions were reached through the model validation analysis.

The comparative tests determine if a model fits better than another specified model, taking likelihood and parsimony into consideration. Goodness of fit tests are used to determine whether or not the candidate fits the data acceptably in the first place. This step should be performed prior to any comparative tests, as comparing two models that do not fit the data well, will result in an unsatisfactory model either way. Many goodness of fit tests apply the test criterion to the difference between the observed data and the expected values under the model. It is common, for example, to work with standardized residuals. For a single process model, with independent, normally distributed innovation terms, the standardized residuals can be tested to assess whether they significantly different, statistically, from the independent normal sample expected from the model. Similarly, if the residuals, ϵ_t in the models described earlier in this Chapter, are assumed to have a Student's t-distribution, the goodness of fit of the model could be assessed by comparing the standardized residuals with the Student's t-distribution. The idea is to compare the estimated residuals with the random innovations assumed in the model; the null hypothesis would be that the residuals are consistent with the model. For regime-switching models it is not quite so straightforward

to determine what the the distribution of the standardized residuals should be, under the null hypothesis. Although the ϵ_t are assumed $\mathcal{N}(0,1)$, the standardized residuals are not necessarily $\mathcal{N}(0,1)$, under the null hypothesis, because of the impact of the regime process.

Hardy, Freeland & Till (2006) did a preliminary, more heuristic analysis of the residuals from each of the hidden Markov models above along with some other non-hidden Markov models, for the S&P 500 and found that only the regime-switching models passed standard residual tests such as the Jarque-Bera test and a QQ-plot inspection. In Chapter 3 we review the heuristic approach of Hardy et al (2006) and develop a more rigorous process for goodness of fit testing of regime switching models.

2.6   Conclusion

In this chapter, the S&P 500 total return index was identified as the data-set of interest, and it was demonstrated that traditional GARCH modeling was inadequate for typical long-term investment guarantee analysis for this index. Many of the popular hidden Markov models used for long-term equity data were then described and fit to the S&P 500 using maximum likelihood estimation. The hidden Markov models showed significant improvement in capturing the long-term tails of the S&P 500 over traditional GARCH modeling.

The GMAB example demonstrated that model choice, even among different types of hidden Markov models, can have a significant impact on long-term investment guarantee analysis, justifying the need for thorough model validation and selection analysis. The current literature for investment guarantee model validation maintains a lot of uncertainty in decision, which paves the way for a more thorough validation analysis developed here in Chapter 3.

Chapter 3

Hidden Markov Residuals

3.1 Introduction

In Chapter 2, different hidden Markov models used for modeling long-term financial data were presented and the long-term implications of each, when fitted to the S&P 500 were discussed. There were very significant differences found between the models with regard to the left tails of the longer term accumulation factors. These differences may have significant impact on the risk management of investment guarantees or other financial derivatives. The need for effective validation of these models was emphasized.

One branch of statistical tests is that of model comparison tests. Examples of such tests include the Akaike Information Criterion (AIC) test (Akaike, 1974), and the Schwarz-Bayes Criterion (SBC) test (Schwarz, 1978). These tests use the maximum log-likelihood of the data under the model, with a penalty for the number if parameters, to rank candidate models for the purpose of model selection.

The popular model comparison tests were developed for single state models. Their application to hidden Markov models is questionable in some cases: for instance, under its standard assumptions the LRT is not valid for selection of the number of underlying states in a hidden Markov for a particular data set (Hamilton, 1994), and in theory should only be used

to compare embedded models. The LRT is widely used as a heuristic assessment outside these theoretical limitations, including in Hardy, Freeland & Till, (2006), where some of the models described in Chapter 2 were compared based on the LRT criterion, *inter alia.*

Another important branch of model tests are goodness-of-fit tests. These type of tests measure whether or not a data set is consistent with a specified model. These tests are often based on the standardized residuals of the data, because it is common that the distribution of the model standardized residuals is well specified. That is, suppose we have a discrete time stochastic process $\{Y_t\}$, $t = 1, 2, ..., T$, and a data set $(y_1, y_2, ..., y_T)$. The model standardized residuals are the random variables

$$\epsilon_t = \frac{Y_t - E[Y_t]}{\sqrt{V[Y_t]}},$$

and the standardized residuals of the data are estimated as

$$\hat{\epsilon}_t = \frac{y_t - \hat{\mu}_t}{\hat{\sigma}_t}$$

where $\hat{\mu}_t$ is the estimated value of $E[Y_t]$ and $\hat{\sigma}_t^2$ is the estimated value of $V[Y_t]$.

For a univariate series, the model residuals are generally independent (even where the process has serial dependence), and follow a standard distribution such as $\mathcal{N}(0,1)$ or the Student's t-distribution. See, for example, Anscombe (1961), or Morgan (1954), or Campbell, Lo & McKinlay (1996).

Residual-based testing has not been extensively developed for hidden Markov models. Although the process is defined to have independent $\mathcal{N}(0,1)$ residuals within each state, these are not recovered using standardized residuals, assuming the state is unknown. A particular problem is that the standardized residuals under a regime switching process are not independent, as they depend, serially, on the underlying regime.

In this chapter, the difficulties of defining residuals for hidden-Markov models are first illustrated. We then demonstrate how, given a hidden Markov model with $\mathcal{N}(0,1)$ residuals within the underlying states, we can generate model residuals that are independent, $\mathcal{N}(0,1)$ distributed, even with a hidden Markov regime process. We can use this information to gen-

erate the distribution of the residuals from the data. The resulting random variables are also independent $\mathcal{N}(0,1)$ under the null hypothesis that the data is consistent with the model, and this allows us to assess the goodness of fit of the data with the model using standard goodness of fit tests for Normal distributions.

3.2 Hidden Markov Residuals

In this section, we use R_t to denote model residuals, which are functions of the random variables Y_t and ρ_t, the state process. We use r_t to denote the data residuals, which are functions of the data and of the estimated model parameters.

For a hidden Markov process with $\mathcal{N}(0,1)$ innovations within regimes, we recover a $\mathcal{N}(0,1)$, independent residual process as

$$R_t^m | \rho_t = \frac{Y_t - E[Y_t|\rho_t]}{\sqrt{V[Y_t|\rho_t]}}$$

where ρ_t is the underlying state at time t as in Chapter 2. However, as the regime process is unobserved, we cannot calculate the equivalent data residuals.

In practice, although this series of residuals is not observable because the underlying state process is also unobserved, there is an observable set of residuals for each of the underlying states. These sets of residuals are denoted $r_{t,k}$, where $k \in \{1, \ldots, K\}$:

$$r_{t,k} = \frac{y_t - \hat{\mu}_{t,k}}{\hat{\sigma}_{t,k}}, \quad t \in \{1, \ldots, N\}, \quad k \in \{1, \ldots, K\}$$

where $\hat{\mu}_{t,k}$ and $\hat{\sigma}_{t,k}$ are the estimates of μ_t and σ_t for regime k.

If the data is consistent with the model, then exactly one of the k residuals calculated at each time point is the 'true' $\mathcal{N}(0,1)$ residual. The problem is, we don't know which one. Our objective is to retrieve a residual set with values which are independent and follow a known distribution (preferably $\mathcal{N}(0,1)$ as that is distribution of the model intra-state residuals), and which does not require knowledge of the regime process ρ_t. If we then apply the same process

to data, we can apply goodness of fit tests, knowing that the data should be consistent with the model residual distribution for an adequate fit.

The challenge then is to select a single path of residuals from the state-specific residuals, which retains the characteristics of the $R_t^m|\rho_t$ process without conditioning. In the remainder of this section different options used previously in the literature will be presented, and then it will be demonstrated that some choices lead to an observed residual distribution which is be quite different from the corresponding intra-state residual process, $R_t^m|\rho_t$, that the residuals are intended to recover.

We then present a new, stochastic method for calculating residuals which does recover the intra-state residual process.

As was the case in Chapter 2, we consider models with random innovations within regimes that are independent $\mathcal{N}(0,1)$ distributed. It would be straightforward to generalize to other innovation distributions.

3.2.1 Unconditional Residuals

One approach to obtaining a single residual time series is to first calculate the unconditional mean and standard deviation of the data observation, given what can be inferred about the underlying state path. This set of residuals will be termed the 'unconditional residuals' for this chapter.

Let μ_t^{UC} and σ_t^{UC} be the unconditional mean and standard deviation respectively of the data observation at time t given the information set up to and including time t. μ_t^{UC} and σ_t^{UC} can be calculated through the conditional expectation and variance formulae as:

$$
\mu_t^{UC} = \sum_{k=1}^{K} E[Y_t|\rho_t = k] \cdot p_k(t)
$$

$$
(\sigma_t^{UC})^2 = \sum_{k=1}^{K} Var[Y_t|\rho_t = k] \cdot p_k(t) + \sum_{k=1}^{K} E[Y_t|\rho_t = k]^2 \cdot p_k(t) - \left(\sum_{k=1}^{K} E[Y_t|\rho_t = k] \cdot p_k(t) \right)^2
$$

where $p_k(t)$ is the probability under the model that data observation t has underlying state k, given the data observations up to and including time t, as described in Chapter 2. The parameters μ_t^{UC} and σ_t^{UC} are the unconditional values of the mean and standard deviation of Y_t, given the data up to $t-1$ – that is

$$\mu_t^{UC} = E_\rho \left[E[Y_t|\rho_t] | \{y_1, ... y_{t-1}\} \right]$$

$$\sigma_t^{UC} = E_\rho \left[V[Y_t|\rho_t] | \{y_1, ... y_{t-1}\} \right] + V_\rho \left[E[Y_t|\rho_t] | \{y_1, ... y_{t-1}\} \right]$$

Given the UC parameters, we consider the residual process

$$R_t^{UC}, = \frac{Y_t - \mu_t^{\hat{U}C}}{\sigma_t^{\hat{U}C}}, \quad t \in \{1, \ldots, N\}, \quad k \in \{1, \ldots, K\}$$

If Y_t follows a hidden Markov process with Normal innovations within regimes, the set R_t^{UC} will not be $\mathcal{N}(0,1)$ distributed. The residual values are likely to be far from the original set, $R_t^m|\rho_t$, which we are seeking to recover.

To illustrate this point, consider a hidden Markov model with two states. Assume that the distributions of observations within the states are $\mathcal{N}(-4,1)$ and $\mathcal{N}(4,1)$ respectively. Now suppose $y_t = 0$. The individual regime residual values are

$$r_{t,1} = +4 \qquad r_{t,2} = -4$$

Suppose, further, that $p_1(t) = p_2(t) = 0.5$. Then $\mu_t^{UC} = 0$ which means that $r_t^{UC} = 0$. So, while $r^m|\rho_t$ must lie in either the left or right tail of the $\mathcal{N}(0,1)$ distribution (either $+4$ or -4), the unconditional residual r^{UC} lies in the centre.

We can demonstrate that the UC residuals do not recover the $R_t^m|\rho_t$ process with a simulated example, where we start with known $R_t^m|\rho_t$ values, and consider the resulting UC residuals.
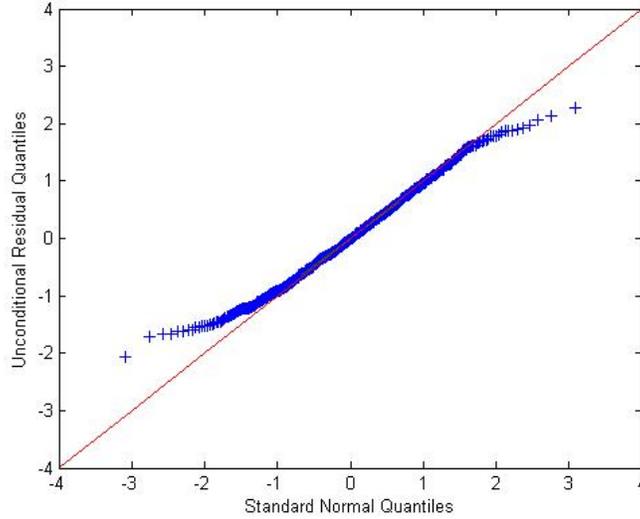
Figure 3.1: QQ-plot of N(0,1) and Unconditional Residual Quantiles for the RSLN-2 Example

We use the RSLN-2 model from Chapter 2, with parameters also from the MLE fit,

$$\mu_1 = \quad 0.01024 \qquad \sigma_1 = 0.03384 \qquad p_{1,2} = 0.0337$$

$$\mu_2 = - \quad 0.01448 \qquad \sigma_2 = 0.06486 \qquad p_{2,1} = 0.1517$$

To demonstrate the distortion of the unconditional residuals, compared with the $R_t^m|\rho_t$ (which are i.i.d. $\mathcal{N}(0,1)$ in this model) we generate a series of 500 values of the log-return from the RSLN process. Instead of random ϵ_t's, we used standard quantiles from the Normal distribution, randomly distributed throughout the time series, without replacement.

$$z_t \in \left(\Phi^{-1}(1/501), \Phi^{-1}(2/501), \ldots, \Phi^{-1}(500/501) \right)$$

By construction, the innovations should appear to be perfectly $\mathcal{N}(0,1)$. However, the residuals $\{r_t^{UC}\}$ are not normally distributed. The QQ-plot of the unconditional residuals is displayed in Figure 3.1, which demonstrates that the obtained residual set is much thinner tailed than the N(0,1) distribution.
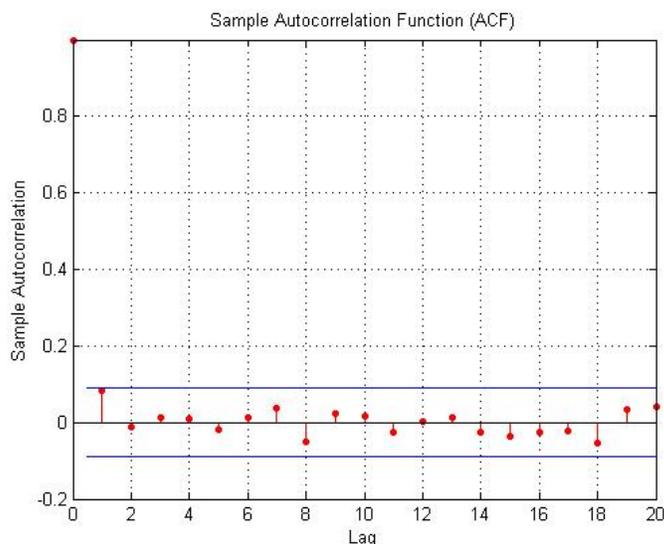
Figure 3.2: ACF of Unconditional Residuals for the RSLN-2 Example

Additionally, for this example, the distribution of the residual set is platykurtic. Both the left and right tails of the unconditional residual distribution are thinner than the standard Normal distribution, which is consistent with the effect in the earlier simpler example. This is to say the unconditional residual set has thinner tails than the residuals $\{R_t^m | \rho_t\}$.

Such a finding represents a hazard when comparing the unconditional residual set to a standard $\mathcal{N}(0,1)$ distribution in the context of long-term equities. The method tends to generate thinner tailed residuals than the underlying model, for situations where the two regimes have different means. If an analyst assumed that this residual set was representative of $R_t^m | \rho_t$, then he or she could be misled into using a model with tails that are significantly thinner than those underlying the data.

Another important assumption of the residual distribution under the null hypothesis is independence. The autocorrelation function of the residual set generated for the RSLN-2 example is shown in Figure 3.2. For the time 1 lag, the estimated autocorrelation is right on the border of rejection of independence at the 95% level. There does appear to be a possibility for a positive correlation effect between two consecutive residuals.

An explanation for this can be found in the model framework. When the state process

41

transitions from the low mean, high volatility state to the high mean, low volatility state, this is generally the period of the greatest state uncertainty (the state probabilities $p_1(t)$ and $p_2(t)$ are furthest away from both 0 and 1). This is due to the area of overlap of the two regime distributions. This means that periods of distortion due to state uncertainty are grouped together, which can create the correlation seen in Figure 3.2.

Together, the distortion of the residuals from the random innovations under the model, and the possible evidence of correlation in the residuals, suggests that the unconditional residual set is not suitable for appropriately testing the goodness-of-fit of the model to the data.

3.2.2 Weighted-Average Residuals

Instead of calculating the unconditional first two central moments of the data observations and then obtaining the residual set using the unconditional method, one can instead first calculate the state-specific residuals and then generate a single set of residuals through a weighted-average calculation.

Mathematically, as before, let

$$R_{t,k} = \frac{Y_t - \mu_{t,k}}{\sigma_{t,k}}, \quad t \in \{1,\ldots,N\}, \quad k \in \{1,\ldots,K\}$$

be the standardized model residual for at time t for state k. Then using the state probabilities, conditional on all information available at time t, $p_k(t), \in \{1,\ldots,K\}$, the weighted average residual at time is defined to be:

$$R_t^w = \sum_{k=1}^{K} R_{t,k} \cdot P[\rho_t = k | Y_t, \ldots, Y_1]$$

Similar to the unconditional case, for a process distributed as the hidden Markov distribution with randomly distributed Normal innovations under the null hypothesis, the weighted average residual set $R_t^w, t \in \{1,\ldots,N\}$ also will not distributed as $\mathcal{N}(0,1)$. Also similar to

the unconditional case, the weighted average residual set will also typically have thinner tails than what should be indicated by the data under the null.

Consider again the example of the two-state hidden markov with the state distributions for the data being $\mathcal{N}(-4, 1)$ and $\mathcal{N}(4, 1)$ respectively. Under the model, the data observation must be associated with one of the underlying states. Any observation in the interval $(-4, 4)$ will have a positive model-associated random innovation from one state, and a negative one for the other. The resulting weighted average residual will always be smaller in magnitude that either of the two associated state-specific random innovations. This is to say, still under the null, that no matter what the $r_t^m | \rho_t$ value is, the weighted average residual will always understate it, if $r_t^m | \rho_t$ lies between the mean values, μ_k, for the two regimes.

Another potential issue is that a deep value from the high-mean state which falls deep in the left tail, will have a residual that resembles a small (positive or negative) random innovation from the low-mean state. The same is true for deep right tail values from the low-mean state. For the typical RSLN model for stock returns, the weighted average residual distribution has much thinner tails than the underlying $R_t^m | \rho_t$ distribution, so, again, we do not recover an independent $\mathcal{N}(0, 1)$ residual sample using the weighted residual process.

Using the same simulation as in the UC example of Figure 3.2, we generated the weighted residual set $r_t^w, t \in \{1, \ldots, N\}$. The results are shown in Figure 3.3.

Similarly to the UC residual set, both the left and right weighted average residual tails are significantly thinner than the $\mathcal{N}(0, 1)$ quantiles used for the $r_t^m | \rho_t$ process, and it is clear that weighted average residuals do not replicate the $R_t^m | \rho_t$ distribution.

The autocorrelation function for the weighted average residuals exhibits the same pattern as that for the unconditional residuals (Figure 3.4). There is some evidence that there is positive correlation for consecutive residuals.

In both cases of averaging, either at the level of central moment calculation for the unconditional residual set, or through averaging of the residuals themselves, the residual sets do not retain the independence nor the $\mathcal{N}(0, 1)$ distribution of the $R_t^m | \rho_t$ process, and can't be assumed to do so for goodness of fit testing.
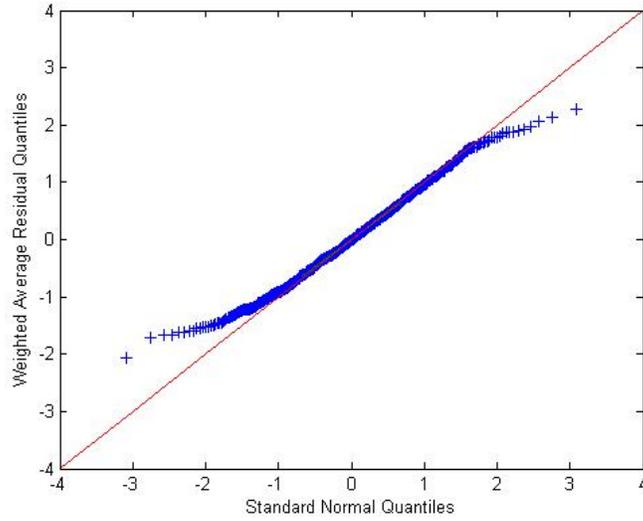
Figure 3.3: QQ-plot of N(0,1) and Weighted-Average Residual Quantiles for the RSLN-2 Example
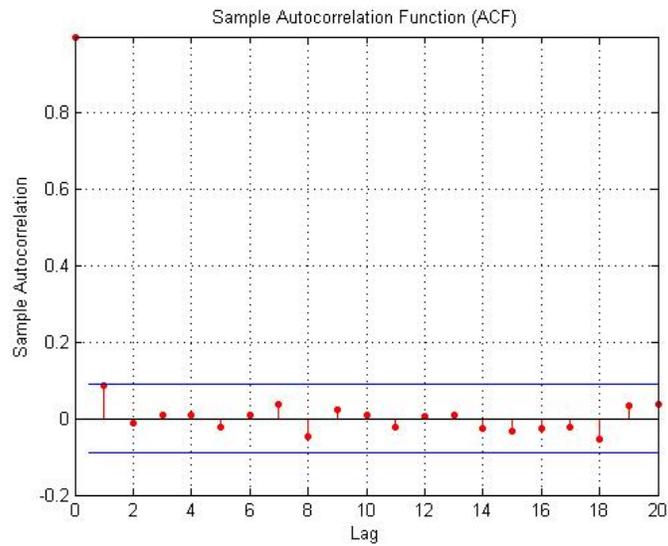
Figure 3.4: ACF of Weighted Average Residuals for the RSLN-2 Example

3.2.3   Indicator Residuals

A different approach is to use the information that the conditional model residual, $R_t^m|\rho_t$ is exactly equal one of the state-specific residuals, $R_{t,k}|\rho_t = k$, but we don't know which. Suppose we have data $\{y_t\}_{t=1}^T$, and we have calculated k sets of residuals, $\{r_{t,k}\}, t = 1, ..., T, k = 1, ..., K$. We may select the value of $r_{t,k}$ for which $p_k(t)$ is greatest.

That is, the model indicator residuals are defined as

$$R_t^i = \sum_{k=1}^K R_{t,k} \cdot I[P[\rho_t = k|Y_t, \ldots, Y_1] = \max_{\forall i}(P[\rho_t = i|Y_t, \ldots, Y_1])]$$

Indicator residuals were used in Hardy, Freeland & Till (2006) to test the goodness-of-fit of many long-term equity models, including hidden Markov models. The main attraction of the indicator residual set is that, for the range of parameters and models commonly used, we are often relatively confident in assessing which regime the process is in – that is, $p_k(t)$ is often near to 1.0 for one regime. In this case, we will have a high probability that $R_t^i = R_t^m|\rho_t$. This probability will depend strongly on the parameters of the model, and, in particular, the overlap between regime distributions. This method will work best if the regime distributions are reasonably distinct.

If $R_t^i = R_t^m|\rho_t$ with high probability, then we might expect the distribution of R_t^i to be close to the distribution of $R_t^m|\rho_t$. There are problems though. For example, in the cases where the wrong regime is selected, so $R_t^i \neq R_t^m|\rho_t$, it is likely that the values are very far apart, and also likely that $|R_t^i| \ll |R_t^m|\rho_t|$, because this method tends to choose smaller residuals over larger ones, from the set $\{r_{t,k}\}$ at each t, because regime probabilities are connected to the residuals, and a large residual will be associated with a lower regime probability than a smaller residual. This results in both the left and the right tails of the overall associated random innovation distribution being thinned by the indicator residual process.

The indicator residual set was calculated for the RSLN-2 model example using the same methodology as for the unconditional and weighted average residual sets, and its QQ-plot relative to the $\mathcal{N}(0,1)$ quantiles is displayed in Figure 3.5. Note that the left and right
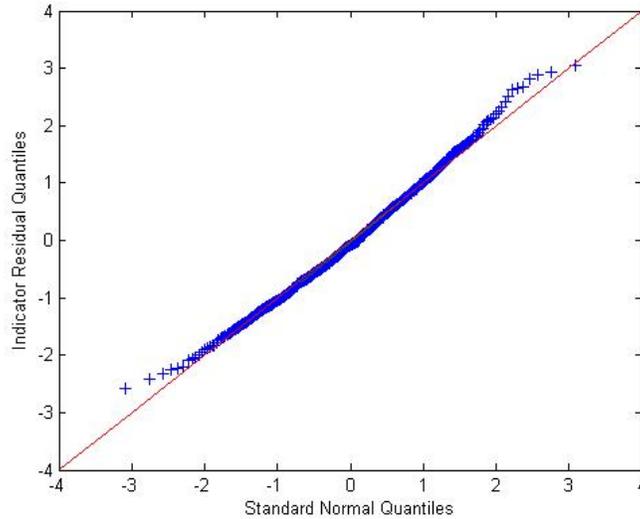
Figure 3.5: QQ-plot of N(0,1) and Indicator Residual Quantiles for the RSLN-2 Example

tails of the indicator residual set are again too thin, although less significantly than for the unconditional and weighted-average residual sets. However, it is still apparent that the indicator residual set does not generate the perfect $\mathcal{N}(0,1)$ set that replaced the random innovation under the model by construction for the example.

The autocorrelation function for the indicator residual, however, behaves very differently from those of the unconditional and weighted average residuals sets. Displayed in Figure 3.6, the lag 1 autocorrelation is significantly negative, representing negative association between consecutive residuals. An explanation for this again centers around the transition from the low-mean state to the high-mean state. The regime probabilities are influenced by the process history, and there is often a lag of 1 to 2 months between the regime shift happening, and the regime probabilities reflecting the shift by moving from less than to greater than 0.5 for the new regime. During the lag period, the residuals will be chosen from the 'wrong' regime. So, if the process shifts from the high mean low volatility state to the low mean high volatility state, and the wrong residual is selected, the indicator residual is likely to be large (and negative), followed by the switch to the right regime one month later, where the indicator residual is likely to be smaller. Similarly, switching from the high-vol regime to the low, if the indicator is wrong for one month, a large residual (from the high vol regime)

46
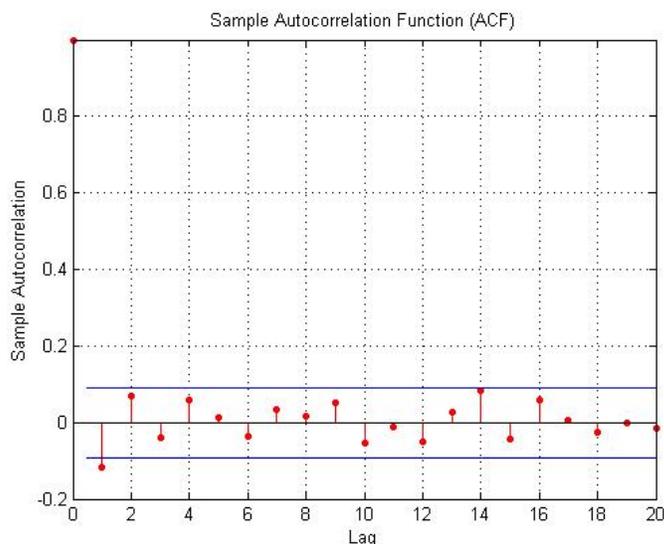
Figure 3.6: ACF of Indicator Residuals for the RSLN-2 Example

wrongly selected, is likely to be followed by a smaller residual correctly selected. Hence, we see a negative auto-correlation generated by the lag in identifying regime changes under the indicator method.

We see from the QQ and ACF plots that the indicator residual set, for this example, is probably not independent and is not normally distributed.

3.2.4 Stochastic Residuals

The main issue with indicator residuals is that the process tends to select the smaller residual, underestimating the possibility of larger residuals, except when transitioning, when the larger residual may be wrongly selected if there is a lag in identifying the regime change. The problem is more critical with more regimes, or with larger overlap between each regime distribution, so the probabilities associated with each state may all be close, and the residual selection becomes less likely to replicate the $R_t^m | \rho_t$ process.

This selection bias can be corrected by sometimes selecting the less likely residual, rather than always selecting the most likely residual. To determine how to do this we use the

regime probabilities, $\{p_j(t) = P(\rho_t = j | y_t, y_{t-1}, \ldots, y_0)\}_{j=1}^K$. Sample regime paths can be generated through Monte Carlo simulation, using the regime probabilities as the underlying distribution.

The stochastic residual set for this path is then the residual set that results from assuming the simulated path is the correct one. Letting τ_t be the simulated regime at time t, then a stochastic residual value at t is selected to be one of the values of $R_{t,k}$, with probability $p_k(t)$ for each regime $k = 1, \ldots, K$. That is, let

$$R_t^s = R_{t,k} \times I \left\{ \sum_{j=1}^{i-1} p_j(t) < U < \sum_{j=1}^{i} p_j(t) \right\}$$
$$t \in (1, \ldots, N) \tag{3.1}$$

where U is a randomly $U(0,1)$ variable, independent of Y_t and ρ_t.

The values of r_t^s are generated by Monte Carlo simulation. We can generate many different paths through the $r_{t,k}$ values. We will come back to this later. First, we show that the model stochastic residuals are independent $\mathcal{N}(0,1)$ random variables where the return process is RSLN-K.

A proof of the Normality of Stochastic Residuals

The following theorem shows that the model residuals generated in 3.1 will have a N(0,1) distribution.

Theorem 1 *Suppose that Y_t, $t = 1, \ldots n$ follows a regime switching process with K regimes, such that*

$$Y_t | \rho_t \sim \mathcal{N}(\mu_{\rho_t}, \sigma_{\rho_t}^2)$$

where ρ_t represents the regime at time t. Let

$$p_j(t) = Pr[\rho_t = j|Y_1, ..., Y_t].$$

Then define the random variable $R_t^s(U)$ as

$$R_t^s(U) = \sum_{i=1}^{K} \frac{Y_t - \mu_i}{\sigma_i} I \left\{ \sum_{j=0}^{i-1} p_j(t) < U < \sum_{j=1}^{i} p_j(t) \right\},$$

where $I\{\}$ is the indicator function and $U \sim U(0,1)$ and $p_0(t) = 0 \quad \forall t$.

Then, conditionally on $p_{i,j}, \mu_k, \sigma_k \quad \forall i, j, k \in \{1, \ldots, K\}$ and measurable $Y_1, \ldots, Y_t$,

$$R_t^s \quad \text{are i.i.d and } \mathcal{N}(0,1) \text{distributed.}$$

Proof 1 *Conditional on $Y_1, \ldots, Y_{t-1}$, the process at time t, Y_t, is simply a mixture of Normal random variables. The mixing weights are given by*

$$\alpha_i = \sum_{j=1}^{K} p_{ji} p_j(t-1) \quad for \quad i = 1, \ldots, K,$$

where p_{ji} is the one-period transition probability from state i to state j.

The conditional density for Y_t is then

$$f(y) = \sum_{i=1}^{K} \alpha_i f_i(y),$$

where $f_i(y)$ is the pdf for a Normal distribution with mean μ_i and variance σ_i^2, that is, the density corresponding to the ith regime distribution.

We also have

$$p_i(t) = \frac{\alpha_i f_i(y)}{f(y)}$$

So that $p_i(t)f(y) = \alpha_i f_i(y)$.

Now we note that

$$E[I\left\{\sum_{j=1}^{i-1} p_j\left(t\right) < U < \sum_{j=1}^{i} p_j\left(t\right)\right\}|Y_t] = p_i(t).$$

We complete the proof by finding the moment generating function for $R_t^s(U)$, conditional on $Y_1, Y_2, ..., Y_{t-1}$.

$$
\begin{aligned}
M_{R^s}(s)|Y_1, Y_2, ..., Y_{t-1} &= E[e^{s \cdot R_t(U)}|Y_1, ..., Y_{t-1}] \\
&= E\left[E\left[e^{s \cdot \sum_{i=1}^{K} \frac{Y_t - \mu_i}{\sigma_i} I\left\{\sum_{j=1}^{i-1} p_j(t) < U < \sum_{j=1}^{i} p_j(t)\right\}}|y_t\right]\right] \\
&= E\left[\sum_{i=1}^{K} e^{s \cdot \frac{y_t - \mu_i}{\sigma_i}} p_i(t)\right] \\
&= \sum_{i=1}^{K} \int_{-\infty}^{\infty} e^{s \cdot \frac{y - \mu_i}{\sigma_i}} p_i(t) f(y) dy \\
&= \sum_{i=1}^{K} \alpha_i \int_{-\infty}^{\infty} e^{s \cdot \frac{y - \mu_i}{\sigma_i}} f_i(y) dy \\
&= \sum_{i=1}^{K} \alpha_i e^{\frac{s^2}{2}} \\
&= e^{\frac{s^2}{2}}
\end{aligned}
$$

Hence $R_t^s(U)$ must have a standard Normal distribution; and, furthermore, since the conditional information on previous values of Y_t does not impact the mgf (it changes the α_i weights only), the stochastic residuals are also serially independent.

This result shows that for hidden Markov models of the form $N(\mu_{\rho_t}, \sigma_{\rho_t})$, the stochastic residual set is independent and identically $\mathcal{N}(0,1)$ distributed. Standard tests of goodness of fit, appropriate for normal models, can be applied to the stochastic residual set.

Stochastic Residual Sampling Error and Averaging

One potential issue with stochastic residuals is sampling error. However, stochastic residual sets can be generated multiple times. However, if one were to average residuals from the same time t over the different generated stochastic residual paths, we simply recover the weighted residuals, and the averaged sample of the stochastic model residuals is no longer normally distributed. On the other hand, if we sort the residual sets, so that we ignore the time argument (which should be uninformative, as the residuals are independent), then each Monte Carlo generated stochastic model residual set is a sample of quantiles of the N(0,1) distribution. To reduce the impact of sampling error, we may average the quantiles from a large number of simulations.

The summarize, a general procedure for an overall set of stochastic residuals that retains the $\mathcal{N}(0,1)$ distribution of the model stochastic residuals is:

1. Generate, say, 10,000 individual stochastic residual paths.

2. Sort the residuals values within each path.

3. Average the order statistics across the different residual sets.

4. Compare the averaged order statistics with the true Normal(0,1) quantiles.

The average ordered stochastic residual set was as above was generated for the RSLN-2 example with the perfectly standard normal model random innovations. The QQ-plot for the calculated average stochastic residual set against the standard Normal quantiles is displayed in Figure 3.7. The QQ-plot demonstrates that the constructed standard Normal quantiles are indeed recovered through the stochastic residual process. The ACF of the average stochastic residual set (Figure 3.8) also does not indicate any significant correlations between consecutive residuals.
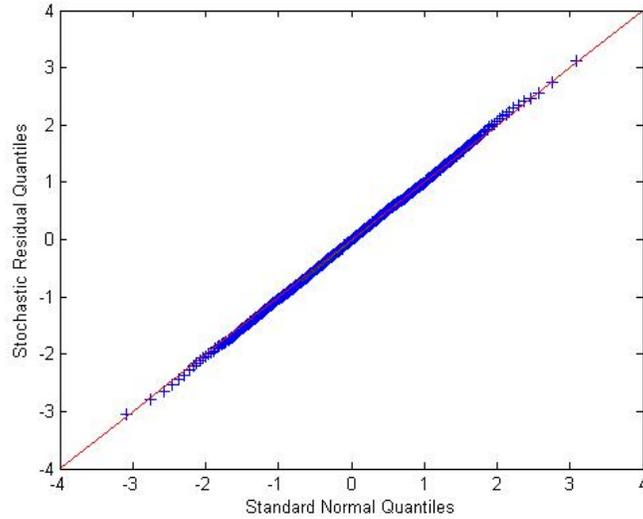
Figure 3.7: QQ-plot of N(0,1) and Average Stochastic Residual Quantiles for the RSLN-2 Example
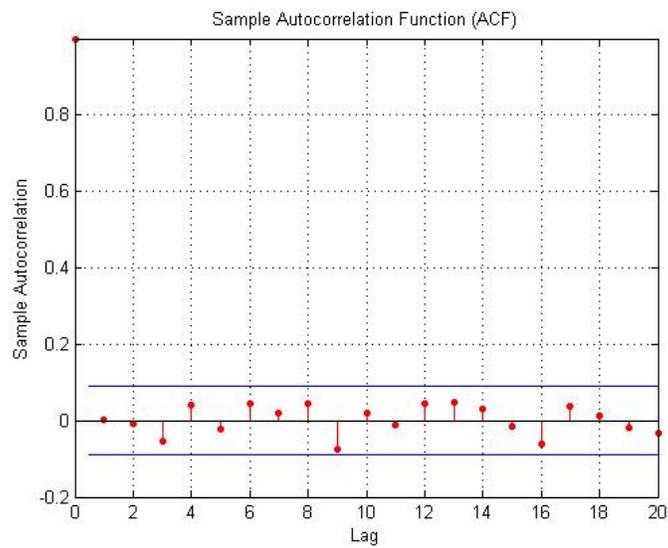
Figure 3.8: ACF of Stochastic Residuals for the RSLN-2 Example

The Power of Stochastic Residual Tests

Theorem 1 has shown that that stochastic model residuals for the regime switching conditionally Gaussian process are $\mathcal{N}(0,1)$ distributed. Hence, if we apply the process to data generated from the model, we will have a $\mathcal{N}(0,1)$ set of residuals, with some sampling variability if the number of simulations is small. However, what happens if the null hypothesis is false – that is, if the data is not regime switching conditionally Gaussian? Will the stochastic residual method, and especially the average stochastic residual method given its multiple simulations, have the power to reject the null hypothesis?

We do not explore this question in any analytic detail; that lies beyond the scope of this work. However, as a preliminary investigation we constructed an example, similar to the RSLN-2 example above. In this case, we again use the RSLN-2 model except that, instead of using the $\mathcal{N}(0,1)$ quantiles for the random innovations, the first 50 and the last 50 were raised to a power of 1.25. The resultant innovation quantiles are shown compared to the standard Normal quantiles using a QQ-plot in Figure 3.9 (a). This example was chosen to emulate a situation where the model is significantly thinner tailed than the data - a worst case scenario for the valuation of many investment guarantees.

The stochastic residual sets were simulated as described above, from the generated heavy tailed data set, and the averaged stochastic residual quantile set was calculated. A QQ-plot of the average stochastic residual quantiles, compared to the quantiles of the standard Normal distribution is shown in Figure 3.9 (b). The nature of deviation of the data from the model is apparent from the plot, although the magnitude of the deviation has been softened somewhat. The dampened power of the test should be kept in mind when the hidden Markov models are tested against data from the real S&P 500, however.
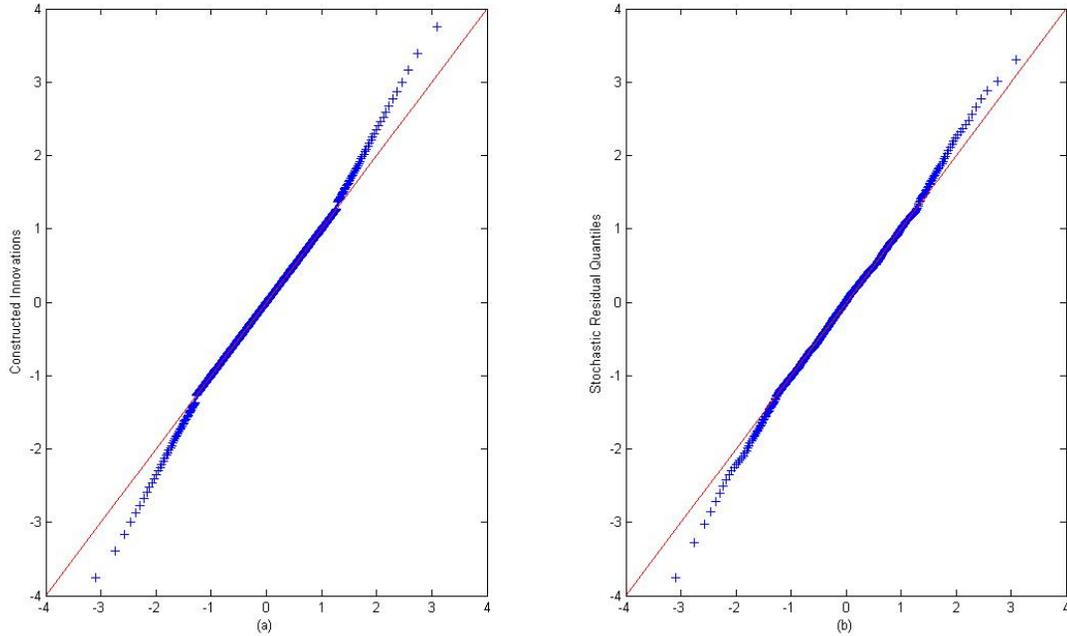
Figure 3.9: The Purposely Constructed Innovation Set (a), and the Resulting Average Stochastic Residual Set for the RSLN-2 Residual Test Power Example

3.3 Stochastic Residuals and the S&P 500 Hidden Markov Models

In this section, the long-term equity hidden Markov models fitted in Chapter 2 will be tested for goodness of fit, using the stochastic residual method. For the stochastic residual method, 10,000 simulations of stochastic sets of residuals were performed, and quantiles averaged. We also consider using the Indicator residuals, again, comparing with Normal quantiles, to demonstrate the differences between the two methods applied to real world data.

3.3.1 RSLN-2

The QQ-plots for the indicator residuals and average stochastic residuals of the RSLN-2 model for the S&P 500 data against standard Normal quantiles are displayed in Figure 3.10.

54
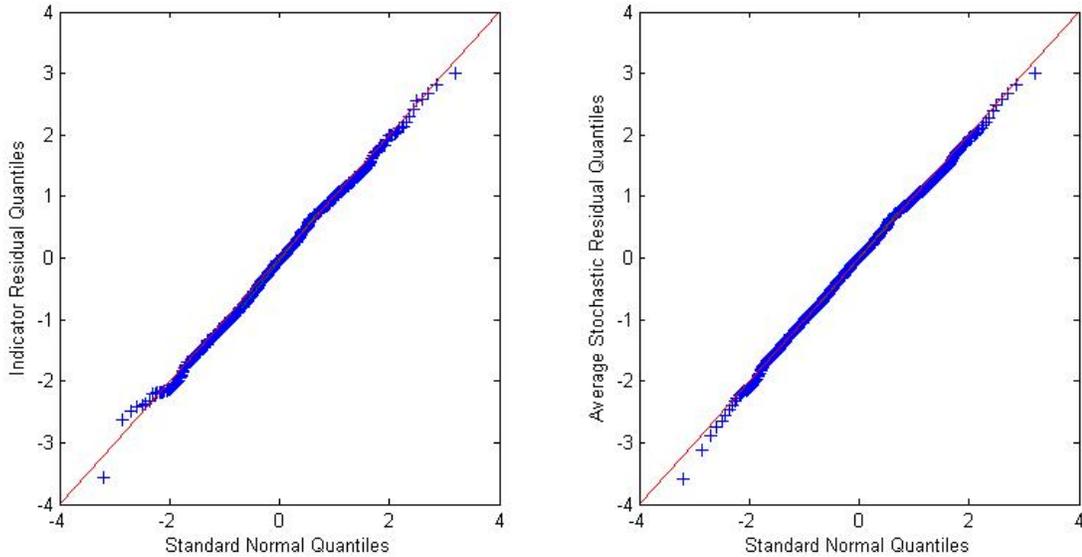
Figure 3.10: QQ-plots for the Indicator and Stochastic Residuals of the RSLN-2 Model for the S&P 500

This is a good example of the improved power of the stochastic residual method over the indicator method. Using the indicator residuals, the left tail in the data appears slightly thinner than the model (apart from October 1987). Using the stochastic residuals, the left tail in the data is somewhat thicker than that of the model. This represents a very significant difference for the purposes of long-term investment guarantee modeling from a qualitative perspective, as underestimating the left tail can have much more profound impact than overestimating it.

The average stochastic residual plot indicates the model does appear to fit reasonably well for the most part. However, one should keep in mind the RSLN-2 example from the previous section, where the average stochastic residual method was shown to dampen deviations. This may suggest that the left tail underestimation could be more severe than indicated. The average Jarque-Bera statistic (see Jarque & Bera, 1980) for the stochastic residual sets for the RSLN-2 model was 3.5184, which has a corresponding p-value of 0.1722, providing a more objective assessment that the stochastic residuals sets are consistent with model assumptions.
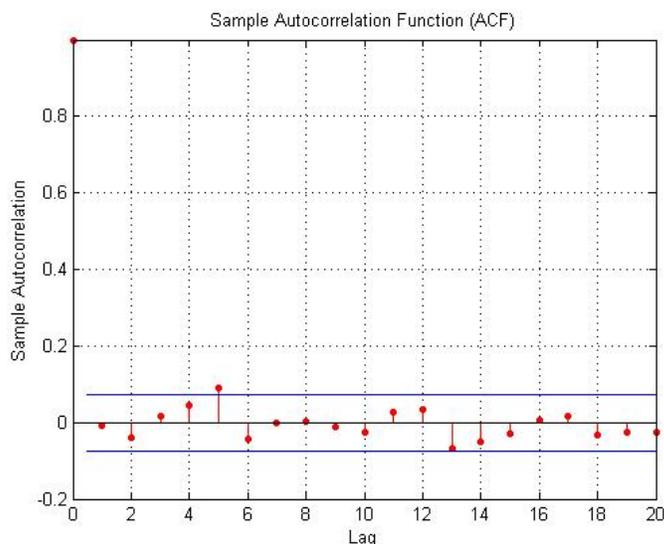
Figure 3.11: Autocorrelation Function for the Stochastic Residuals of the RSLN-2 Model for the S&P 500

The autocorrelation function for the RSLN-2 stochastic residuals does not indicate anything troubling. There is a significant correlation estimated for lag 5, but this is within the realm of statistical noise. The lag 1 estimated correlation is close to 0.

3.3.2 RSLN-3

The RSLN-3 stochastic residual QQ-plot is similar to that of the RSLN-2 model (Figure 3.12). The plot indicates that the model may slightly underestimate the left and right tails of the data. As was the case for the RSLN-2 model, the indicator residual plot suggested the model slightly overestimated the left tail; the distortion effect of the indicator residual method relative to the stochastic method can be seen again. Worth noting is that the October 1987 market crash is adequately captured on the RSLN-3 residual QQ-plot. The average Jarque-Bera statistic for the stochastic residual sets for the RSLN-3 model was found to be 2.3480, which has a corresponding p-value of 0.3091. The average JB statistic for the RSLN-3 model was significantly lower than that for the RSLN-2 model.

The ACF of the RSLN-3 stochastic residuals (Figure 3.13) also behaves similarly to that of
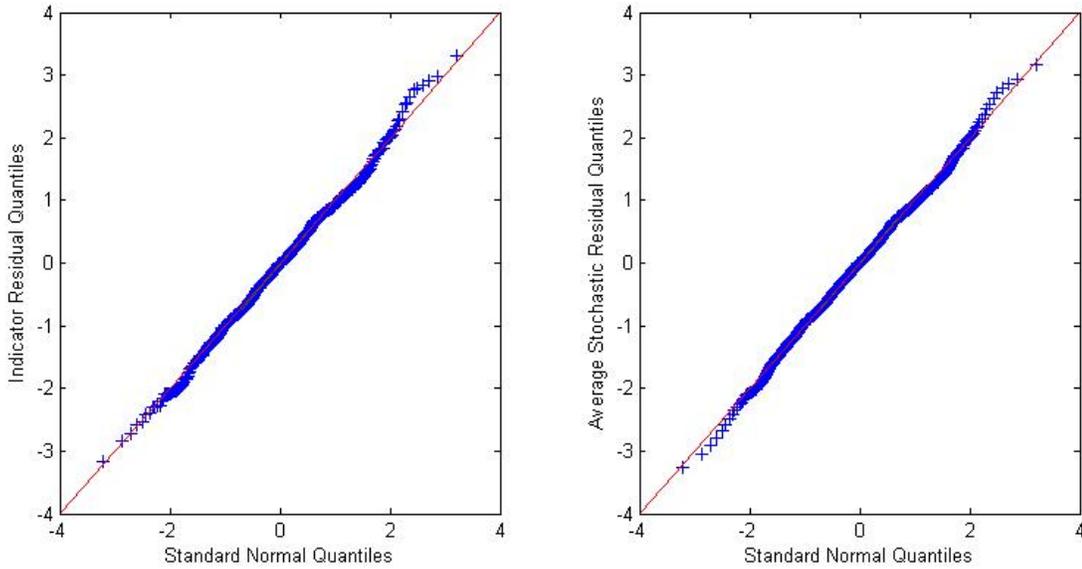
Figure 3.12: QQ-plots for the Indicator and Stochastic Residuals of the RSLN-3 Model for the S&P 500
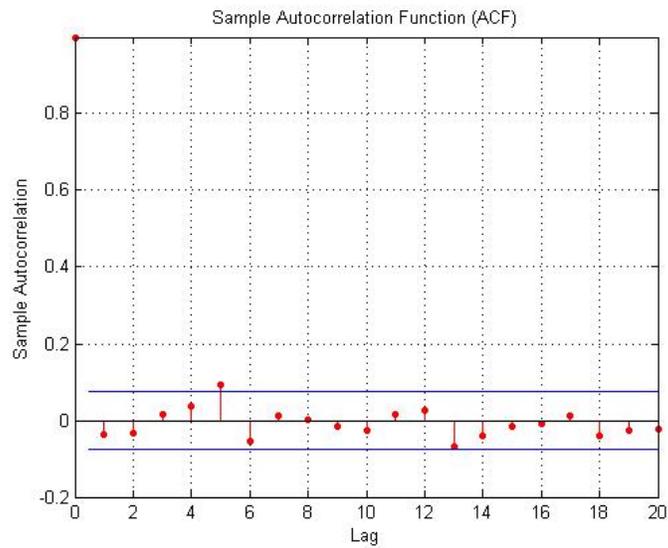
Figure 3.13: Autocorrelation Function for the Stochastic Residuals of the RSLN-3 Model for the S&P 500
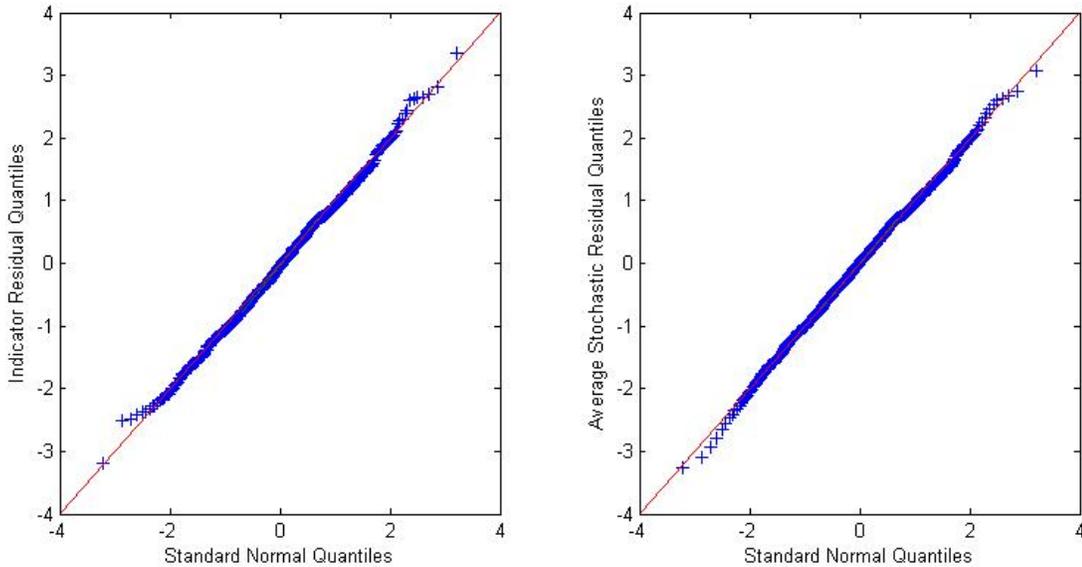
Figure 3.14: QQ-plots for the Indicator and Stochastic Residuals of the RSDD-2 Model for the S&P 500

the RSLN-2 average stochastic residual plot.

3.3.3   RSDD-2

The RSDD-2 model also fits the data well. The QQ-plot and the ACF plot of the stochastic residual set are displayed in Figures 3.14 and 3.15 respectively. The average Jarque-Bera statistic for the stochastic residual sets for the RSDD-2 model was 1.6514, which has a corresponding p-value of 0.4379. The stochastic residual QQ-plot behaves similarly to the RSLN models, and the October 1987 crash is also adequately captured under the RSDD-2 model.

3.3.4   RS-GARCH

The stochastic residual QQ-plot for the RS-GARCH model indicates there are potentially some problems with this model in the right tail (Figure 3.16). The QQ-plot indicates that
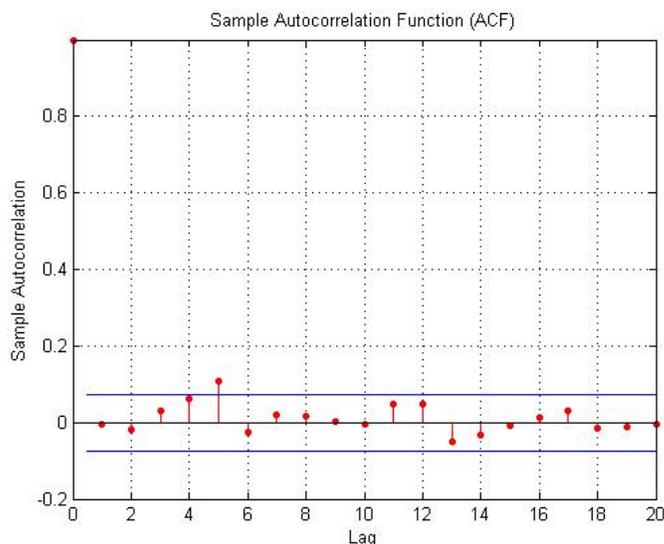
Figure 3.15: Autocorrelation Function for the Stochastic Residuals of the RSDD-2 Model for the S&P 500

the model may overestimate the deep right tail. The left tail of the data appears to fit well. Again the distorting effect of the indicator residual method can be seen. The average Jarque-Bera statistic for the stochastic residual sets for the RSGARCH model was 5.5430, which has a corresponding p-value of 0.0625. This result would still pass at the 5% level, but not by much. The RS-GARCH model performed the worst among the candidate models under the J-B test.

The autocorrelation plot of the RS-GARCH model looks identical to the other hidden Markov models. Overall, it appears the fit of the model may be slightly worse than those of the other models due to the right tail deviation.

3.3.5 MARCH

For the MARCH, a significant change in the QQ-plot of the residuals under the indicator method to the residuals under the stochastic method can clearly be seen in Figure 3.18. The indicator residual QQ-plot indicated that the left tail did not fit the data well, overestimating the deep right tail, while underestimating the shallow portion of the tail. Under the
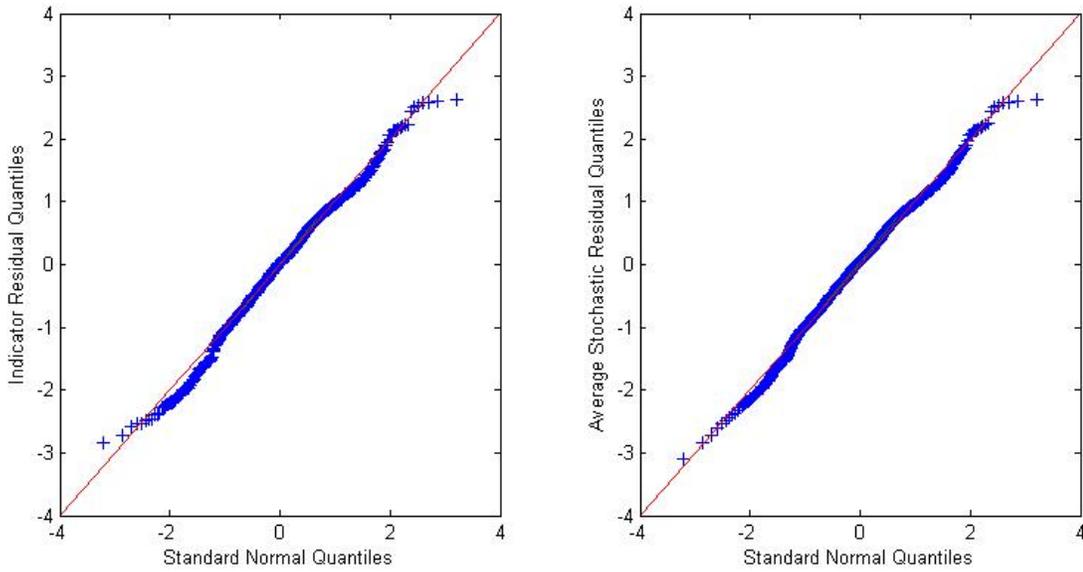
Figure 3.16: QQ-plots for the Indicator and Stochastic Residuals of the RSGARCH Model for the S&P 500
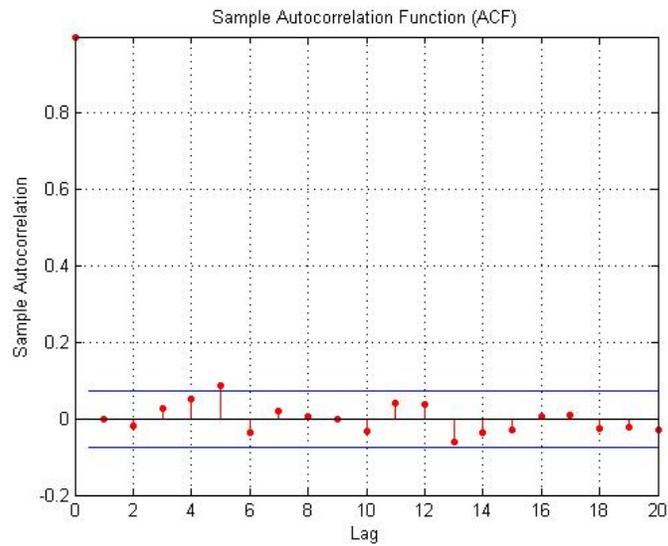
Figure 3.17: Autocorrelation Function for the Stochastic Residuals of the RSGARCH Model for the S&P 500
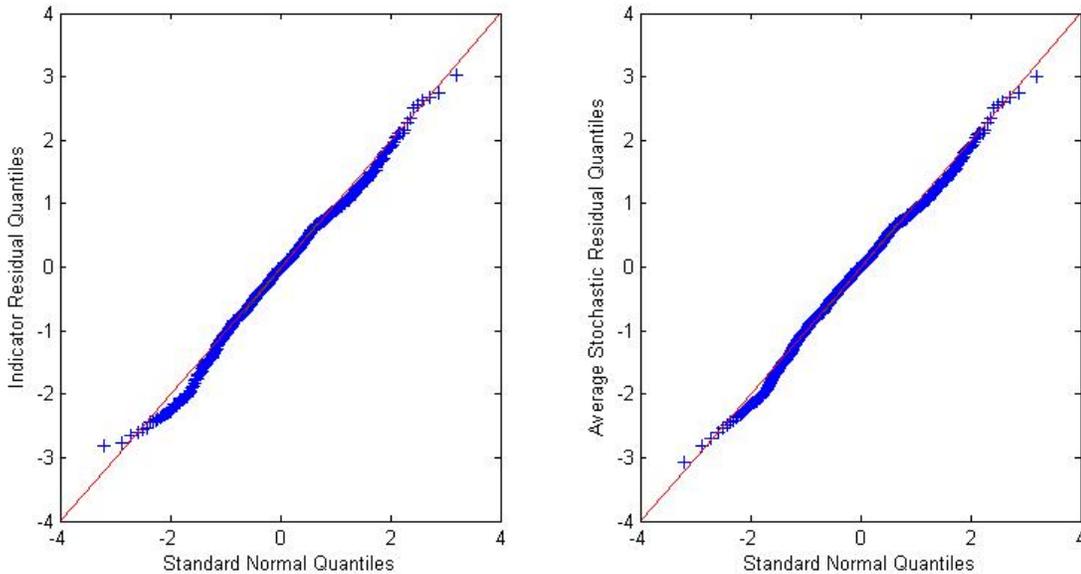
Figure 3.18: QQ-plots for the Indicator and Stochastic Residuals of the MARCH Model for the S&P 500

stochastic method, however, the left tail of the model fits the data quite well. The right tail of the data is also fit well by the MARCH model. The average Jarque-Bera statistic for the stochastic residual sets for the MARCH model was 2.5213, which has a corresponding p-value of 0.2834, providing further evidence that the obtained residuals closely resemble the assumed innovations of the model. The autocorrelation function for the MARCH model using the stochastic residuals performs as the other hidden Markov models.

3.3.6 A Summary of the Stochastic Residual Testing of the S&P 500 Models

All of the hidden Markov models for the S&P 500 performed well in the goodness-of-fit testing using the stochastic residual method. This provides evidence that hidden Markov models in general represent a sound fit for S&P 500 monthly data. Comparison tests between the models (as in Hardy, Freeland & Till, 2006) can now be performed knowing the comparisons will be between models that adequately fit the data.
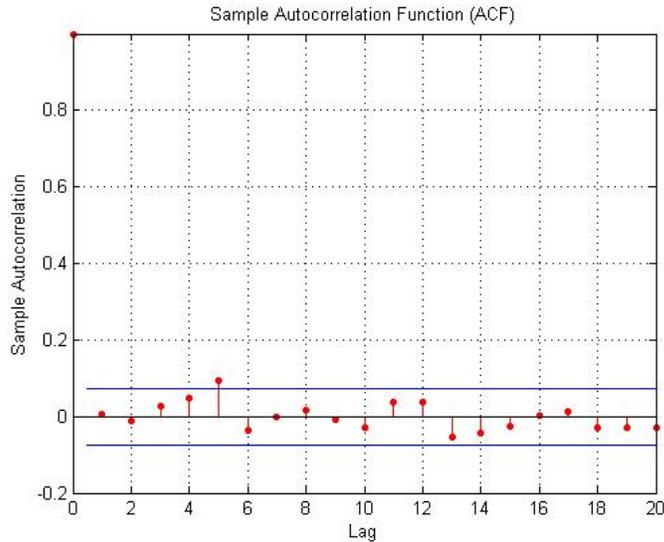
Figure 3.19: Autocorrelation Function for the Stochastic Residuals of the MARCH Model for the S&P 500

This section also demonstrated the benefit of using the stochastic residual method. For all five of the models, there were noticeable differences between the indicator residual set and the average stochastic residual set that change the results of the testing and the inference drawn from it.

3.4   Conclusion

In this chapter, goodness-of-fit testing for hidden Markov models was the focus. Several methods for generating residuals for hidden Markov models were presented, and distortion effects where shown for some of the methods previously used in the literature. A brand new hidden Markov residual method, the stochastic residual method, was introduced and it was proven that the stochastic residual method generates standard Normal residuals under the null hypothesis of the model. This method allows hidden Markov models to be properly tested for goodness of fit using standard Normal tests. The power of the stochastic residual method to reject under the null hypothesis was also explored, and it was shown that stochastic residual tests have the power to do this.

The stochastic residual method was used to test the fit of the hidden Markov models for the S&P 500 data from Chapter 2. All of the models performed reasonably well. Further, a significant difference was noticeable between the indicator residual set and the stochastic residual set for all models, indicating that usage of the stochastic residual method will have an impact on the results obtained through testing.

Chapter 4

Bayesian Long-Term Equity Hidden Markov Modeling

4.1   Introduction

The maximum likelihood fitting of the two-state hidden Markov models in Chapter 2 all resulted in a similar state framework: one state with a high frequency, a positive expected return and a relatively low volatility, and one state with a low frequency, negative expected returns and a relatively high volatility. These results were consistent with Hardy (2001), Wong & Chan (2005) and Panneton (2004). This nature of hidden Markov models when fitted to long-term equity data creates a specific problem for modeling long-term deep out-of-the-money options. The reserving of such contracts is concerned with tails of the model used, and the accuracy of those tails in the model depends significantly on the accuracy of the parameter estimation for the high volatility state. However, because that state has a relatively low frequency, it is likely that the high volatility state parameters will have high standard errors.

Using standard asymptotic results for assessing uncertainty of the maximum likelihood estimated parameters is problematic because the sample size for the high volatility state parameters is quite small, and because parameters may approach the bounds of the parameter

space.

One could stress-test candidate models, which involves using different parameter sets and observing the different reserving results for each. However, the different parameter sets would not have an assigned distribution, and thus only qualitative inference is feasible.

An alternative approach is to fit the hidden Markov models to long-term equity data using a Bayesian methodology. Bayesian modeling incorporates parameter uncertainty into the estimation structure.

This chapter will present some techniques from the literature that are available for Bayesian estimation of hidden Markov models, and then fit the models from Chapter 2 using Bayesian estimation. The chapter contributes to the literature, by demonstrating the results of two different Bayesian techniques for the RSLN-2 model when fitted to the S&P 500 monthly data. We also compare the models using the additional insight available from the Bayesian framework. A further objective of this chapter is to provide a framework for the analysis done in Chapters 5 and 6, where we focus on portfolio optimization and replication, respectively, under hidden Markov models, and assess the impact of parameter uncertainty on those processes.

4.2 Traditional Bayesian Estimation

Conventional Bayesian estimation (see, for example, Gelman et al., 2004) first requires one to assign a probability distribution to the model parameters, also known as the prior distribution. Let Θ represent the vector of parameters of interest, and $\pi(\Theta)$ represent the prior distribution of the parameters Θ. Many factors should be taken into account when deciding on $\pi(\Theta)$. For example, a thinner tailed prior distribution might be used when more certain about the likelihood of values for the parameter set Θ, whereas a thicker tailed prior might be more suitable when there is less certainty about the values of Θ.

Once the data is observed, and the prior distribution selected, the distribution of the parameters conditional on the data observations can then be calculated. Using the conditional

distribution of the data given the parameters, $f(\vec{y}|\Theta)$, where $\vec{y} = \{y_1, \ldots, y_N\}$, Bayes theorem gives:

$$\pi(\Theta|\vec{y}) = \frac{\pi(\Theta) \cdot f(\vec{y}|\Theta)}{f(\vec{y})}$$

The distribution $\pi(\Theta|\vec{y})$ is the posterior distribution, and represents the updated belief about the distribution of the parameters given the sample; $f(\vec{y})$, the unconditional distribution of the data, can be derived through the relationship

$$f(\vec{y}) = \int_{\Theta} \pi(\Theta) \cdot f(\vec{y}|\Theta) d\Theta$$

For simple models, using conjugate distributions for π and $f()$, the posterior distributions can be calculated analytically. For more complex models with many interdependent parameters, such as hidden Markov models, we can not determine posterior distributions analytically, but, using the Markov Chain Monte Carlo (MCMC) framework described below, we can use simulation to generate a sample from the joint posterior distribution, and, provided the sample is large enough, that will give us insight into the parameter uncertainties, as well as a resource for quantifying the impact of that uncertainty on calculations.

4.2.1 Markov Chain Monte Carlo

Markov Chain Monte Carlo algorithms (thoroughly discussed in Gilks, Richardson and Spiegelhalter, 1996) generate a Markov sequence of parameter vectors $[\Theta^{(0)}, \Theta^{(1)}, \Theta^{(2)}, \ldots, \Theta^{(r)}, \ldots]$, where $\Theta^{(r)}$ is the set of parameter values after the rth iteration of the algorithm. The stationary distribution of the sequence is the posterior distribution of Θ.

The sequence is serially correlated, so that the initialization, $\Theta^{(0)}$ has an impact on the resulting Markov chain. To minimize this impact, typically the first b observations of the chain are omitted from the sample for the purpose of inference. This is referred to as the 'burn-in' period. If there are R simulations performed during the algorithm, then the

sequence of random variables $[\Theta^{(b)}, \Theta^{(b+1)}, \ldots, \Theta^{(R)}]$ converges to the desired posterior distribution.

The choice of the starting parameters should typically be estimates that are relatively central for the posterior parameter distribution, such as the maximum likelihood parameter estimates. Smart choices will typically result in the chain converging more quickly than if radical choices for parameters were input. However, even very central choices for starting parameters will still have an impact on the Markov chain, and thus a 'burn-in' period is always advisable.

For multi-parameter models, one approach for each iteration is to update the parameters one at a time (see, for example, Robert and Titterington, 1998), with each parameter being generated conditional on the data and the latest parameter values in the chain. Let $\Theta^{(r)}$ represent the vector of parameters generated during the rth iteration of the Markov chain, and $\theta_d^{(r)}, d \in \{1, \ldots, D\}$ be the generated dth parameter of the rth iteration. The $(r+1)^{th}$ update of the parameter vector $\Theta^{(r+1)}$ is generated by iterating through the individual parameters consecutively, that is

$$
\begin{aligned}
\theta_1^{(r+1)} &\sim \pi(\theta_1|\vec{y}, \theta_2^{(r)}, \ldots, \theta_D^{(r)}) \\
\theta_2^{(r+1)} &\sim \pi(\theta_2|\vec{y}, \theta_1^{(r+1)}, \theta_3^{(r)}, \ldots, \theta_D^{(r)}) \\
&\ldots \\
\theta_D^{(r+1)} &\sim \pi(\theta_1|\vec{y}, \theta_1^{(r+1)}, \ldots, \theta_{D-1}^{(r+1)})
\end{aligned}
$$

With this framework, the task of simulating Θ from its joint distribution $\pi(\Theta|\vec{y})$ reduces to single simulations of the θ_d's from their conditional distributions $\pi(\theta_d|\vec{y}, \Theta_{-d})$, which can be simulated from the proportions

$$
\pi(\theta_d|\vec{y}, \Theta_{-d}) \propto f(\vec{y}|\Theta) \cdot \pi(\theta_d)
$$

where Θ_{-d} is the vector of parameters Θ without the parameter θ_d.

It is important to note that the parameter vector generated at each iteration is set of dependent parameters, as each iteration is a simulation of the vector of parameters from the joint posterior distribution $\pi(\Theta|\vec{y})$. Maintaining this dependence is important for model accuracy when simulating predictive distributions, which will be described in Section 4.2.2.

4.2.1.1 The Metropolis Hastings Algorithm

A popular technique for simulating from the posterior distribution using MCMC is the Metropolis-Hastings algorithm (MHA) (first introduced in Metropolis and Ulam, 1949 and Metropolis et al, 1953, and then generalized in Hastings, 1970). The main advantage of the MHA is that it can be used when the posterior parameter distribution has no closed form. The method involves simulating parameters from another distribution, known as the candidate distribution, and then using an acceptance-rejection methodology to shape the accepted candidate parameters such that their distribution follows the posterior distribution.

A joint candidate distribution can be specified for all D model parameters, or separate candidate distributions can be specified for each parameter. Candidate distributions are usually defined conditional on the value in the previous iteration of the chain. The objective when specifying candidate distributions is to achieve an acceptance rate that falls in an acceptable range, typically 35%-50%. If there are too many acceptances, there is a risk that the chain is missing areas of the parameter space. If there are too many rejections, the algorithm will converge to the posterior distribution too slowly, and most likely the candidate distributions can be adjusted for a more efficient algorithm.

Let $\lambda_d^{(r)}, d \in \{1, \ldots, D\}$ be the candidate parameter for parameter θ_d for iteration r. Let $q_d(\cdot|\theta_d), d \in \{1, \ldots, D\}$ be the candidate distribution function for parameter θ_d, conditional on the algorithm's previous iteration's parameter value. Therefore, the candidate distribution function of $\lambda_d^{(r)}$ will be $q_d(\lambda_d^{(r)}|\theta_d^{(r-1)})$.

Under the MHA, the candidate parameter $\lambda_d^{(r)}$ is accepted with probability α, defined by

$$\alpha = \min\left(1, \frac{\pi_d(\lambda_d^{(r)}|\vec{y}, \Theta_{\sim d}^{(r-1,r)}) \cdot q(\theta_d^{(r-1)}|\lambda_d^{(r)})}{\pi_d(\theta_d^{(r-1)}|\vec{y}, \Theta_{\sim d}^{(r-1,r)}) \cdot q(\lambda_d^{(r)}|\theta_d^{(r-1)})}\right) \tag{4.1}$$

and rejected with probability $1 - \alpha$.

$\Theta_{\sim d}^{(r-1,r)}$ is the set of current values of the parameter set Θ, without parameter θ_d, at the time that the candidate parameter $\lambda_d^{(r)}$ is being simulated (assuming parameters are simulated in the order $1, 2, \ldots, d, \ldots, D$):

$$\Theta_{\sim d}^{(r-1,r)} = [\theta_1^{(r)}, \ldots, \theta_{d-1}^{(r)}, \theta_{d+1}^{(r-1)}, \ldots, \theta_D^{(r-1)}]$$

The ratio of the posterior distributions evaluated at $\lambda_d^{(r)}$ and $\theta_d^{(r-1)}$ respectively in equation (4.1) is equal to the ratio of the likelihood of the data times the prior distributions, evaluated at $\lambda_d^{(r)}$ and $\theta_d^{(r-1)}$ respectively.

Let $L(\hat{\theta}_d, \Theta_{\sim d})$ represent the likelihood function under the model of the data set of interest using $\hat{\theta}_d$ for parameter θ_d and the parameter set $\Theta_{\sim d}$ as normal. Then α can be calculated through the formula

$$\alpha = \min\left(1, \frac{L(\lambda_d^{(r)}, \Theta_{\sim d}^{(r-1,r)}) \cdot \pi_d(\lambda_d^{(r)}) \cdot q(\theta_d^{(r-1)}|\lambda_d^{(r)})}{L(\theta_d^{(r-1)}, \Theta_{\sim d}^{(r-1,r)}) \cdot \pi_d(\theta_d^{(r-1)}) \cdot q(\lambda_d^{(r)}|\theta_d^{(r-1)})}\right)$$

Finally, the likelihood function $L(\hat{\theta}_d, \Theta_{\sim d})$ can be computationally uncomfortable to work with due to its size, so it is common to work instead with the natural logarithm of the likelihood, $l(\Theta) = \log(L(\Theta))$.

The general MHA used in this thesis can be summarized by the following steps:

1. Initialization:

   - Specify prior distributions for all model parameters, $\pi_d(\theta_d), d \in \{1, \ldots, D\}$
   - Specify candidate distributions for all model parameters, $q_d(\cdot|\theta_d), d \in \{1, \ldots, D\}$
   - Specify initial values for all parameters, $\Theta^{(0)} = [\theta_1^{(0)} \ldots \theta_D^{(0)}]'$

2. For the rth iteration of the process, perform the following for each parameter $d, d \in \{1, \ldots, D\}$:

   (a) Sample a candidate parameter $\lambda_d^{(r)}$ from its candidate distribution function $q_d(\cdot|\theta_d^{(r-1)})$

(b) Calculate the acceptance probability α using the equation

$$\alpha = \min\left(1, \frac{L(\lambda_d^{(r)}, \Theta_{\sim d}^{(r-1,r)}) \cdot \pi_d(\lambda_d^{(r)}) \cdot q(\theta_d^{(r-1)}|\lambda_d^{(r)})}{L(\theta_d^{(r-1)}, \Theta_{\sim d}^{(r-1,r)}) \cdot \pi_d(\theta_d^{(r-1)}) \cdot q(\lambda_d^{(r)}|\theta_d^{(r-1)})}\right)$$

(c) Generate a Uniform(0,1) random number U

(d) If $U < \alpha$, then set $\theta_k^{(r)} = \lambda_d^{(r)}$. Otherwise, set $\theta_d^{(r)} = \theta_d^{(r-1)}$

3. Repeat step 2 for R total iterations.

There will be two versions of MHA methods explored for the RSLN-2 model in this chapter: one using state simulation, and the other using the marginal likelihood method used in Chapter 3 for maximum likelihood estimation. For the state simulation method, there is a slight deviation from the above steps for the transition probability parameters. This will be further explained in Section 4.3.1.

4.2.2 The Predictive Distribution

Once the joint posterior distribution of the parameter set Θ has been obtained, a distribution of future values of the process can be obtained conditional on the data observations. This distribution is known as the predictive distribution.

Assume the process has T observed data points, $\vec{y} = \{y_1, \ldots, y_T\}$. The predictive distribution of $y_r|\vec{y}$, for any $r > T$, has probability density function

$$f(y_r|\vec{y}) = \int_{\Theta} f(y_r|\Theta) \cdot \pi(\Theta|\vec{y})d\Theta$$

When Markov Chain Monte Carlo has been used to obtain the joint posterior distribution of the parameter set Θ, then the predictive distribution can be obtained by simulation. Each of the simulated values of the parameter set under the chain, $\Theta^{(s)}, s \in \{b+t, \ldots, S\}$ is an equally like set of parameter observations under the posterior distribution. Therefore, for any $r > T$, the predictive distribution of y_r, (or of any function which depends on future

71

values of y_t) can be obtained by simulation, where each simulation uses a different parameter set $\Theta^{(s)}$, generated using MCMC.

The values for y_r simulated using the predictive distribution incorporates both parameter and idiosyncratic uncertainty, since the simulations use parameters from across the range of values from the posterior distribution. This will be different from values simulated assuming fixed parameter values, such as the maximum likelihood parameter estimates. The predictive distribution is a very useful tool for Bayesian model adequacy analysis (Robert et al. (1999)), and will be further explored for hidden Markov long-term equity models in this chapter and beyond.

Specifically related to applications in this thesis, where the tails of the data are of paramount interest, the Bayesian predictive distribution of future data observations will incorporate the higher parameter uncertainty relating to the low-frequency state, implicitly, in applications where we can select different parameter sets from the MCMC algorithm in our simulations of output measures such as guarantee payments.

4.3 Markov Chain Monte-Carlo for Hidden Markov Models

For the ML estimation in Chapter 2, the likelihood of any particular data point is defined to be the average of the regime-specific likelihoods, with the data dependent regime probabilities, $p_k(t), k \in \{1, \ldots, K\}$, as weights. That is, the contribution of y_t to the likelihood function, $L(\Theta)$, is for the 2-state process,

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

$\quad = f(y_t | \rho_t = 1, \Theta) \, Pr[\rho_t = 1 | y_{t-1}, ..., y_1, \Theta] + f(y_t | \rho_t = 2, \Theta) \, Pr[\rho_t = 2 | y_{t-1}, ..., y_1, \Theta]$

$\quad = f(y_t | \rho_t = 1, \Theta) p_1(t) + f(y_t | \rho_t = 2, \Theta) p_2(t)$

$\quad = f_1(y_t) \, p_1(t) + f_2(y_t) \, p_2(t)$

These contributions are combined to give the likelihood; for the general $K-$state case,

$$L_{ML} = \prod_{t=1}^{N}\sum_{k=1}^{K}f_k(y_t)\,p_k(t) \tag{4.2}$$

The resulting likelihood is sometimes known as the marginal likelihood. Marginal likelihoods were used for Bayesian estimation of hidden Markov models in Hardy (2002).

Another approach, which implicitly treats the (unknown) states as parameters, is to simulate the state path, and then calculate the likelihood given the simulated path. See, for example, Robert & Titterington (1998). The simulation of the underlying state path, $\rho_1, \ldots, \rho_N$, is conditional on both the data set and the parameter set being tested. Once the state path has been simulated, the likelihoods are calculated conditional on the state path, which makes the calculation straightforward. Recall that, given $\rho_t = k$, say, y_t is normally distributed, with mean and variance μ_k and σ_k^2, and is independent of the other values in $\vec{y}$.

That is, let $\vec{\rho} = (\rho_1, ..., \rho_n)$ denote the simulated state path, then the state simulation likelihood is calculated as

$$L_{SS} = \prod_{t=1}^{N}\sum_{k=1}^{K}f_k(y_t)I(\rho_t = k) \tag{4.3}$$

where I is the indicator function, and $f_k(y_t)$ is defined, as above, as the density for y_t given that the process is in state k at t.

State simulation is quite popular for hidden Markov model MCMC in the literature (see, for example, Shephard, 1994, Chib, 1996 or Robert & Titterington, 1998), typically using Gibbs sampling (Geman & Geman, 1984), which is a special case of the MHA that does not use acceptance-rejection. One reason that state-simulation MCMC is popular is that the sample state paths can be used as a tool for diagnostic checks. Another reason is that using this approach, state-specific parameters will only be affected by data points with the corresponding simulated state. Therefore, state simulation could give a more 'model-oriented' picture of what different parameter sets would look like given the different plausible underlying state paths of the series. The simulated states also have been used as a diagnostic

tool for model checking in the MCMC Bayesian literature.

Section 4.4.1 will compare the the marginal likelihood and state simulation methods for the RSLN-2 model using the S&P 500 monthly data set. The objective will be to identify any noticeable differences in the posterior and predictive distributions of the simulations in the context of long-term equity models. The algorithms will be now further described in Sections 4.3.1 and 4.3.2.

4.3.1 Marginal Likelihood MCMC

The marginal likelihood MCMC method, used in Hardy (2002), does not simulate the underlying regime, but instead defines the log-likelihood by averaging across the regimes using the data-dependent regime probabilities $p_k(t)$. Mathematically, the log-likelihood of the data observation set y is defined as

$$ l = \sum_{t=1}^{N} \log\left(\sum_{k=1}^{K} p_k(t) \cdot f_k(y_t) \right) $$

which is simply the log of equation (4.2). The probabilities $p_k(t)$ are dependent on the transition probability parameters $p_{1,2}$ and $p_{2,1}$, meaning these parameters are defined within the marginal likelihood. Therefore, they can be treated as any other parameter under the MHA and can be subject to the standard acceptance rejection procedure.

The algorithm used for the marginal likelihood method is identical to the algorithm described in Section 4.2.1.1. For each iteration of the algorithm, the transition probability parameters, $p_{1,2}$ and $p_{2,1}$, are simulated first, followed by the state-specific parameters (μ_1, σ_1, μ_2 and σ_2 for the RSLN-2 model), though changing the order of simulation did not result in any noticeable differences in the posterior distributions. The specific prior and candidate distributions used for each of parameters were tailored based on the specific models and the S&P 500 data set, and will be presented in Section 4.4.

4.3.2 State Simulation MCMC

Under the state simulation MCMC method, the simulated state process, which is generated using the candidate parameters for the transition probabilities, is used in the likelihood calculation. The acceptance-rejection step is applied simultaneously, for the transition parameters and the resulting simulated regime paths.

For iteration r in the MHA under the state simulation method for the RSLN-2 model, the algorithm is adapted as follows. Let $\vec{\theta}^{(r-1)}$ denote the parameter vector of state-specific parameters, without the transition probabilities, from the $r-1$th iteration. Let $\vec{P}^{(r-1)}$ denote the vector of transition probabilities from the $r-1$th iteration, and let $\vec{\rho}^{(r-1)}$ denote the simulated state path from the $r-1$th iteration. Then the likelihood based on the simulated states for the $r-1$th iteration is $L_{SS}(\vec{\theta}^{(r-1)}, \vec{\rho}^{(r-1)})$

Consequently for the rth iteration:

1. We select candidate values for the transition probabilities, $\vec{\lambda_P}$, say.

2. Use the candidate transition probabilities to simulate a candidate state path $\vec{\lambda_\rho}$, using the method described in equations (4.4), (4.5) and (4.6) below.

3. Calculate the joint acceptance-rejection probability for the transition probabilities and state path:

$$\alpha = \min\left(1, \frac{L_{SS}(\vec{\theta}^{(r-1)}, \vec{\lambda_\rho})\,\pi(\vec{\lambda_P})\,q(\vec{P}^{(r-1)}|\vec{\lambda_P})}{L_{SS}(\vec{\theta}^{(r-1)}, \vec{\rho}^{(r-1)})\,\pi(\vec{P}^{(r-1)})\,q(\vec{\lambda_P}|\vec{P}^{(r-1)})}\right)$$

4. If the transition probabilities are accepted, set the state path $\vec{\rho}^{(r)} = \vec{\lambda_\rho}$, and set $\vec{P}^{(r)} = \vec{\lambda_P}$. If the transition probabilities are rejected, then $\vec{\rho}^{(r)} = \vec{\rho}^{(r-1)}$, and $\vec{P}^{(r)} = \vec{P}^{(r-1)}$

5. Sample and accept/reject candidates for the other parameters of the model, μ_1, σ_1, μ_2 and σ_2, using the likelihood based on the state path $\vec{\rho}^{(r)}$.

We use the local updating procedure of Robert et al. (1993) to simulate the underlying states. Let $\rho_t^{(r-1)}$ be the simulated regime for time point t for the rth iteration. Then regimes for

all time observations for iteration r are simulated according to the following probabilities. We use the transition probabilities and state process ρ_t in these equations, although in the algorithm, these would be replaced with the candidate transition probabilities and simulated states.

$$Pr[\rho_1^{(r)} = k] \quad \propto \quad \pi_k \, p_{k,\rho_2^{(r-1)}} \, f_k(y_1|\theta^{(r-1)}) \tag{4.4}$$

$$Pr[\rho_t^{(r)} = k] \quad \propto \quad p_{\rho_{t-1}^{(r)},k} \, p_{k,\hat{\rho}_{t+1}^{(r-1)}} \, f_k(y_t|\theta^{(r-1)}) \tag{4.5}$$

$$P[\hat{\rho}_n^{(r)} = k] \quad \propto \quad p_{\rho_{T-1}^{(r)},k} \, f_k(y_n|\theta^{(r-1)}) \tag{4.6}$$

4.3.3 Switching states

One complication for the μ, σ and other state-specific parameters is that the states, and thus their associated parameters, can switch places. Specifically this means that at some point during the algorithm, the state that had previously represented the high volatility state switches to the low-volatility state, and vica-versa. This will cause a mingling of both transition probability and state-specific parameters that would most likely result in them all having large cluttered posterior distributions. The resulting posterior distributions would not be very useful for analysis of the hidden Markov model, as one would obtain little understanding about the uncertainty about the different states. This is traditionally handled in the literature (Robert & Titterington, 1998 and Hardy, 2002) by continually simulating candidate parameters until the desired signs of $\mu_1 - \mu_2$ or $\sigma_1 - \sigma_2$, for example, are preserved. Specifically, if one desired state 'one' to be the high expected return, low volatility state, candidate parameters for μ_1 are simulated until a parameter that is greater than the current iteration's value for μ_2 is generated. After this check is met, one would proceed with the standard MHA acceptance-rejection for the candidate parameter. The downside to this method is the longer computation time. However, for this thesis, it was found that this method did not increase algorithm time to beyond reasonable limits, and thus this method was used for the Bayesian estimation of all models in this chapter.

4.4   S&P 500 Bayesian Estimation

This section presents the results for the candidate models described in Chapter 2 when fitted to the S&P 500 monthly data set (from January 1950 to October 2010) using the Bayesian approaches described in Sections 4.1-4.3. Assessments of the convergence of the parameter posterior distributions for each of the models be found in Appendix A.

4.4.1   RSLN2

The RSLN-2 model has six parameters: μ_1, μ_2, σ_1, σ_2, $p_{1,2}$, and $p_{2,1}$. The MCMC algorithm was run using both the state simulation and marginal likelihood methods for the RSLN-2 model. For both methods, the MLE parameter estimates were used as the starting parameters of the chain. The choices here for prior distributions for the μ and σ parameters are consistent with those found throughout much of the hidden Markov MCMC literature referenced already, largely due to their being conjugate priors, though this is not a requirement of Metropolis-Hastings MCMC.

The mean parameters, μ_1 and μ_2, have identically defined $N(0, 0.02^2)$ prior distributions. The candidate distribution for μ_1, conditional on the previous iteration's value of $\mu_1^{(r-1)}$, is $N(\mu_1^{(r-1)}, 0.005^2)$, and the candidate distribution for μ_2, conditional on the previous value $\mu_2^{(r-1)}$, is $N(\mu_1^{(r-1)}, 0.015^2)$. Both candidate distributions resulted from trial and error to yield MHA acceptance probabilities of approximately 40%.

The volatility parameters, σ_1^2 and σ_2^2, have inverse gamma distributions for both their prior and candidate distributions. Using inverse gamma distributions, the lower parameter values lead to the larger the variance in the corresponding candidate distribution, and vice-versa. This results in the process being less likely to be stuck with repeatedly low candidate parameters for the processes volatility. The prior distribution for σ_1^{-2} is Gamma$(1.038, 833.30)$ and the prior distribution for σ_2^{-2} is Gamma$(0.0361, 5263.2)$. Both of these prior distributions are designed to have means equal to the maximum likelihood estimates of the parameters, along with large enough variances such that the prior distribution will not have too large an impact on the resulting posterior. The candidate distributions for σ_1^{-2} and σ_2^{-2}, conditional
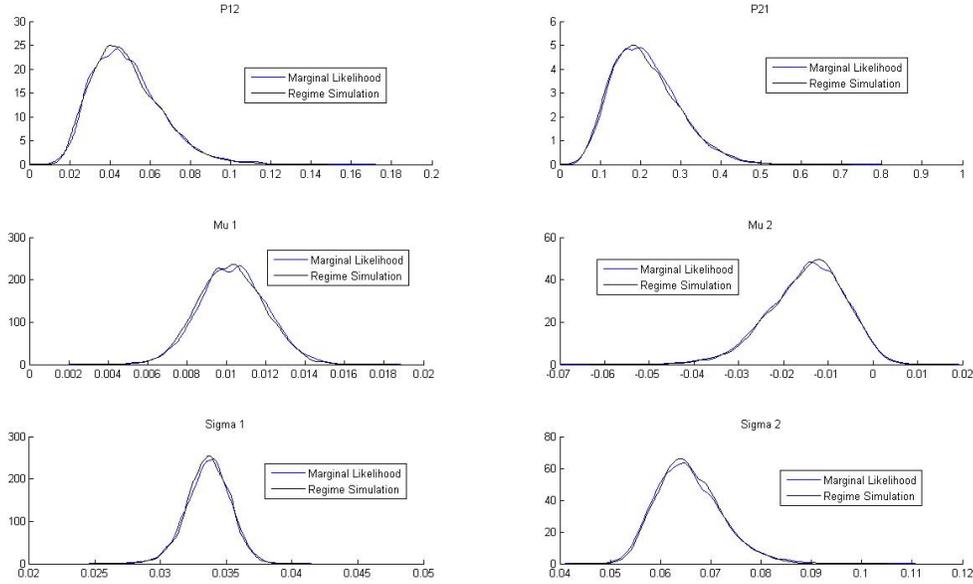
Figure 4.1: Distributions of the RSLN-2 parameters under Regime-Switching Bayesian Estimation

on the previous iteration's parameters values are Gamma$(5^2, \frac{\sigma_1^{-2(r)}}{5^2})$ and Gamma$(2.5^2, \frac{\sigma_2^{-2(r)}}{2.5^2})$ respectively, again designed to yield acceptance ratios of approximately 40%.

The prior distributions for $p_{1,2}$ and $p_{2,1}$ were Beta$(4, 48)$ and Beta$(2, 5)$ respectively.

The posterior distributions of the parameters under each of the regime-switching Bayesian MCMC methods are displayed in Figure 6.3. The type of hidden Markov MCMC (marginal likelihood versus state simulation) was found to have very minimal impact on the obtained posterior distributions. The estimated posterior distributions of each of the parameters were nearly identical for each method. The bivariate relationships between the state-specific parameters, shown in Figure 4.2, are similar for either MCMC method.

Because the marginal likelihood and state simulation methods generate such similar results for the RSLN-2 model, it was decided to use the marginal likelihood method for the Bayesian estimation of the other popular long-term equity models. The reason for this is that the marginal likelihood method is computationally less taxing due to not simulating underlying regime paths.

78
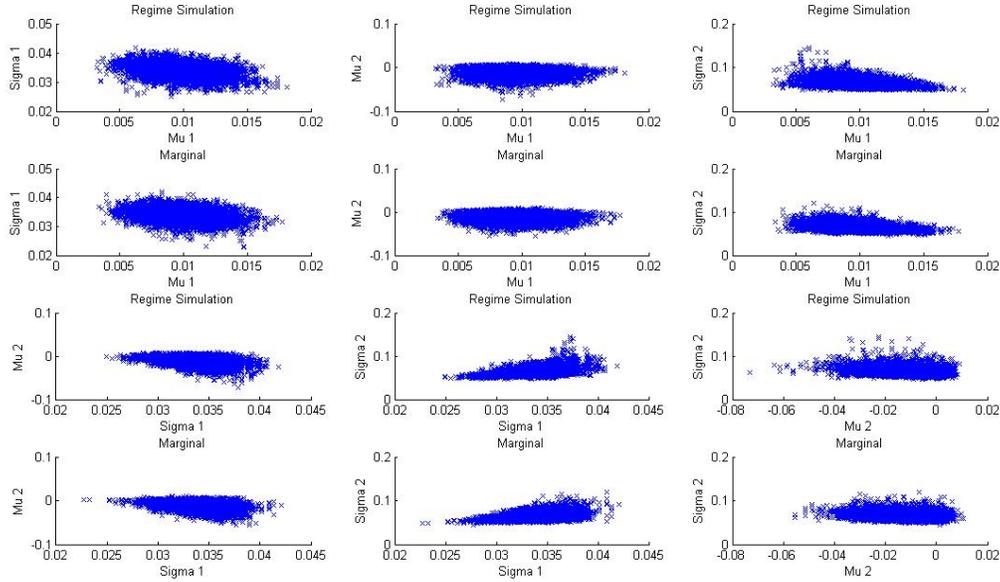
Figure 4.2: Relationships of the regime-specific RSLN-2 parameters under Regime-Switching Bayesian Estimation

The posterior means under both methods are compared with the maximum likelihood estimates in Table 4.1. The Bayesian estimates of the transition probability parameters are higher in both cases, meaning the Bayesian estimation of the model experiences more state transitions. Under the Bayesian estimates, the model spends slightly more relative time in state two. The state one parameters are quite similar across all three estimates. The state two variance is higher under the Bayesian approach.

These factors together generate thicker tails for the longer term accumulation factors for the

Method	$p_{1,2}$	$p_{2,1}$	μ_1	σ_1	μ_2	σ_2
Marginal Likelihood	0.0483	0.2133	0.0102	0.0336	-0.0147	0.0656
(Standard Deviations)	*0.0177*	*0.0812*	*0.0017*	*0.0017*	*0.0086*	*0.0067*
Regime Simulation	0.0489	0.2147	0.0104	0.0336	-0.0149	0.0658
(Standard Deviations)	*0.0171*	*0.0835*	*0.0017*	*0.0016*	*0.0085*	*0.0064*
MLE	0.0475	0.2017	0.00990	0.03412	-0.01286	0.06353

Table 4.1: The Posterior Distribution Means of the RSLN-2 Parameters Fitted to the S&P 500 Monthly Log-Return from Jan 1950 to Oct 2010 under the Bayesian Approach
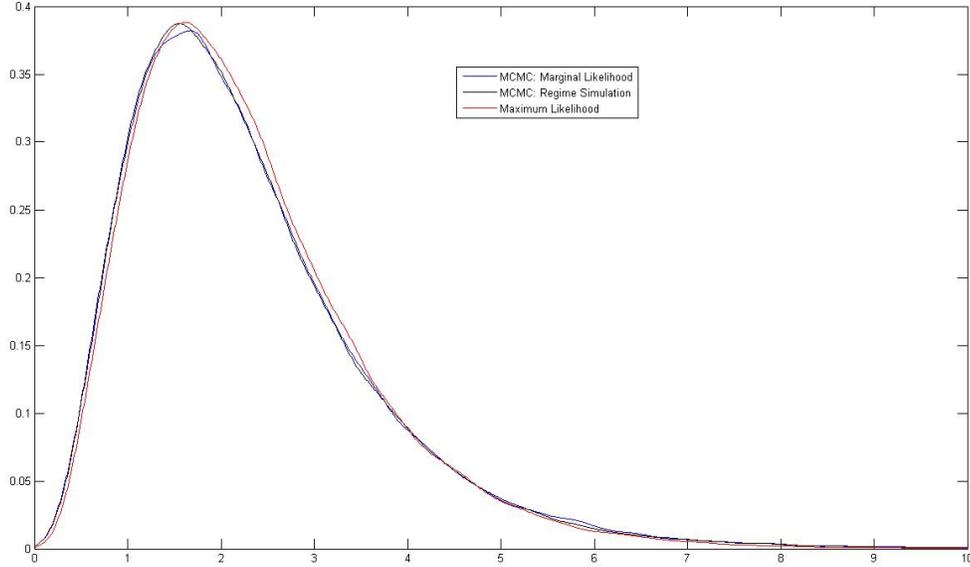
Figure 4.3: Simulated 10-Year Accumulation Factor Distributions of the RSLN-2 Model under Different Model Estimations

RSLN-2 model. The 10-year accumulation factors are shown in Figure 4.3.

4.4.2 RSDD2

The Bayesian estimation of the RSDD-2 model is very similar to the RSLN-2 model, with the addition of the drawdown parameters D_1 and D_2. The prior and candidate distributions for the transition probabilities were the same as those used for the RSLN-2 estimation, as were the state one parameters μ_1 and σ_1. The prior distributions for μ_2 and σ_2 were widened a little. The D parameters were estimated using the same inverse gamma approach as is done for the σ^2's.

The RSDD-2 parameter posterior distributions are displayed in Figure 4.4. For the parameters that are also included in the RSLN-2 model, their respective posteriors from the RSDD-2 model are quite similar. The D parameters, especially D_2, have larger uncertainty relative to the state means μ_1 and μ_2, and the D_2 posterior has a thicker left tail than a right one.
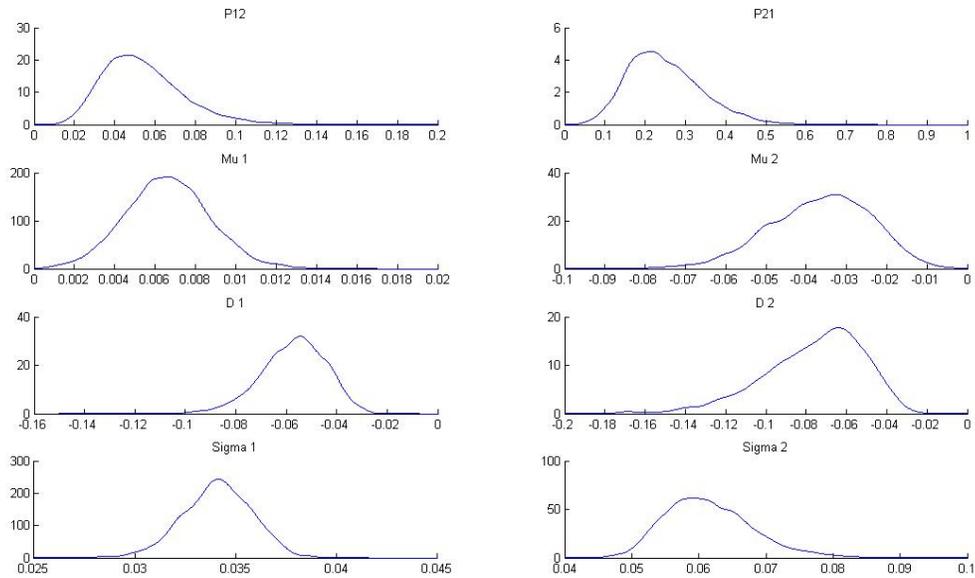
Figure 4.4: Distributions of the RSDD-2 parameters under Bayesian Estimation

The bivariate parameter plots for the RSDD-2 model are displayed in Figure 4.5. There is an positive relationship between D_2 and μ_2, indicating that the more negative the mean, the more powerful the mean-reversion effect.

Relative to the maximum likelihood estimate, the probability of transitioning from state one to state two on average under Bayesian estimation is twice as high (Table 4.2). There is also a larger probability assigned to the opposite transition. For the other parameters, the Bayesian posterior means and the ML estimates are relatively close together.

The inclusion of paramater uncertainty via Bayesian estimation did not have as profound effect on the long-term left tail of the RSDD-2 model as it did for the RSLN-2 model. As shown in Figure 4.6, the left tails for the 10-year accumulation factors from the maximum likelihood and Bayesian estimates of the model are very close together. The relationship between D_2 and μ_2 is clearly playing a role.
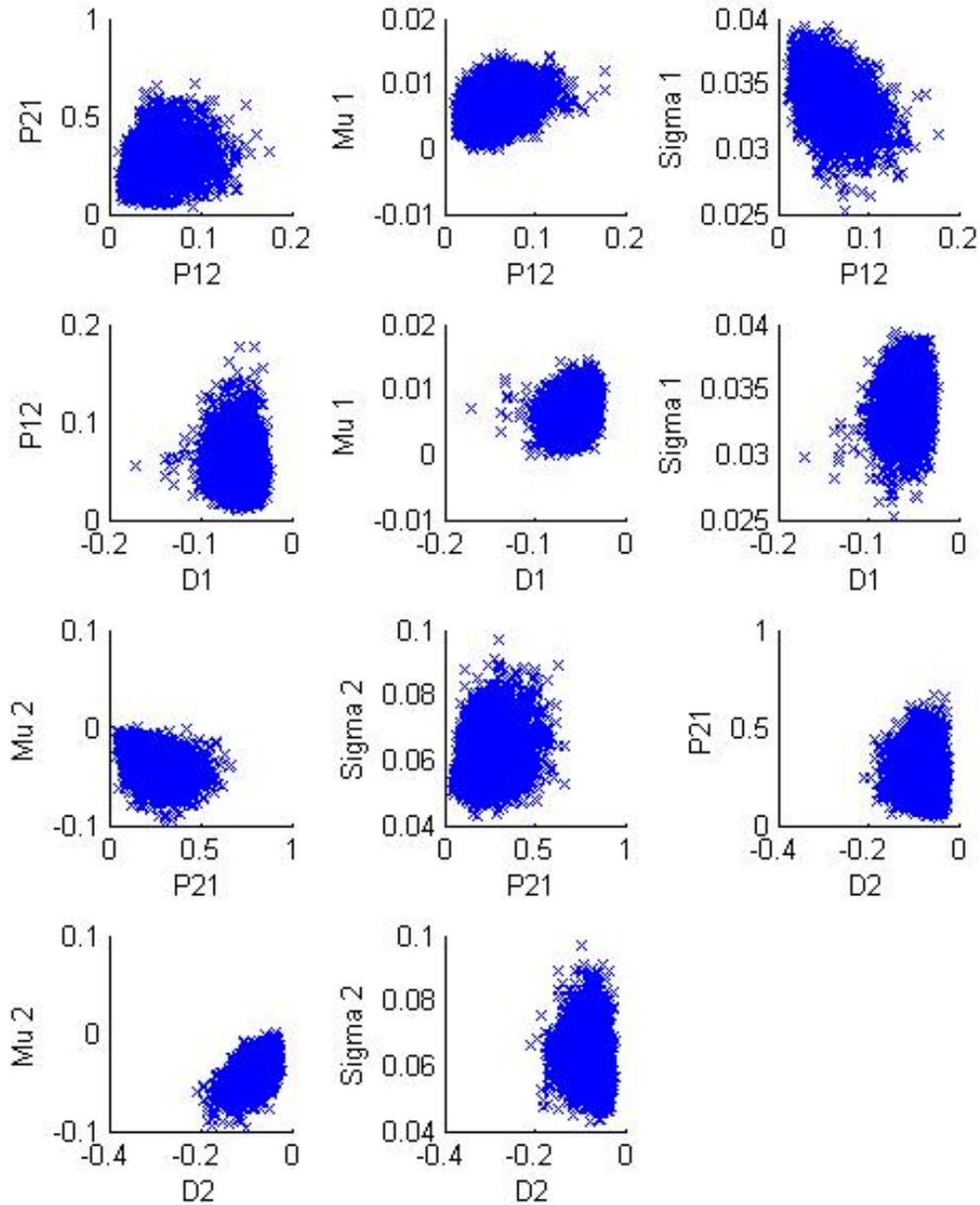
Figure 4.5: Relationships of the RSDD-2 parameters under Bayesian Estimation

Methd	Bayesian	S.D. (Bayesian)	MLE
$p_{1,2}$	0.0408	0.0142	0.0284
$p_{2,1}$	0.2296	0.0755	0.1870
μ_1	0.0064	0.0018	0.0068
φ_1	-0.0544	0.0107	-0.0429
σ_1	0.0341	0.0013	0.0352
μ_2	-0.0427	0.0125	-0.0488
φ_2	-0.0717	0.0231	-0.0997
σ_2	0.0644	0.0059	0.0623

Table 4.2: The Posterior Distribution Means of the RSDD-2 Parameters Fitted to the S&P 500 Monthly Log-Return from Jan 1950 to Oct 2010 under Bayesian estimation
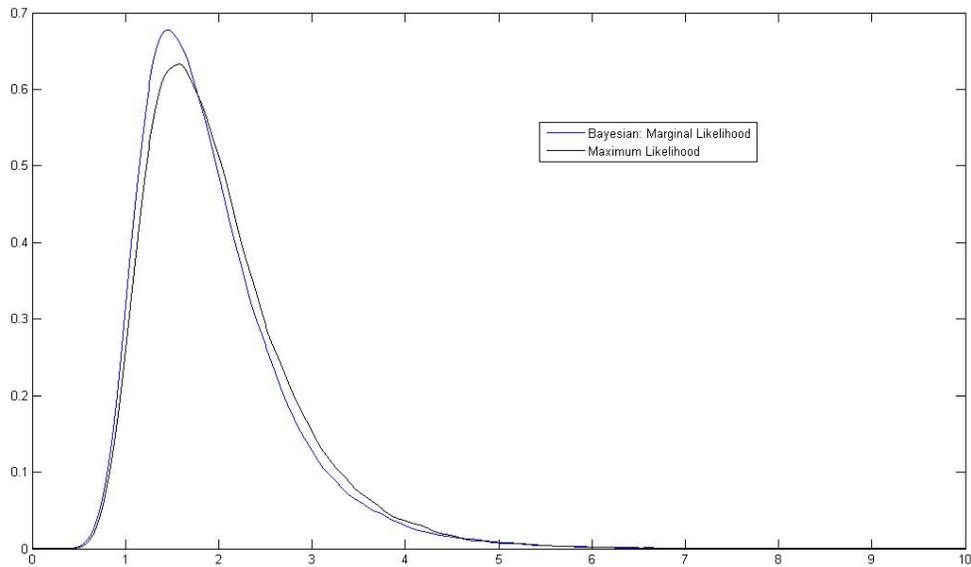
Figure 4.6: Simulated 10-Year Accumulation Factor Distributions of the RSDD-2 Model under Different Model Estimations

4.4.3 RSGARCH

The RSGARCH model fitted using maximum likelihood estimation in Chapter 2 has ten parameters: μ_1, μ_2, $\alpha_{1,0}$, $\alpha_{2,0}$, $\alpha_{1,1}$, $\alpha_{2,1}$, β_1, β_2, and the transition probabilities $p_{1,2}$ and $p_{2,1}$. However, under the maximum likelihood fitting, all α parameters were found to be very close to zero. A very small move in any of these parameters reduces the likelihood by very large amounts. This suggests that great care in design will be needed for the MCMC algorithm to work effectively. However, even with very tight priors for these parameters, simulations could not produce acceptance ratios higher than 5%. So, for the final Bayesian estimation of the RSGARCH model, we only considered the six other parameters.

An interesting result from the MLE fitting of the RSGARCH model in Chapter 2 was that state two was very low frequency ($p_{2,1}$ was estimated at effectively 1). However, lowering the value of $p_{2,1}$ does not significantly impact the likelihood. This fact, together with the fact that state two is low frequency, suggests that the state two parameters should be given prior and candidate distributions that incorporate this uncertainty, as very little was learned about them through maximum likelihood estimation.

For prior distributions, $p_{1,2}$ was assigned a Beta(4,60) distribution, and for $p_{2,1}$ a Beta(3,2) distribution was used. For μ_1, a $N(0.01, 0.02)$ was used for a prior distribution, and a $N(\mu_1^{(r-1)}, 0.003)$ distribution was used to simulate candidate parameters. For β_1, a Beta(20,4) prior was defined and a Beta($24 \cdot \beta_1^{(r)}, 24 \cdot (1 - \beta_1^{(r)})$) was used as its candidate distribution. The state-two parameters were allowed to range much more broadly, given the relatively uncertainty about those. $N(-0.08, 0.08)$ and $N(\mu_2^{(r-1)}, 0.02)$ were used as the prior and candidate distributions for μ_2 respectively, and Beta(4,4) and Beta($5 \cdot \beta_2^{(r)}, 5 \cdot (1 - \beta_2^{(r)})$) were defined as the prior and candidate distributions for β_2.

The posterior distributions for the six parameters are displayed in Figure 4.7, and the bivariate scatterplots are shown in Figure 4.8. There is much uncertainty exhibited by the results for the transition probability parameter $p_{2,1}$, as well as the parameter β_2. The posterior for the β_2 parameter was difficult to generate, as acceptance rates were high.

Interestingly, however, there was no discernible relationship between these two parameters in the bivariate scatterplot. There was, as expected, a relationship between μ_2 and the
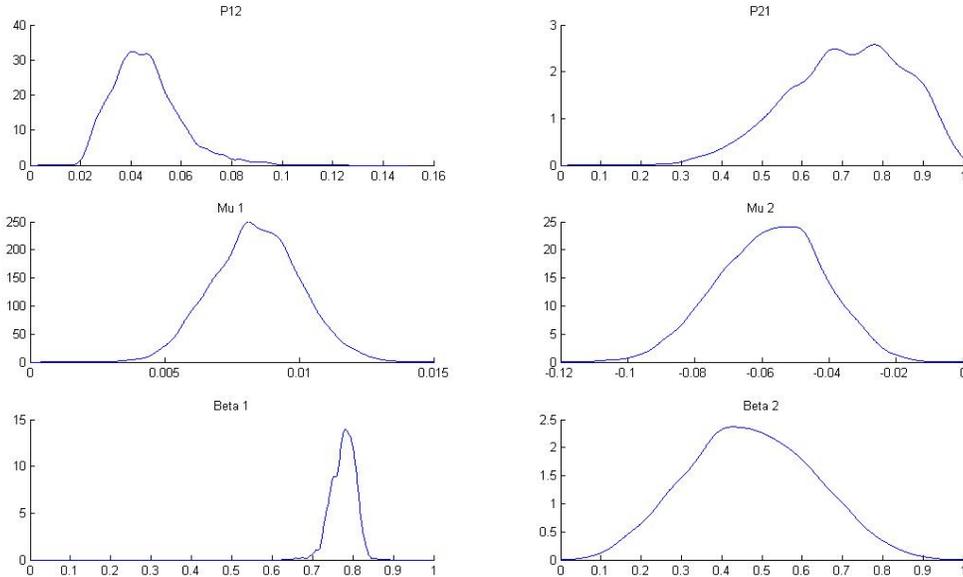
Figure 4.7: Distributions of the RSGARCH parameters under Bayesian Estimation

transition probability parameters. The more likely the process is to spend time in the second state, the higher its mean return, counterbalancing the parts of the distribution from which negative returns can come.

When comparing the Bayesian estimates to the MLE estimates, the differences again centre around the second state. The maximum likelihood estimates and the means of the Bayesian posterior distributions for each of the parameters is displayed in Table 4.3. The relative uncertainty about $p_{2,1}$ and β_2 meant that the ML estimates and the posterior means were significantly different.

The impact of parameter uncertainty is quite profound for the RSGARCH model when comparing the 10-year accumulation factor distributions under both the Bayesian and maximum likelihood model estimations (Figure 4.9). The Bayesian approach yielded much thicker model tails, providing evidence that not accounting for parameter uncertainty can lead to underestimation of the long-term left tail - the more dangerous side for investment guarantee contracts.
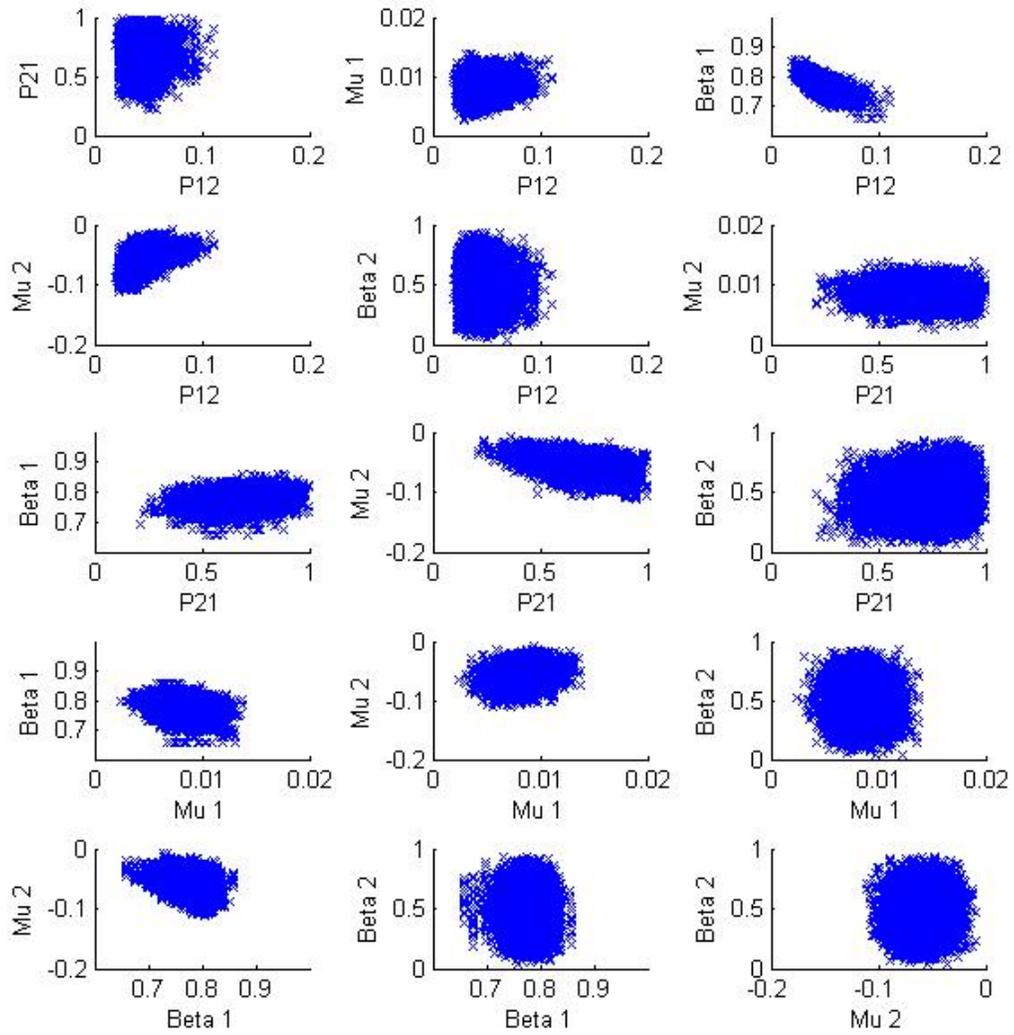
Figure 4.8: Relationships of the RSGARCH parameters under Bayesian Estimation

Parameter	Bayesian	*S.D. (Bayesian)*	MLE
$p_{1,2}$	0.0458	0.0103	0.0331
$p_{2,1}$	0.7183	0.1400	1.0000
μ_1	0.0083	0.0014	0.0079
β_1	0.7776	0.0251	0.8091
μ_2	-0.0582	0.0158	-0.0870
β_2	0.4535	0.1569	0.2939

Table 4.3: The Posterior Distribution Means of the RSGARCH Parameters Fitted to the S&P 500 Monthly Log-Return from Jan 1956 to Sep 2004 under the Bayesian Approach compared to the ML Estimates
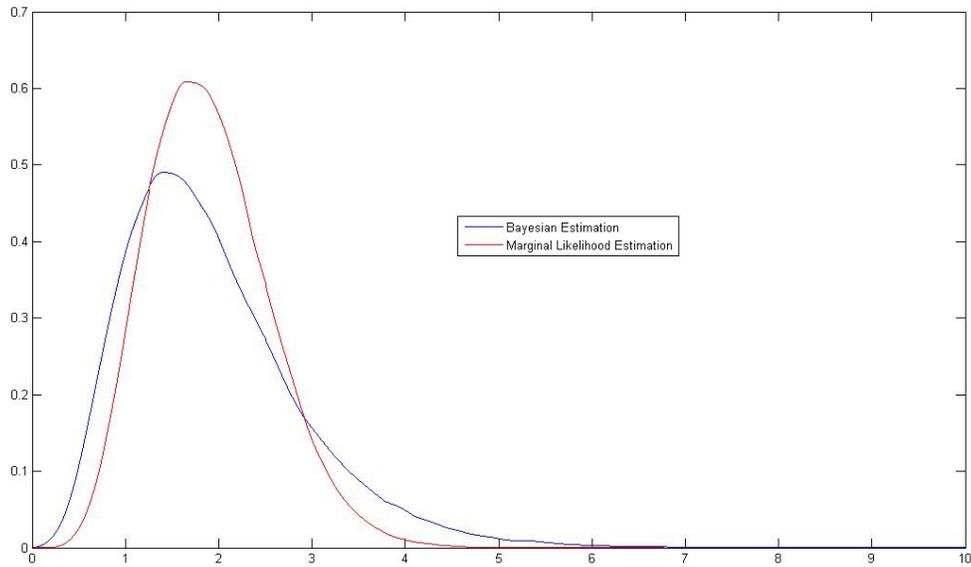
Figure 4.9: Simulated 10-Year Accumulation Factor Distributions of the RSGARCH Model under Different Model Estimations

4.4.4   MARCH

The MARCH model from Chapter 2 when fitted using maximum likelihood estimation was estimated to have 7 significant parameters: the means from the two states, μ_1 and μ_2, the three alpha parameters from the high mean state $\alpha_{1,1}$, $\alpha_{1,2}$ and $\alpha_{1,3}$, the zero-lag alpha parameter from the second state, $\alpha_{2,0}$, and the mixing weight P. As was the case for the RSGARCH model, the second state was visited much less often for the MARCH model than the RSLN or RSDD models, and rarely does the model spend consecutive periods in state 2 given the P estimate of 96.7%. Similar to the RSGARCH case, significant parameter uncertainty pertaining to the state 2 parameters is expected from the Bayesian fitting.

The mixing weight P was assigned a Beta(48,4) prior distribution, and its candidate distribution also a Beta distribution that was tightened somewhat and centered on the current iteration's value of the parameter. The means from the two states were assigned normal priors and candidate distributions similar to those for the previous models. The state 1 α parameters were assigned identical $Beta(2, 16)$ prior distributions, which were wider prior distributions than others, and their candidate distributions were all $Beta(18 \cdot \alpha^{(r)}, 18 \cdot (1 - \alpha^{(r)}))$ distributions which are dependent on the previous iteration's value. Finally, $\alpha_{2,0}$ was assigned a $Beta(2, 24)$ prior distribution, and a corresponding centered Beta candidate distribution.

The posterior distributions for the seven MARCH parameters are displayed in Figure 4.10. The posterior distribution of P is almost entirely over (0.9,1), so the Bayesian results confirm state one to be very high frequency for the MARCH model. The posterior distributions of the state 1 alpha parameters were interesting. The first and second lag parameters, $\alpha_{1,1}$ and $\alpha_{1,2}$ respectively, have identical posterior distributions, whereas the posterior distribution $\alpha_{1,3}$ was much further displaced from 0 and has a thicker right tail. This result suggests there is more of an effect from the 3-month deviation lag than for the first two. Finally, the posterior distribution of $\alpha_{2,0}$ indicated that the parameter had an effective range of (0.05,1), and was relatively centered over that range.

The bivariate scatter plots of select MARCH parameter are shown in Figure 4.11. Of interest, the state 1 α parameters have counter-balancing relationships. When one parameter has a relatively large value, the other two tend to scale back, resulting in the overall ARCH effect
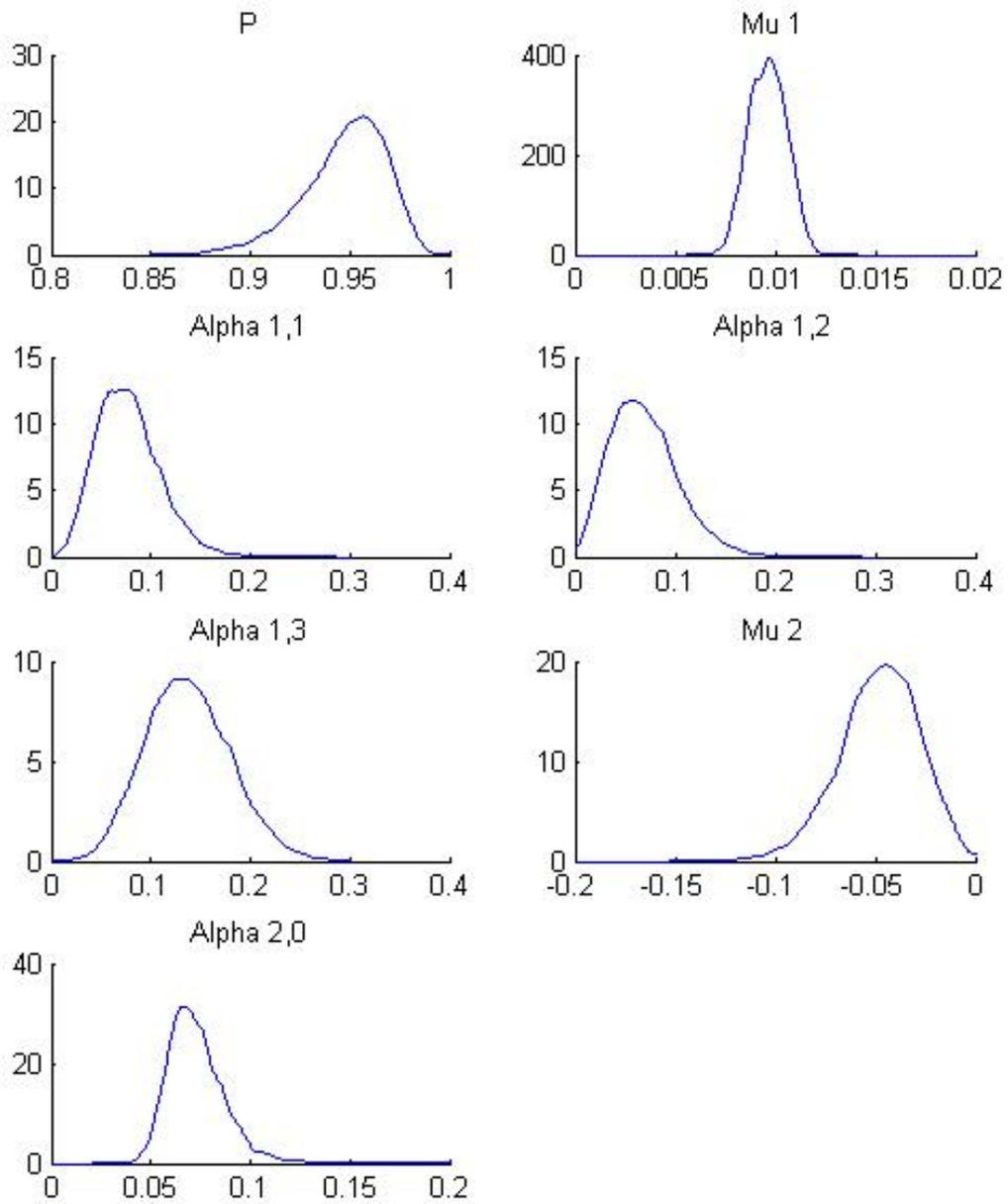
Figure 4.10: Distributions of the MARCH parameters under Bayesian Estimation
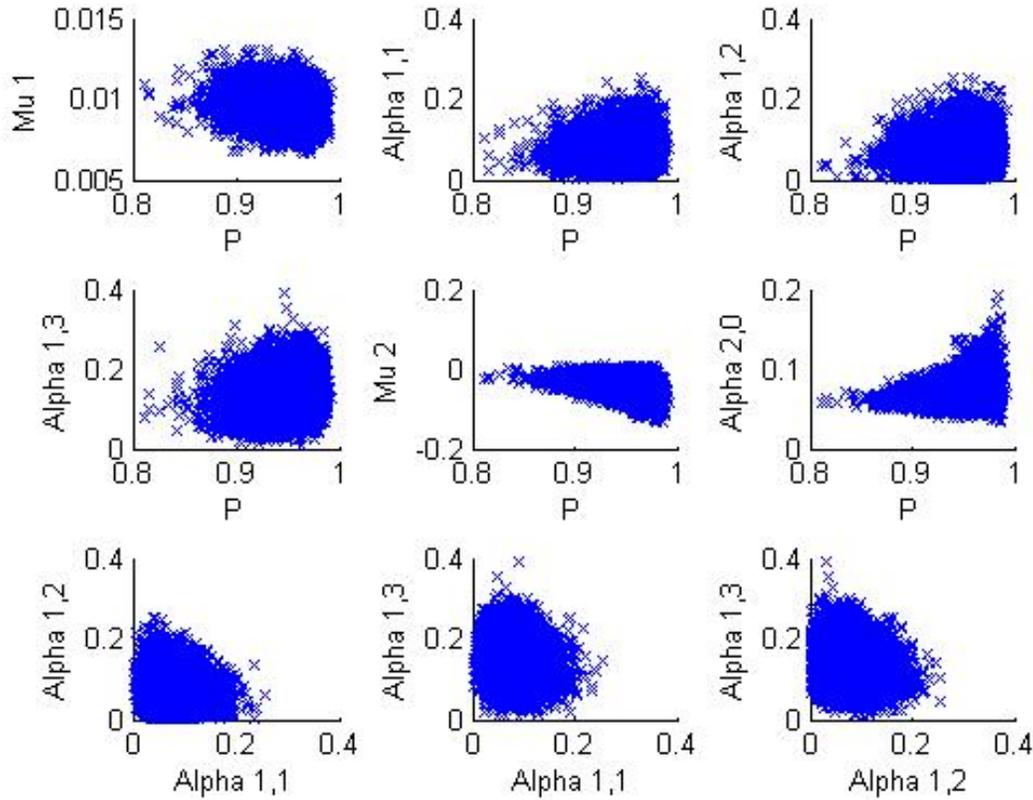
Figure 4.11: Relationships of the MARCH parameters under Bayesian Estimation

from previous data observations deviation from the mean remaining somewhat constant. Also worth noting is the effect of the mixing parameter P on the state two parameters μ_2 and $\alpha_{2,0}$. As P increases, resulting in the second state being visited less often, the distribution of the second state becomes more negative and more volatile, indicating that Bayesian framework assigns the second state for the most volatile negative observations, but the uncertainty lies in how much of the left tail the state captures.

The maximum likelihood estimates and the mean of the posterior distributions under the Bayesian estimation of the MARCH parameters are displayed in Table 4.4. Compared to the other models, the MARCH parameters are rather close together under the Bayesian and MLE methods. Under the Bayesian approach, the model spends a little bit more time in state 2, but the state 2 mean is smaller in magnitude to compensate. The Bayesian method

Parameter	Bayesian	S.D. (Bayesian)	MLE
P	0.9475	0.0213	0.9667
μ_1	0.0096	0.0009	0.0095
$\alpha_{1,1}$	0.0772	0.0312	0.0714
$\alpha_{1,2}$	0.0696	0.0340	0.0570
$\alpha_{1,3}$	0.1377	0.0428	0.1636
μ_2	-0.0477	0.0205	-0.0659
$\alpha_{2,0}$	0.0730	0.0141	0.0700

Table 4.4: The Posterior Distribution Means of the MARCH Parameters Fitted to the S&P 500 Monthly Log-Return from Jan 1950 to Oct 2010 under the Bayesian Approach compared to the ML Estimates

also shifts some of the autoregressive effect to the 1st and 2nd lags, and away slightly from the third.

Similar to the RSGARCH model, the 10-year accumulation factor left tail of the MARCH model is significantly thicker under Bayesian estimation than under maximum likelihood (Figure 4.12). This is a consequence of the left tail of the posterior distribution of P being thicker than the right. The right tail of the 10-year accumulation factors for the MARCH model are similar under maximum likelihood and Bayesian estimation.

4.5 Tail Capturing Revisited

In Chapter 2, the fitted hidden Markov models were compared by assessing their ability to capture the more extreme negative return periods of the S&P 500. In particular, the probability of experiencing the 37% drop in the S&P 500 from January 1999 to January 2009. Table 4.5 displays these probabilities again for the maximum likelihood fitted models, but also includes the probabilities of experiencing the crisis under the Bayesian model estimates. Figure 4.13 shows the distributions of the 10-year accumulation factors for the models under Bayesian estimation.

For the RSLN-2, RSGARCH and MARCH models, there was a significant increase in the models' capacity to capture the experience when fitted using Bayesian estimation over maximum likelihood. For the MARCH model, the probability of capturing the 10-year drop
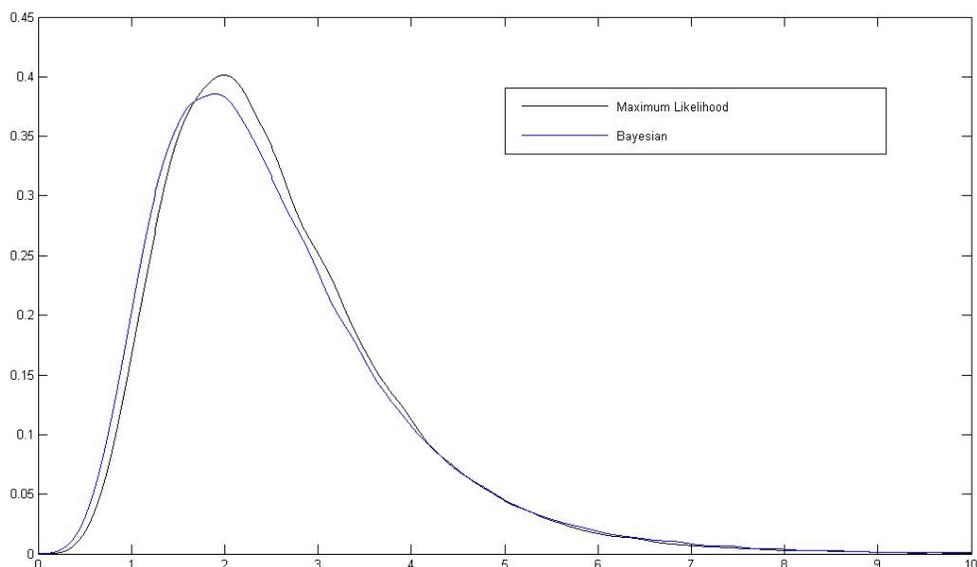
Figure 4.12: Simulated 10-Year Accumulation Factor Distributions of the MARCH Model under Different Model Estimations

Model	Maximum Likelihood	Bayesian
RSLN-2	0.0276	0.0328
RSDD-2	0.0016	0.0016
RSGARCH	0.0182	0.0283
MARCH	0.0042	0.0075

Table 4.5: Probability of experiencing the 1999-2009 S&P Return under the Maximum Likelihood and Bayesian fitted Hidden Markov Models

nearly doubled. The crash was still most likely to be experienced under the RSLN-2 model, although the difference between the probability of capturing the drop under the RSLN-2 and RSGARCH models is noticeably smaller.

For the RSDD-2 mode, however, the probability remained unchanged. The counterbalancing effect seen in the parameter bivariate distributions displayed in Figure 4.5 meant that long-term negative returns are just as unlikely under the model when estimated under a Bayesian framework as they were for maximum likelihood. The model's mechanics are not conducive to capturing the deeper left tails of the other models.
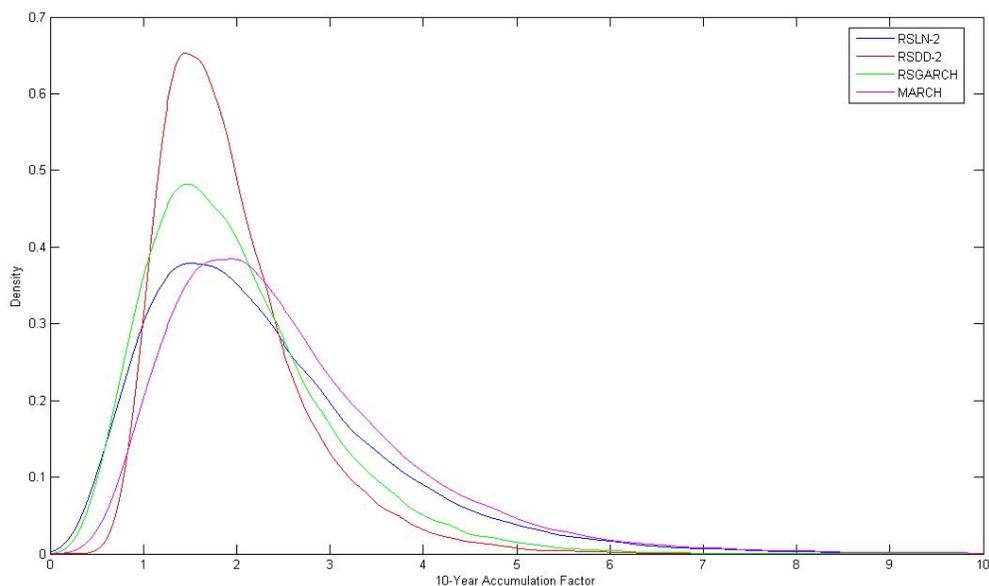
Figure 4.13: Density Functions of the 10-Year Accumulation Factors of the Popular Hidden Markov Models Under Bayesian Estimation

4.6 Conclusion

In this Chapter, the hidden Markov long-term equity models from Chapter 2 were estimated using a Bayesian approach. There were a few notable results from the analysis performed:

- For the RSLN-2 and the S&P 500, there was no observable difference in the obtained posterior distributions for the parameters between the marginal likelihood and state simulation methods under Metropolis-Hastings MCMC.

- The Bayesian predictive distributions of all of the hidden Markov models in this thesis indicated thicker left tails that the respective maximum likelihood distributions, with the sole exception being the RSDD-2 model. The posterior paramater distributions also indicated that uncertainty was more prevalent in the parameters associated with the left tail of the log-return distribution.

- Incorporating the parameter uncertainty with Bayesian estimation did not override

93

model mechanics however. For the RSDD-2 model, the Bayesian analysis indicated the long-term left tail of the log-returns was similar to the corresponding tail under maximum likelihood estimation. This meant a further deviation of the RSDD-2 model long-term returns from the other hidden Markov models.

The methodology from this Chapter will also be used to investigate the impact of parameter uncertainty pertaining to hidden Markov portfolio optimization in Chapter 5, and portfolio replication in Chapter 6.

Chapter 5

Hidden Markov Portfolio Optimization

5.1 Introduction

The optimal portfolio is one that meets a target expected return with minimum risk. Portfolio optimization in finance typically uses a measure of portfolio variation as its definition of risk, such as portfolio variance or portfolio value at risk. Risk can also be indirectly minimized through use of a utility function. A portfolio that maximizes an investor's expected utility is considered optimal under such a design (See Panjer et al. (1998)).

Markowitz (1952, 1959) used a single-period approach that provided a foundation for the current state of optimization research in finance. The Markowitz model uses variance as the risk criterion to be minimized, while at the same time aiming to maximize terminal wealth, which is indirectly fixed through specification of a risk tolerance level.

Alexander, Coleman and Li (2006) explored the use of value at risk (VaR) and conditional tail expectation (CTE) as the measure of risk to be minimized. These risk measures can be more attractive than variance in some cases because the measure is only directly dependent on one tail of the portfolio performance distribution. They found the CTE to be preferable

to VaR due to it being a coherent risk measure in the sense of Artzner et. al (1999), and because they found minimizing portfolio CTE typically resulted in minimizing portfolio VaR anyway.

A hidden Markov setting offers a challenge to portfolio optimization in finance: how best to handle the unobserved nature of the underlying state process. All of the above measures of risk can be derived unconditionally from a hidden Markov distribution (for the expected utility case, see Boyle and Liew (2008)). The regime switching framework can be conditioned away to yield the stationary distribution and its corresponding measures of risk, and portfolio optimization can be performed as it would for any single state distribution of returns. Such an approach does not take advantage of the information contained in the data about the plausible underlying state paths under the model. More informed optimization processes, when using the hidden Markov structure for equity returns, would dynamically incorporate all information from which one could obtain inference about the underlying state process: that contained in both the model and the data.

The literature pertaining to hidden Markov portfolio optimization has focused on a few areas. Yin and Zhou (2004) looked at the topic from a continuous time perspective. They showed that if the current underlying state of the market were known, then a portfolio that was mean-variance optimized according to the mean and variance of the current regime is asymptotically optimal as the length of the period in a discrete time setting approaches zero. The paper looked at the problem from a solely theoretical perspective.

Ang & Bekaert (2004) investigated the problem of hidden Markov optimization for a selection of international markets and funds. Their approach to the unobserved regime problem was an approach similar to the indicator method described in Chapter 3. Theirs was a two-regime model, and if the estimated probability of the process being in regime 1 for the **next** stock return realization (based on the history of observed stock returns and the transition probabilities) was greater than 0.5, the portfolio would be mean-variance optimized according to the regime 1 parameters. If that probability were less than 0.5, the portfolio would be optimized according to the regime 2 parameters. They found their regime-switching strategy dominated non-regime switching strategies in out-of-sample tests.

Boyle & Liew (2008) investigated the use of hidden Markov portfolio optimization for ana-

lyzing hedge-fund based portfolios. They optimized over the unconditional stationary distribution of the market (in this case, the hedge fund and the S&P 500) to demonstrate the relative volatility of portfolios comprised of various hedge funds.

This chapter will explore optimization strategies using different approaches to including the information about the regime switching process. The objectives will be to:

1. Determine if a particular method is favored for hidden Markov optimization over the S&P 500 and NASDAQ indexes when the deep left tail of the investment portfolio is the primary risk measure, and

2. Determine the robustness of the different methods when parameter uncertainty is taken into account.

5.2   Multivariate Long-term Equity Modeling

The relationship that financial indexes have with each other can be both dynamic and complex. Stocks in a particular index may be significantly financially linked with other stocks in a different index, due, for instance, to the complex ownership and stakeholder arrangements in many of the world's large financial institutions. Good performance by a single bank can have positive results on many of the world's indexes. Likewise, in periods of economic crisis, such as October of 1987 and the 2007-2010 crisis, virtually all financial markets were in retreat and thus returns across different indexes were very highly correlated with each other. The positively correlated relationship does not always exist, however, as events in one nation may shape an index specific to that nation, but other indexes not directly related may be less affected or perhaps completely unaffected. It becomes important to capture as best possible these complex relationships when fitting joint models for the world's indexes when analyzing the risk of instruments based on these indexes.

It has been argued (Hamilton, 1989) that the rationale behind the use of hidden Markov models for equity returns is that the market can be thought of as having different phases, related to the state of investor opinion in general. This was discussed for the choice of hidden

Markov models for single-variate equity returns, and the same line of reasoning can hold for the use of multi-variate hidden Markov models for modeling multiple equity returns. The hidden Markov process is still thought to represent the underlying state of the market as a whole, and the extension of single-variate hidden Markov models for equity returns to the multivariate case is quite natural.

Since the hidden Markov process is interpreted as representative of investor confidence and general market perception, the interpretation can move beyond that of overall market stability measurement. The hidden Markov structure can still be representative of worldwide index advances and retreats, yet can also model instances of index decoupling, where indexes move independently of each-other. As an example, a fit of a hidden Markov model could yield an underlying state for which the TSX index moved with relatively high volatility and at the same time the S&P moved with relative stability - such a state could be representative of uncertainty in Canada while steady investment holds in the United States.

The price paid for this flexibility of hidden Markov models for multivariate returns is that of scope: the number of variables of the model grows exponentially with the number of regimes, and generally the added model parameters have relatively more parameter uncertainty than do the base parameters of the smaller model.

5.2.1   The S&P 500 and the NASDAQ

In this chapter, a portfolio optimization example using the S&P 500 and NASDAQ indexes will be used to demonstrate some different hidden Markov optimization approaches and the relative impact of parameter uncertainty on those approaches. A first step towards such a goal, once the market of instruments over which an investor wishes to optimize is selected, is the fitting of the hidden Markov class of models to the index data. In this section, some brief observations about the two indexes will be made, and then a hidden Markov model will be fit to two indexes using both maximum likelihood and Bayesian methods.

Thirty eight years (1971 to 2009) of monthly index levels and log-returns of the S&P 500 and NASDAQ indexes are displayed in Figure 5.1. A few qualities of the two indexes and their interaction with each other over time can be directly observed from these graphs. The
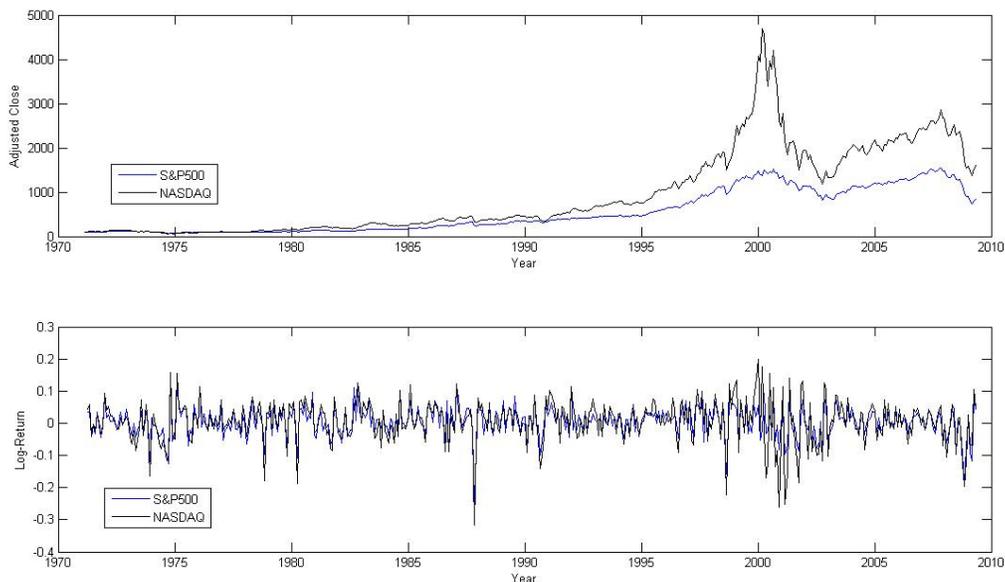
Figure 5.1: Monthly Closes and Log-Returns of the S&P 500 and NASDAQ Indexes

first is that the two indexes are very highly correlated with each other. Especially visible from the log-return plot, this is true both when the indexes are experiencing steady-return periods and when both indexes are experiencing more volatile return periods.

A second observation is that in periods of relative uncertainty, the NASDAQ's returns are much more volatile than those of the S&P 500. The two indexes were valued at almost the same level in 1990, but the NASDAQ climbed to three times the value of the S&P by the turn of the millennium. Successful models of the joint indexes will need to be reflective of this observation.

One potential question in the use of hidden Markov models to capture the indexes is whether or not there is enough support in the data for additional regimes beyond the two traditional regimes of one state of low volatility and steady index returns, and the other of high volatility and on average negative index returns. Additional market states are not outright apparent from the index return plots. An investigation of the value and data support for additional states is warranted during the model fitting stage.

99

5.2.2 The Bivariate Regime-Switching Lognormal Model

Akin to the single-variate RSLN model, the bivariate regime-switching log-normal (BRSLN) model uses a very simple modeling approach within the underlying states. For the two index case, the equity log-returns are jointly modeled using a bivariate Gaussian distribution; one defined for each underlying state. The bivariate normal distribution has five parameters: the two series' means, the two variances and the correlation coefficient parameter. Including the transition probability parameters, the BRSLN-K model has $5K + K(K-1) = K^2 + 4K$ parameters, where K is the number of underlying states in the model. The number of parameters grows quite quickly with the number of regimes.

Mathematically, under the BRSLN model, the joint distribution of X_t and Y_t, the log-returns of the indexes of interest, given underlying state ρ_t at time t, can be defined by:

$$
\begin{bmatrix} X_t \\ Y_t \end{bmatrix} \bigg| \rho_t \sim MVN \left(\begin{bmatrix} \mu_{\rho_t}^X \\ \mu_{\rho_t}^Y \end{bmatrix}, \begin{bmatrix} (\sigma_{\rho_t}^X)^2 & \sigma_{\rho_t}^X \sigma_{\rho_t}^Y \tau_{\rho_t} \\ \sigma_{\rho_t}^X \sigma_{\rho_t}^Y \tau_{\rho_t} & (\sigma_{\rho_t}^Y)^2 \end{bmatrix} \right)
$$

where $\rho_t \in \{1, \ldots, K\}$ represents the unobserved underlying state of the process state at time t, and K is the number of of underlying states of the model. $\mu_1^X, \ldots, \mu_K^X$ and $\mu_1^Y, \ldots, \mu_K^Y$ represent the mean log-returns of indexes X and Y respectively, while $\sigma_1^X, \ldots, \sigma_K^X$ and $\sigma_1^Y, \ldots, \sigma_K^Y$ represent the standard deviations. The correlation coefficient parameters, $\tau_1, \ldots, \tau_K$, are defined on $[-1, 1]$, with $\tau = 1$ representing perfect positive correlation between the two indexes, and $\tau = 0$ representing independence of the two indexes.

This base regime-switching multivariate model can capture relation between the two indexes through two facets: directly through the correlation coefficient parameters $\tau_1, \ldots, \tau_K$, and indirectly through the hidden Markov framework as described at the beginning of this section. As with all hidden Markov models, the underlying states and their respective bi-variate normal distributions resulting from the fit are more likely to be acceptable if they make intuitive sense.

Regime 1		Regime 2	
S&P 500	NASDAQ	S&P500	NASDAQ
$\mu = 0.00976$	$\mu = 0.01277$	$\mu = -0.02035$	$\mu = -0.02727$
$\sigma = 0.03516$	$\sigma = 0.04474$	$\sigma = 0.07210$	$\sigma = 0.11561$
$\tau = 0.85704$		$\tau = 0.84649$	
$p_{1,2} = 0.03039$		$p_{2,1} = 0.15080$	

Table 5.1: Maximum Likelihood BRSLN-2 Parameters of the S&P 500 and NASDAQ

5.2.3 Maximum Likelihood Estimation of the S&P 500 and NAS-DAQ

The S&P 500 and NASDAQ indexes were first estimated with a BRSLN-2 model using maximum likelihood estimation. The likelihood is as easily defined here as it was for the RSLN model in the single-variate case, with the multivariate normal probability distribution function taking the place of the single-variate normal pdf. The Generalized Reduced Gradient optimization algorithm, part of the Microsoft Excel package, performed the optimization quite quickly, and obtained the same optimal parameter set with the use of several different sets of starting parameter vectors.

The MLE BRSLN-2 parameter set for the S&P 500 and NASDAQ indexes are listed in Table 5.1. The resulting underlying states are very similar to the single-variate case found in Chapter 2, and quite common for a hidden Markov model when fitted to long-term equity data. One state is very investor friendly: the two indexes have positive expected log-returns and relatively low volatility. The other state is quite the opposite, with both indexes having negative expected log-returns and relatively high volatility. The transition parameters are also standard for a two-state hidden Markov long-term equity model; the process spends about 83% of the time in the calm, profitable state.

There are two additional observations about the optimal parameter set worth noting. The first is that the NASDAQ is always the more volatile index of the two, which is consistent with the observation made from the historical log-return plot earlier. Under the first under-lying state, the higher volatility of the NASDAQ is compensated with a higher expected rate of return. However, under the more chaotic state, the higher volatility is coupled with a ex-

Regime 1		Regime 2		Regime 3	
S&P 500	NASDAQ	S&P500	NASDAQ	S&P500	NASDAQ
$\mu = 0.00969$	$\mu = 0.01270$	$\mu = -0.02063$	$\mu = -0.02772$	$\mu = 0.05603$	$\mu = 0.06785$
$\sigma = 0.03527$	$\sigma = 0.04482$	$\sigma = 0.07170$	$\sigma = 0.11534$	$\sigma = -$	$\sigma = -$
$\tau = 0.85618$		$\tau = 0.84671$		$\tau = 1$	
$p_{1,2} = 0.02596$ $p_{1,3} = 0.00257$		$p_{2,1} = 0.14126$ $p_{2,3} = 0.00000$		$p_{3,1} = 0.00000$ $p_{3,2} = 1.00000$	

Table 5.2: Maximum Likelihood BRSLN-3 Parameters of the S&P 500 and NASDAQ

pected return even worse than that of the S&P, providing a rather inhospitable environment to investors.

The second observation is that the correlation coefficients of the two indexes together are nearly identical under both underlying states. Hence, under the model, there is no decoupling effect between the two indexes when the market shifts to a more uncertain outlook. This result is consistent with Boudreault & Panneton (2009), who found there to be no significant evidence for differing covariance matrixes across regimes. Special attention will be paid to this point when fitting the model using a Bayesian approach.

A BRSLN-3 model was also fitted to the S&P 500 and NASDAQ indexes to see the effect and shape of the additional underlying state. The maximum likelihood estimates were much more difficult to arrive at using the Generalized Reduced Gradient optimizer in Microsoft Excel with the additional parameters (in this case there were 21 model parameters), and there were several different parameter sets with log-likelihoods very close to that of the found global maximum, even though the third regime for each of those sets would look drastically different from case to case.

The parameter set yielding the global maximum log-likelihood found is displayed in Table 5.2. While the first two regimes are very reminiscent of the two regimes from the 2-state model, the third regime, as evidenced by the zeroed volatility parameters for both indexes, captures a single outlying data observation. In this case, the point captured is a high month of returns that is experienced as the market transitions from the investor-friendly regime to the chaotic one.

In many of the local maxima found for these two data sets for the BRSLN-3 model, the third

regime acted to capture a single observation, and in all cases was the average tenure in the additional regime only one month. Regimes of this kind represent inefficient parameterization in the sense that the data observations that these regime capture can be captured with much fewer parameters than those needed to do under a regime switching model. For instance, a single jump parameter can capture the frequency of such an observation, as opposed to several transition parameters needed to capture the data point with the same frequency. As a result of the potential third regime being difficult to focus, the two regime model will be presented for Bayesian comparisons.

5.3 Bayesian Estimation of Bivariate Hidden Markov Models

Bayesian estimation of multivariate Hidden Markov models can be approached quite similarly to the single-variate case in Chapter 4. Since the underlying states are defined similarly, the estimation of the transition probability parameter posterior distributions can be programmed much the same way; the main difference is that the likelihood used must now be the joint likelihood of all of the modeled indexes in the market. Beta distributions remain suitable choices for prior distributions (and candidate distributions in the case of use of the Metropolis Hastings algorithm) of the transition parameters. Individual index mean and standard deviation posteriors can be again estimated quite well with respective normal and inverse gamma prior distributions.

The correlation coefficient parameters are the only new parameters to the estimation relative to that from Chapter 4. The correlation coefficient parameters are defined on $[-1, 1]$, and so prior distributions would normally need to be defined on that range. However, if there is sufficient evidence to believe that negative correlation between indexes is highly unlikely, then the prior distributions that are defined on $[0, 1]$, such as the beta distribution, can be used. Uniform prior distributions are also an option, which would represent a prior position of absolute uncertainty across the distribution range.

5.3.1 Bayesian Estimation of the BRSLN-2 Model to the S&P 500 and NASDAQ

For the two indexes used for demonstration purposes for this chapter, the BRSLN-2 model will be estimated using the Bayesian approach. The Metropolis-Hastings MCMC algorithm will be employed similarly to that done in Chapter 4 for the single variate case. Likewise, the marginal likelihood method of hidden Markov MCMC will be used as opposed to the regime simulation method, under the assumption that the result obtained in Chapter 4 will carry through (ie. that both the marginal likelihood and state simulation methods will perform comparably).

Relative restriction of candidate parameters, such as the log-return means in the first state always be higher than the log-return means in the second, that serve the purpose of preventing the hidden states from assuming each others positions and effectively creating identical states that cover the entire model will also be incorporated in the estimation. This restriction can be easily adapted to the joint S&P 500 and NASDAQ model, since both the log-return plots and maximum likelihood estimates of the model indicated that the indexes are highly correlated during periods of both index stability and index instability. State one will be the investor-friendly state, and state two the highly volatile state.

Since the level of correlation between the two indexes is evidenced to be quite high (from both the index level and log-return plots and from the results of the maximum likelihood estimates) beta prior distributions will be used for the correlation coefficient parameters initially. There was no evidence from the posterior distribution results for the correlation coefficient parameters that a distribution defined outside of $[0, 1]$ would be needed.

Several iterations of different prior and candidate distribution parameters were performed, with the objective of achieving suitable candidate parameter acceptance rates. The prior distributions for each of the parameters for the final MCMC simulation are displayed in Table 5.3. The candidate distributions were very close to the priors in most cases, excepting parameters for whom the acceptance rate was too low, for which the candidate distributions were a bit tighter. The acceptance rates for each of parameters were all between 20% and 40%.

Regime 1		Regime 2	
S&P 500	NASDAQ	S&P500	NASDAQ
$\mu \sim N(0.01, 0.02)$	$\mu \sim N(0.01, 0.02)$	$\mu \sim N(-0.02, 0.04)$	$\mu \sim N(-0.025, 0.04)$
$\sigma^{-2} \sim G(1.04, 833)$	$\sigma^{-2} \sim G(1.04, 833)$	$\sigma^{-2} \sim G(0.036, 8000)$	$\sigma^{-2} \sim G(0.036, 8000)$
$\tau \sim Beta(10, 1.7647)$		$\tau \sim Beta(10, 1.7647)$	
$p_{1,2} \sim Beta(2, 48)$		$p_{2,1} \sim Beta(2, 6)$	

Table 5.3: Parameter Prior Distributions for the Bayesian Estimation of the BRSLN-2 to the S&P 500 and NASDAQ

The posterior distributions for the BRSLN-2 model parameters are displayed in Figures 5.2 (transition probability parameters), 5.3 (state 1 parameters), 5.4 (state 2 parameters), and 5.5 (correlation coefficient parameters). There is clearly more uncertainty surrounding the NASDAQ parameters than those of the S&P, especially in the second more volatile state.

Also interesting is that the correlation coefficient parameter for state one has a very tight posterior, with feasible values only ranging from 0.8 to 0.9. This indicates that the certainty about this parameter is quite strong. This is somewhat less true for the state two correlation coefficient parameter, whose posterior tails are thicker, though still concentrated above 0.7. The argument that the two indexes are very highly correlated no matter the state of the market is supported by the Bayesian results.

5.4   Portfolio Optimization

Under the optimization problem, let T denote the time horizon of the investment to be made, and let $S_t \in \mathbf{R}^N$ be the vector of asset prices at time t, where the investor selects a portfolio from a universe of N instruments whose values at time t based on the asset prices are $V_t = [V_1(S_t) \ldots V_n(S_t)]'$. Finally, let $x_t \in \mathbf{R}^n$ denote the decision variable of the portfolio from time unit t to $t + 1$. The portfolio value at time t after reallocation is then $V_t' x_t$.

Two standard constraints are often enforced in optimization problems and will be implemented throughout this chapter. The first is a budget constraint, which serves to say that any investor has a limited amount of money to invest. B_t will denote the investor's budget at time t, $\quad t \in \{0, \ldots, T\}$ and is mathematically represented in the optimization problems
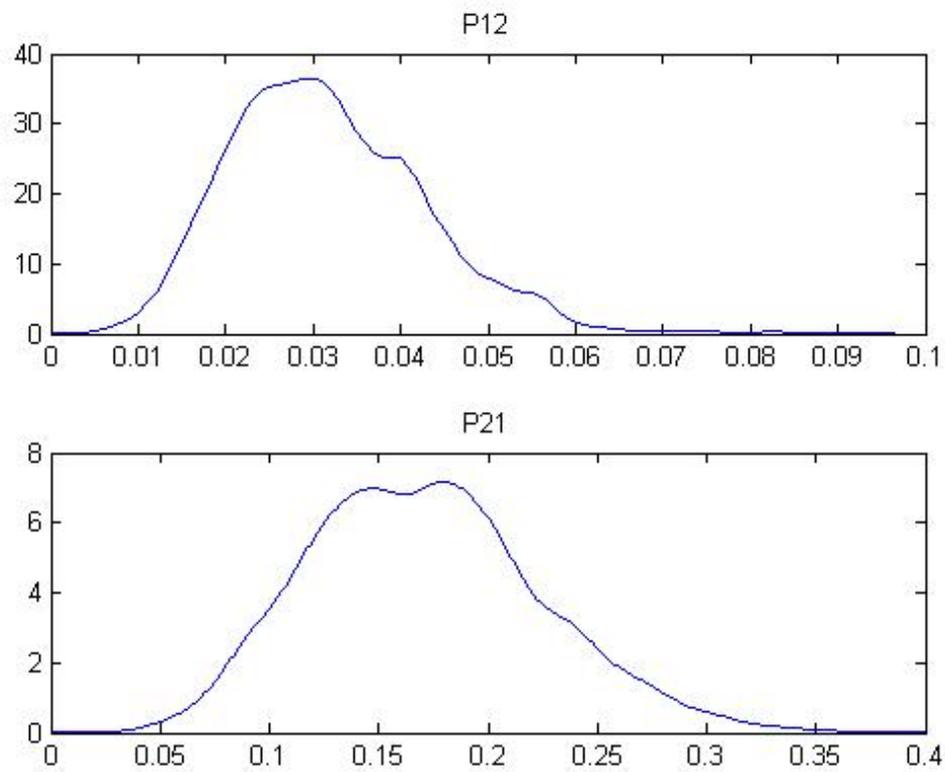
Figure 5.2: Transition Parameter Posterior Distributions for the Bayesian Estimation of the BRSLN-2 to the S&P 500 and NASDAQ
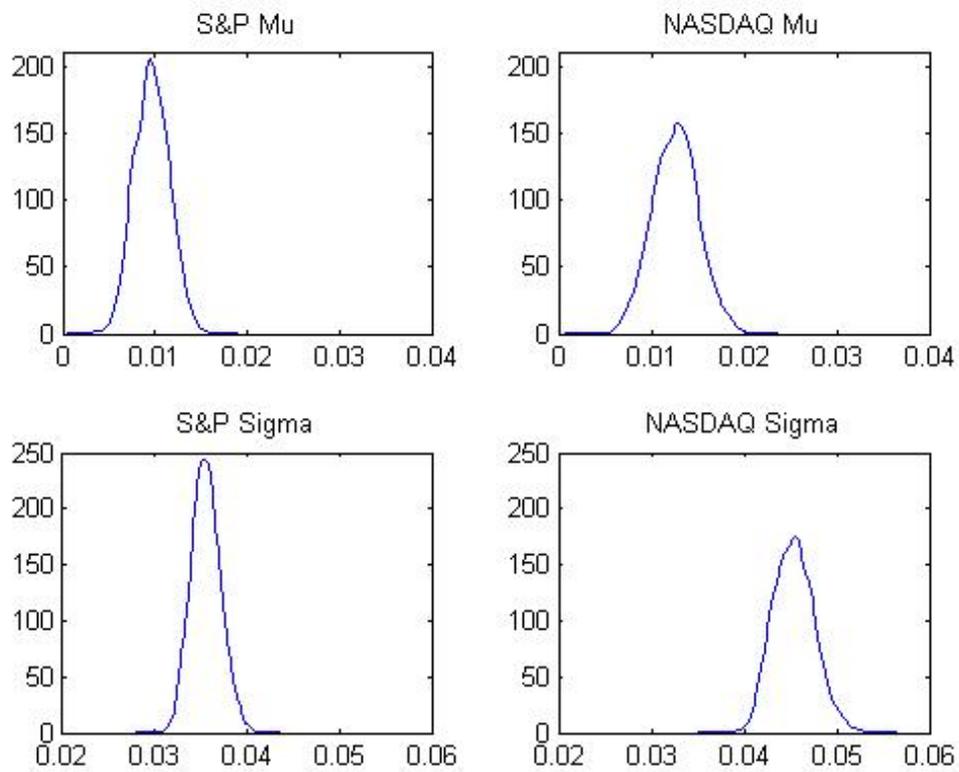
Figure 5.3: State 1 Posterior Distributions for the Bayesian Estimation of the BRSLN-2 to the S&P 500 and NASDAQ
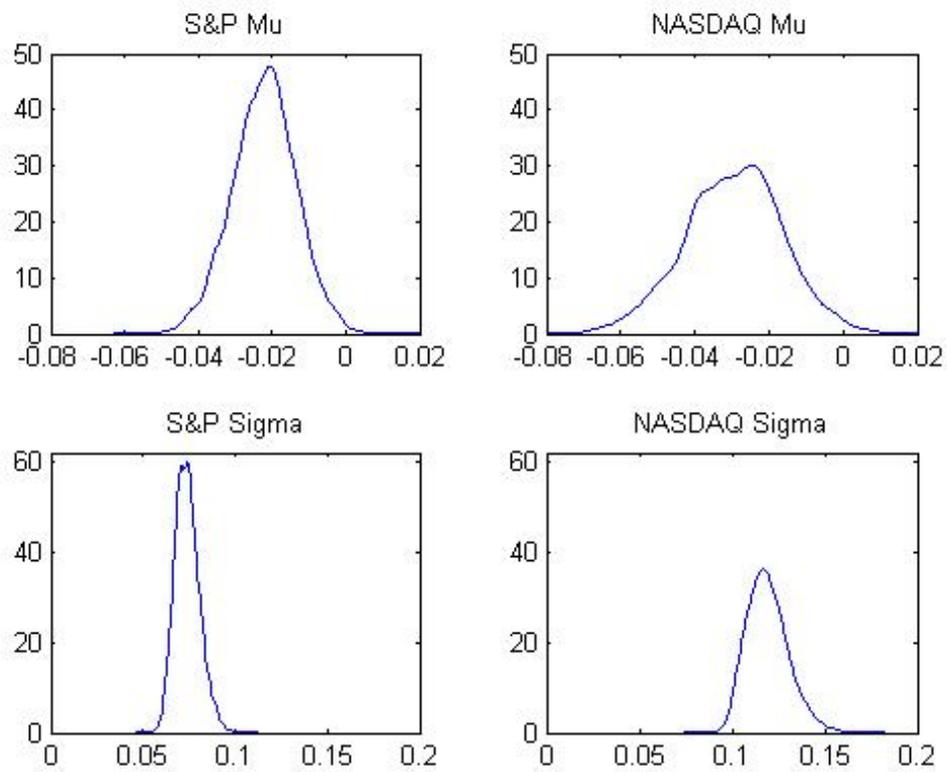
Figure 5.4: State 2 Posterior Distributions for the Bayesian Estimation of the BRSLN-2 to the S&P 500 and NASDAQ
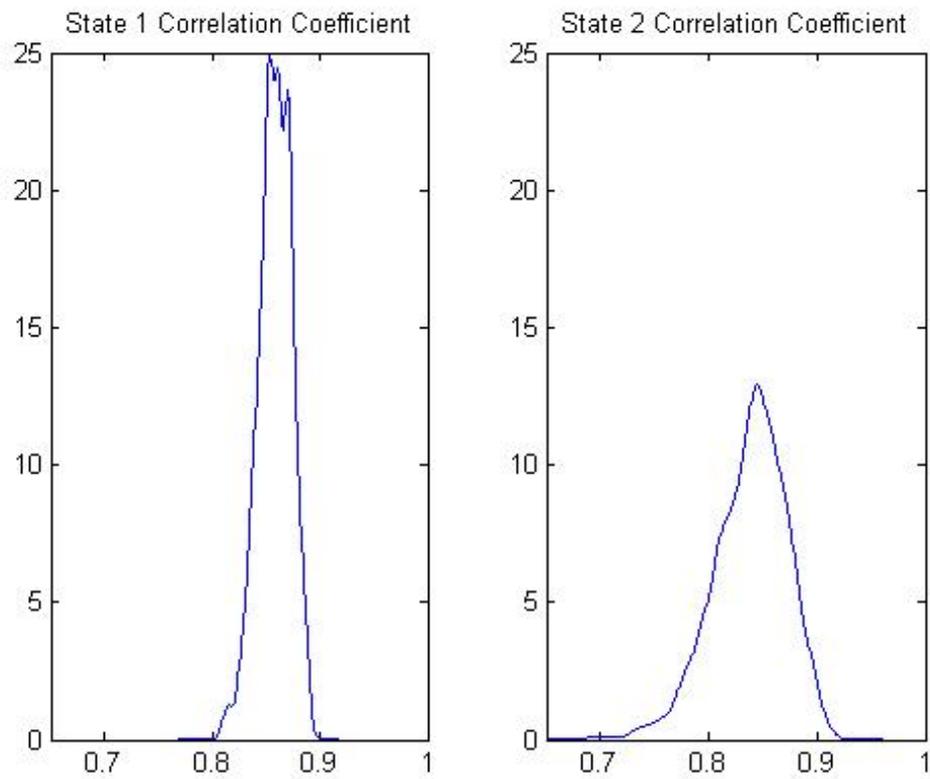
Figure 5.5: Correlation Coefficient Posterior Distributions for the Bayesian Estimation of the BRSLN-2 to the S&P 500 and NASDAQ

by the equation $V_t' x_t = B_t$. For all of the optimizations in this thesis, the level of B_t, so long as it is finite, will not change the portfolio weights.

The second constraint is a cap on short selling. This constraint prevents investors from borrowing an infinite amount of money against one market instrument to invest in another. The cap for each of the instruments can vary. Under the optimization problem, this is mathematically represented as $x_t \geq l_t$, where l_t is an n-dimensional vector of real numbers, typically less than or equal to zero, that represents the absolute minimum position an investor can take in each of the instruments at time t.

5.4.1 Mean-CTE Optimization

The risk measure that will be minimized for optimization here is the Conditional Tail Expectation (CTE) of the portfolio. The CTE is a measure of tail risk, and in that sense is consistent with the tail focus of the rest of this thesis. For a given confidence level β, the value at risk (VaR) of a portfolio's loss random variable is the loss value that is exceeded with probability $1 - \beta$. The CTE, for a continuous random variable, for that confidence level beta is the expected value of the loss given the loss is above the VaR.

Minimizing the CTE of portfolio loss has the attractive feature of only minimizing the risk of the tail of the distribution concerned with portfolio loss, if that is indeed what the investor is concerned about. Risk associated with the left tail of the loss random variable, the tail concerned with portfolio gains, is not (directly) minimized under this risk measure. This has potential benefit over minimizing a portfolio using a risk measure that has presence in both tails of the loss random variable, such as variance.

Mathematically, let $f(x_t, S_{t+1})$ denote the loss of the portfolio from time t to time $t+1$. The portfolio loss then satisfies:

$$f(x_t, S_{t+1}) = -(V_{t+1} - V_t)' x_t$$

The Value at Risk for a specified confidence level β is given by:

$$\alpha_\beta(x_t) = \inf\{\alpha \in \mathbf{R} : F_{f(x_t, S_{t+1})}(\alpha) \geq \beta)\}$$

where $F_{f(x_t, S_{t+1})}(\alpha)$ represents the cumulative distribution function of portfolio loss at level α for a given portfolio x_t.

For a portfolio defined on the real line with no jumps, the conditional expectation of portfolio loss for a confidence level β is:

$$\theta_\beta(x_t) \;=\; (1-\beta)^{-1} \int_{F(x_t, S_{t+1}) \geq \alpha_\beta(x_t)} f(x_t, S_{t+1}) \cdot p(S) dS \tag{5.1}$$

where $p(S)$ is the probability distribution function of the underlying asset prices.

The minimization problem for mean-CTE optimization for the single time period t to $t+1$ is then:

$$\min_{x_t \epsilon X_t} \quad \{ \quad \theta_\beta(x_t) \quad \} \tag{5.2}$$

$$\texttt{subject to} \quad E[V_{t+1}'x_t] = \mu$$

$$V_t'x_t = B_t$$

$$x_t \geq l$$

The integral in equation (5.1) can be calculated through methods such as Monte Carlo simulation or delta-gamma approximations. Alexander, Coleman and Li (2006) showed that the minimization problem in (5.2) is also a convex programming problem, which opens up the optimization to more powerful convex programming problem machinery and software.

5.4.2 Portfolio Optimization Under a Hidden Markov Setting

As was the case for model validation in Chapter 3, portfolio optimization has an added layer of difficulty when the stock prices as assumed to follow a regime switching process. The underlying state structure can also present an opportunity for an investor to make more informed investment decisions and thus experience better portfolio performance.

The unobserved regime switching process provides some choices with regard to proceeding with optimization. One could simply derive the stationary unconditional measures of risk and distributions and proceed accordingly.

Where the investor knows exactly what the regime the process will be in for the period of investment, the optimization becomes much more precise, as the other regimes do not matter. The investor can simply optimize within the correct regime for the overall optimal solution.

The information the investor has in the hidden Markov setting lies somewhere in the middle. Given the data observations up to and including the observation at time t, the investor can generate the probabilities of the process being in each regime at time t; the probabilities: $P[\rho_t = 1|\mathbf{F_t}], \ldots, P[\rho_t = k|\mathbf{F_t}]$ for a regime switching process with k regimes. The probability that the process will be in regime i for the next data observation (the period the investor wishes to optimize over) is then $\sum_{j=1}^{k} P[\rho_t = j|\mathbf{F_t}] \cdot p_{j,i}$ where $p_{j,i}$ is the transition probability from state j to state i.

Three different options for optimization over a single period are presented here.

1. **Static Optimization**

   This method is as described above and does not make use of the probabilities

   $$P[\rho_t = 1|\mathbf{F_t}], \ldots, P[\rho_t = k|\mathbf{F_t}]$$

   It is included mainly for demonstration purposes.

   Unconditional optimization is akin to the weighted average residual option from Chap-

ter 3. The unconditional risk measures and utility distribution are obtained through conditional expectation:

$$E[X] = E[E[X|\rho_{t+1}]] \qquad\qquad Var[X] = E[Var[X|\rho_{t+1}]] + Var[E[X|\rho_{t+1}]]$$

For the stationary option of this type of optimization, the probabilities

$$P[\rho_t = 1|\mathbf{F_t}], \ldots, P[\rho_t = k|\mathbf{F_t}]$$

are assumed unknown, and can simply be replaced with the stationary probabilities

$$\pi_1, \ldots, \pi_k$$

and likewise the probabilities $P[\rho_{t+1} = 1|\mathbf{F_t}], \ldots, P[\rho_{t+1} = k|\mathbf{F_t}]$ will also be approximated by the stationary probabilities.

2. **Dynamic Optimization**

   This method is identical to the stationary unconditional optimization method, except we now assume

   $$P[\rho_t = 1|\mathbf{F_t}], \ldots, P[\rho_t = k|\mathbf{F_t}]$$

   are known. We then use the probabilities

   $$\left(\sum_{j=1}^{k} P[\rho_t = j|\mathbf{F_t}] \cdot p_{j,1}, \quad \ldots \quad , \sum_{j=1}^{k} P[\rho_t = j|\mathbf{F_t}] \cdot p_{j,k} \right)$$

   for the unconditional risk measure and utility distribution calculation.

   For example, the unconditional expected utility function will be

   $$E[u(V'_{t+1}x_t)] = \sum_{i=1}^{k} E\left[u(V'_{t+1}x_t)|\rho = i\right] \sum_{j=1}^{k} P[\rho_t = j|\mathbf{F_t}] \cdot p_{j,i}$$

   The variance and CTE risk measures can likewise be calculated through the conditional

expectation and variance formulas.

One potential issue with this optimization method is, ever when the hidden Markov model is known, the optimizer never delivers the optimal portfolio consistent with the optimal portfolio over one of the two regimes. Under the model, the next data observation's distribution is going to be the same as one of the underlying state distributions, not a weighted average of the states. A weighted average optimization is never going to produce an optimized portfolio consistent with the model mechanics (unless one of the estimated time $t + 1$ regime probabilities is exactly 1).

Another potential issue relative to the other methods presented is that this is the only method that requires re-optimization for each time period, as the incoming information will update the next period regime probabilities. This method could potentially become relatively computationally taxing depending on the frequency and the scope of the investment.

3. **Indicator Optimization**

Similar to the indicator method for residual selection, indicator optimization selects the portfolio that is optimized to the most likely regime for the next data observation.

Let $g(x|\rho = i)$ represent the objective function for optimization under the ith regime. Then the optimization objective line is:

$$\max/\min_{x \epsilon X} \quad g(x|\rho = h)$$

where

$$h = \operatorname*{argmax}_{\texttt{all i}} \sum_{j=1}^{k} P[\rho_t = j | \mathbf{F_t}] \cdot p_{j,i}$$

The indicator optimization method has the attractive quality of correctly optimizing when, under the model, the distribution of the market for the time period in question is consistent with that of the selected regime h. However, when h's distribution is not consistent with that of the market, the chosen portfolio is optimized for a different distribution.

As opposed to non-stationary unconditional optimization, optimization under the indicator method does not need to be performed at each time period, assuming the fitted regime switching distribution has not changed. Once the model has been fitted and selected, the optimizations within each regime can be performed right away. The only factor that changes with the incoming data are the probabilities $p_k(t)$, which only impact which optimized portfolio is selected. Indicator optimization will therefore require much less computational time than non-stationary unconditional optimization.

Although the above methods are listed for single-period optimization, they all can be applied to multi-period optimization through iteration. The latter two methods depend on the data observations, and should therefore make use of the most up-to-date information.

Static optimization does not depend on the data. For multi-period optimizations, this method can be changed to a single optimization using the distribution of the market returns over the entire investment period. This strategy would be a buy and hold strategy, versus an iterated single-period strategy, which would require re-balancing to the unconditional stationary weights after each period.

The effectiveness of each of these hidden Markov optimization methods will be investigated using the S&P 500 and NASDAQ data set as an example. They will first be evaluated under the hypothetical circumstance that the RSLN-2 model estimated through maximum likelihood is fully and completely representative of the market. Next, the effects of parameter uncertainty on the performance of the hidden Markov mean-CTE optimization methods will be evaluated, and conclusions about the important factors for optimization in practice for long-term equity data consistent with hidden Markov models will be made.

5.4.3 Portfolio Optimization under the Maximum Likelihood Estimates for the S&P 500 and NASDAQ Indexes

As discussed in Sections 6.2 and 6.3, the indexes of interest for this example are the S&P 500 and the NASDAQ. A BV-RSLN2 model has already been fitted to this data using maximum likelihood estimation, and the parameters are given in Table 5.1. For this first example,

this model is assumed to be completely representative of the market, and market stock movements will be simulated from the model. The model is assumed to be fully known to the investor.

The investor's objective will be to minimize the conditional tail expectation of her portfolio loss at the 95% level, given a preset desired portfolio expected return, and subject to budget and short-selling constraints. The investor will be able to invest in both the S&P 500 and the NASDAQ, in addition to a bond that yields an annual interest rate of 6%. The expected portfolio monthly return of the investor will be 0.75% of the investment, and no short-selling will be allowed. The investor's budget will be set to $1, although as shown in Alexander, Coleman and Li (2006), the optimal portfolio weights will found through the optimization will be independent of the set budget level. The portfolio performance over a single one-month period will be analyzed.

It is important to note that, assuming the model used is perfectly representative of the market, the dynamic method will optimize as intended. The objective, then, of this preliminary analysis will be to evaluate the performance of the indicator strategy relative to the perfectly optimized strategy of the dynamic/static approach.

The optimizations will be performed using the CVX convex optimization tool for MATLAB. CVX is a programming tool that solves convex optimization problems, and min-CTE optimization is a convex optimization problem as shown by Alexander, Coleman and Li (2006). The method for optimization involves simulating market movements under the probability measure defined by the model, and then deriving the optimal portfolio weights, treating the simulations as the distribution of market movements. One hundred thousand simulations of one-month market movements were used as input for the optimization procedure, and it was found that repeated (different) market simulation sets yielded almost identical optimal portfolio weights.

The obtained portfolio weights for both the indicator and dynamic approaches under the model's stationary distribution are listed in Table 5.4. The dynamic portfolio is significantly riskier than the indicator portfolio, and this makes sense. The indicator portfolio assumes the data are distributed according to the first regime, where expected returns are positive and market volatility is relatively low. This means that to achieve the expected monthly

Method	S&P 500	NASDAQ	Bond
Indicator	0.3344	0.0792	0.5864
Dynamic	0.3299	0.4504	0.2197

Table 5.4: Optimal Portfolio Weights for the Indicator and Dynamic Methods Under the Stationary Distribution for the S&P 500 and NASDAQ Maximum Likelihood Model

Method	E[Return]	$CTE_{95\%}$
Indicator	1.0061	0.9609
Dynamic	1.0075	0.9172

Table 5.5: Performance of the Indicator and Dynamic Methods Under the Stationary Distribution for the S&P 500 and NASDAQ Maximum Likelihood Model

return of 0.75%, the investor only need invest approximately 40% of her portfolio in stock, and the rest can be placed in bonds. For the portion invested in stock, the S&P is favored for its lower estimated volatility.

Conversely, for the dynamic portfolio, which is optimized over the unconditional distribution of the market and includes the possibility of negative returns from the second regime, we obtain the result that a more aggressive portfolio is required to achieve the expected portfolio return of 0.75%. A significantly larger portion of the portfolio is invested in stock (close to 80%), with a much heavier NASDAQ position due to its higher expected return.

Now that the portfolio weights have been calculated, market returns are simulated using the maximum likelihood BRSLN-2 distribution and portfolio performance can be assessed. The achieved expected returns under each of the methods, and the estimated CTE at 95% loss (displayed in the form of the portfolio value) are shown in Table 5.5. The dynamic method, by construction, achieves the expected portfolio return of 0.75% at the minimum portfolio loss CTE at 95%, which corresponds to a portfolio value of $0.9172. The indicator method only achieves an expected return of 0.61%, which is only 81.3% of the expected return of the dynamic method. This significant difference, under this example of the maximum likelihood estimates perfectly representative of the data, suggests that the indicator method will not achieve desired portfolio performance.

5.4.4 Portfolio Optimization under the Bayesian Estimates for the S&P 500 and NASDAQ Indexes

For the second half of this example, parameter uncertainty will be the focus, and its effect on the different hidden Markov model portfolio optimization methods will be explored. Now the market will be assumed to follow the BRSLN-2 model under its Bayesian estimation from Section 5.3.

Recall from Section 5.3 that parameter uncertainty was greatest for the transition probability parameter $p_{2,1}$ and the state 2 volatility parameters for both the S&P 500 and the NASDAQ. The NASDAQ parameters in general also exhibited wider posteriors than those for the S&P 500, indicating that there was more uncertainty surrounding the NASDAQ as well. It stands to reason that a portfolio heavy in the NASDAQ (such as the MLE Dynamic optimized portfolio) would be more sensitive in terms of portfolio performance to the greater uncertainty in these parameters than a portfolio that was relatively lighter in the NASDAQ index.

The performance of three different optimized portfolios will be measured against the BRSLN-2 distribution under the Bayesian estimates. The MLE indicator and dynamic optimized portfolios from the first half of the example will be considered, and the third will be another dynamically optimized portfolio, but this time using the Bayesian BRSLN-2 distribution as the input distribution for the optimization procedure. The optimization was calculated in the same fashion as for the dynamic MLE portfolio, simply using a different input distribution.

The Bayesian dynamic optimal portfolio weights are shown in Table 5.6, along with the old weights that used the MLE distribution. Interestingly, the Bayesian dynamic portfolio weights resemble the MLE indicator weights more closely than they do the MLE dynamic weights. Investigation into this result yielded at least two factors that played a role. The first is that the posterior distribution of the parameter $p_{2,1}$ had more probability concentrated in the right tail than in the left (Figure 5.2). This asymmetry meant that under the Bayesian distribution of the model, less time was spent in the second regime than under the MLE model. This is turn meant that the market returns in general were less likely to be negative, and thus less of the portfolio needed to be allocated to stock to achieve the desired expected return. The second factor is that the greater uncertainty surrounding the

Method	S&P	NASDAQ	Bond
MLE Indicator	0.3344	0.0792	0.5864
MLE Dynamic	0.3299	0.4504	0.2197
Bayesian Dynamic	0.2168	0.2594	0.5238

Table 5.6: Optimal Portfolio Weights for the Indicator and Dynamic Methods Under the Stationary Distribution for the S&P 500 and NASDAQ Using Different Estimation Methods

NASDAQ parameters meant that the stock portion of the Bayesian dynamic portfolio was allocated more towards the S&P 500 to minimize risk.

Once the Bayesian dynamic optimized portfolio was calculated, the three portfolios performance was measure against market simulations, which this time were generated from the Bayesian estimated BRSLN-2 distribution. In this case, the Bayesian dynamic portfolio is the portfolio perfectly optimized against the market, and the performance of the MLE indicator and MLE dynamic portfolios are of primary interest.

The expected returns achieved, and the 95% CTE of the loss function (shown again in the form of portfolio value) are shown in Table 5.7. The MLE dynamic portfolio was a very risky portfolio under this market. Though compensated with a higher expected return, its portfolio loss at the 95% CTE level was almost double that of the Bayesian dynamic portfolio. That parameter uncertainty could have as strong an effect as doubling the required reserve to cover the potential portfolio loss should be a warning to any investor using hidden Markov portfolio optimization without taking this uncertainty into account.

The indicator optimized portfolio, on the other hand, performed quite similarly to the Bayesian dynamic optimized portfolio. This is a function of the nature of the parameter uncertainty, as described above. The distributions of the final portfolio values of the three methods are shown in Figure 5.6. The plots show that parameter uncertainty can have just as large an effect as the optimization method selected under hidden Markov optimization. There is also a benefit seen here in using the MLE indicator optimized portfolio over the MLE dynamic optimized portfolio, due it beng less reliant on the parameters of greatest uncertainty.

Method	E[Return]	$CTE_{95\%}$ (MLE Est.)
MLE Indicator	1.0070	0.9554 (0.9609)
MLE Dynamic	1.0091	0.8999 (0.9172)
Bayesian Dynamic	1.0075	0.9472

Table 5.7: Performance of the Indicator and Dynamic Methods Under the Stationary Distribution for the S&P 500 and NASDAQ Bayesian Model
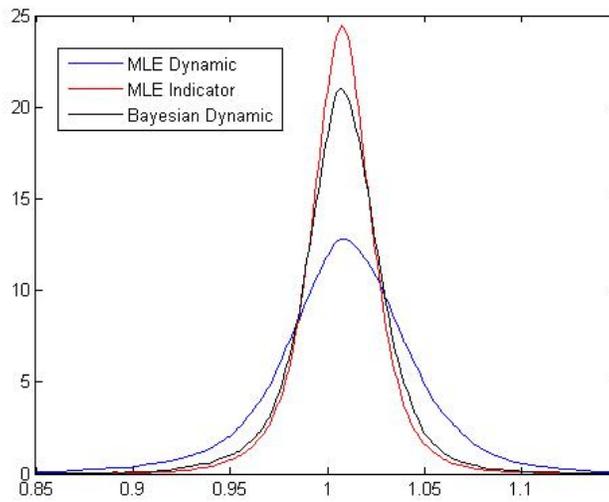
Figure 5.6: Distribution of the Single Period Returns for the Different Optimzation Methods under the Bayesian BRSLN-2 Model

5.4.5   Summary of the BRSLN-2 Portfolio Results

The preceding example demonstrated some important findings for hidden Markov portfolio optimization:

- Parameter uncertainty can have a profound effect on portfolio optimization under hidden Markov assumptions for equity returns. The uncertainty surrounding the parameters from the low mean, high volatility state was the cause of this effect when optimizing over a BRSLN-2 model fitted to the NASDAQ and S&P 500 indexes.

- The dynamic hidden Markov optimized portfolio, when calculated using the maximum likelihood estimates of the market, was found to be less robust to risk associated with parameter uncertainty than the indicator optimized portfolio. A main reason for this is that the optimal portfolio under the dynamic method places more portfolio weight with the risky assets. The difference in portfolio performance can be significant, and it is advisable that portfolion optimization using the hidden Markov dynamic method should always take parameter uncertainty into account.

- The indicator method provided an effective alternative for the major American indexes, due to it being less reliant on the parameters of the model for which parameter uncertainty is greatest, and is significantly less computationally taxing than the dynamic method.

5.5   Conclusion

This chapter explored hidden Markov portfolio optimization from the model fitting stage to the optimized portfolio and the effect of parameter uncertainty. The estimated hidden Markov model's underlying state framework for the S&P 500 and NASDAQ indexes resembled that of the single index model: one state with high expected returns and low volatility, and another with the low expected returns and relatively high volatility. The posterior distribution from the Bayesian estimation indicated that, similar to the single index case, the greatest uncertainty surrounded the high volatility regime.

Different methods of hidden Markov portfolio optimization were described and compared using an example involving the S&P 500 and the NASDAQ indexes. The example demonstrated the importance of taking parameter uncertainty into account for hidden Markov portfolio optimization, and that optimizations that are less reliant on the parameters with the greater uncertainty, namely indicator optimization, will be more robust to its effects.

Chapter 6

Hidden Markov Portfolio Replication

6.1　Introduction

In this chapter, the problem of interest changes from hidden Markov portfolio optimization to hidden Markov portfolio replication. More specifically, the focus will be on the delta hedging of an equity guarantee under a hidden Markov state framework.

Under the famous delta hedging strategy of Black & Scholes (1973), one of the assumptions made by the Black-Scholes framework is that stock prices follow a geometric Brownian motion with a constant volatility parameter. As argued in Chapters 2 through 4 of this thesis, such an assumption is not realistic when looking at long-term equity data. Models with clustering volatility, such as hidden Markov models, can provide much better fits to indexes such as the S&P 500.

The literature concerning hedging under hidden Markov stock assumptions is limited. Alizadeh and Nomikos (2004) used a hidden Markov strategy for hedging an asset using its futures value on the S&P 500 and FTSE 100. Their approach was to calculate the mean-variance hedged portfolios for each of the underlying states, and then average the respective positions using the data-dependent regime probabilities. They compared their regime-switching results against a portfolio hedged using the volatility estimated from a GARCH

process, and found that the regime-switching approach performed better for some cases, and the GARCH approach better for others.

This chapter will first discuss delta hedging under the Black-Scholes framework and discuss issues with the volatility assumption. Then, under the assumption of a hidden Markov model for the data, this chapter will present different options for hedging through different selection processes for the volatility parameter. It will then discuss the implications of each of the strategies, and the circumstances under which they best perform. The focus will be on deep out-of-the-money options because they are the most relevant to investment guarantees in insurance. A deep out-of-the-money put option example, under the RSLN-2 model fitted in Chapter 2, will demonstrate the effectiveness of the different replicating strategies. Finally, the same example will adjusted to include the Bayesian analysis of Chapter 4 to demonstrate the impact of parameter uncertainty on the hidden Markov hedging strategies.

This Chapter contributes to the literature on hidden Markov hedging by:

- Presenting and comparing different options for portfolio replication in a hidden Markov setting;

- Demonstrating the implications of these different options for a deep out-of-the-money put option for the S&P 500; and

- Assessing the impact of model parameter uncertainty on those implications.

6.2 Black-Scholes Delta-Hedging

Developed in the 1970's, the Black-Scholes delta-hedging strategy is still widely used to fund liabilities that are or resemble equity derivatives. The strategy, based on some assumptions made about both the market and the stock price, creates a portfolio consisting of positions in both the option's underlying equity and a risk-free instrument. The portfolio will then approximately replicate the movement of the value of option relative to changes in the underlying stock's price.

The assumptions made about the stock and the market by the Black-Scholes framework are:

- The asset price, S_t, follows a geometric Brownian motion (GBM) with constant volatility parameter σ.

- Trading in the market is assumed to be done continuously, and the replicating portfolio will be continuously re-balanced.

- Short selling is allowed in the market without restriction.

- No transaction costs or taxes are present in the market.

- All market investors can borrow and lend money at the same fixed force of interest r.

Each of the above assumptions above is unrealistic in at least some sense in trading markets. The ability to continuously re-balance every replicating portfolio for every block of business that every insurer has is obviously infeasible. Stock prices in the long-run, as demonstrated in Chapter 2, do not resemble geometric Brownian motions, nor do they have constant volatilities. There are limits to short-selling even for the largest of financial institutions. Transaction costs and capital gains taxes can significantly affect returns for market investors. Interest rates fluctuate, and there is almost always a bid-ask spread for risk-free interest rates that banks will offer investors.

However, the assumptions can be relaxed or their costs be quantified. For this chapter, the underlying asset price can be assumed to follow more complicated processes, such as those described in Chapter 2. Trading will be assumed to be done at discrete intervals, which will introduce a difference in value between the maturing portfolio and the new replicating portfolio at the end of each interval. This difference is also known as hedging error or tracking error. This error will be tracked and valued over time as a cost of hedging the liability. Transaction costs will be incorporated.

The two assumptions that will be kept for the analysis of this chapter will be the lack of restriction on short selling and the constant interest rate assumptions. Further investigations about the impact and possible relaxing of these assumptions in the context of replicating portfolios under hidden Markov models will be left as subject for further research.

6.2.1 Hedging Under the Black-Scholes Model

A major result from the Black-Scholes-Merton framework is that security derivative's price is equal to the derivative's expected value under the risk-neutral measure Q, discounted at the risk-free force of interest r.

$$P_t = e^{-r(T-t)} E_Q[V_T]$$

Under the no-arbitrage assumption, P_t must also be the price of the replicating portfolio at time t.

The Black-Scholes-Merton framework also describes how the hedging portfolio is constructed. Let Ω_t represent the partial derivative of the security price P_t with respect to the asset price S_t

$$\Omega_t = \frac{\partial P_t}{\partial S_t}$$

The portfolio that will exactly replicate the security holds $\Omega_t \cdot S_t$ units in the risky asset, and $P_t - \Omega_t \cdot S_t$ in the risk-free asset. Moreover, this portfolio will be self-financing, meaning that after each infinitesimal time step, the resulting hedged portfolio value will equal the value needed to hedge the security over the next infinitesimal time step.

The Hedge of a Simple Guaranteed Minimum Maturity Benefit

A well known result under the Black-Scholes-Merton framework is the price of European put and call options. Let t represent the current time, T the time of maturity of the contract and K the strike price of the put option. Assume the asset price S_t has variance $\sigma^2 \cdot t$ over a period of length t. The Black-Scholes European put option price, $BSP(t)$ is given by

$$BSP_t = K \cdot e^{-r(T-t)} \cdot \Phi(-d_2) - S_t \cdot \Phi(-d_1)$$

where d_1 and d_2 are given by

$$
\begin{aligned}
d_1 &= \frac{\log(S_t/K) + (T-t)(r + \sigma^2/2)}{\sqrt{T-t}\sigma} \\
d_2 &= d_1 - \sigma\sqrt{T-t}
\end{aligned}
$$

and $\Phi(x)$ is the standard Normal cumulative distribution function evaluated at x.

The hedged portfolio holds $\Omega_t \cdot S_t$ in the risky asset, where Ω_t has been proven to be

$$
\Omega_t = -\Phi(-d_1)
$$

under the Black-Scholes-Merton framework (see Hull, 1989 for details).

6.3 Regime-Switching Portfolio Replication

The Black-Scholes replication strategy requires a single value input for the process volatility, which corresponds to the assumed volatility of the Brownian motion process for the underlying stock. However, it has been argued in this thesis that such a model for the stock process provides an unsatisfactory fit for long-term equity data. Moreover, even the more relaxed assumption of constant volatility in the stock is shown to be non-trivially violated. The dynamic nature of volatility of the equity process led to the fitting and validation of the class of hidden Markov models, for which the volatility of the process can be quite different from time period to time period under the model.

In this chapter, we explore the efficiency of hedging assuming an underlying hidden Markov model. Given the significant effort put forth into validation and assessment of hidden Markov equity models for the valuation of various investment guarantee contracts, a natural step is the incorporation of these models into the financing strategy for these contracts. The challenge becomes finding the most efficient way of utilizing both the hidden Markov model

mechanics and the available information contained in the data to hedge the contracts.

Under the hidden Markov model structure, volatility parameters for each of the underlying states are defined. However, the Black-Scholes replication formula requires that one value for stock volatility be used for calculating portfolio values and hedging requirements. The remainder of this chapter will introduce different types of volatility selection given a hidden Markov model for the underlying equity, discuss the implications of each, and investigate and assess their performance using examples.

6.3.1   The Static Unconditional Volatility Approach

A simple approach to deriving a single volatility estimate from the hidden Markov model structure is to average the volatility parameters of each of the regimes using the invariant distribution for the regimes as the probability measure. This can be done using the formula for conditional variance. The volatility parameter used at time t for portfolio replication can be expressed as:

$$
\begin{aligned}
\sigma_t^{STAT} &= \sqrt{Var[Y_t]} \\
&= \sqrt{Var_\pi[E[Y_t|\rho_t]] + E_\pi[Var[Y_t|\rho_t]]}
\end{aligned}
$$

Given the underlying state ρ_t, the mean and variance of the process are defined by the model, $E[Y_t|\rho_t] = \mu_{\rho_t}$ and $Var[Y_t|\rho_t] = \sigma_{\rho_t}^2$ respectively. Under the invariant distribution, where $\pi_k, k \in (1, \ldots, n)$ represents the probability that the underlying state is state k at any time given no additional information, the unconditional process volatility can be more explicitly expressed as

$$
\sigma_t^{STAT} = \sqrt{\sum_{k=1}^{n} \mu_k^2 \cdot \pi_k - \left(\sum_{k=1}^{n} \mu_k \cdot \pi_k\right)^2 + \sum_{k=1}^{n} \sigma_k^2 \cdot \pi_k}
$$

This approach will be called the static unconditional approach. The σ_t value never changes,

as the processes invariant distribution never changes, and is not impacted at all by the incoming observed data of the time series.

This approach is not expected to perform significantly different than the standard Black-Scholes replication strategy using a simple log-normal model, as they both estimate a long-run constant volatility of the process and plug that into the Black-Scholes equations. Moreover, this approach eliminates the main attraction of using hidden Markov models in the first place: capturing the dynamic nature of the volatility of the stock process. Examples using this method will be performed, but mainly so they can be used as a reference point from which the relative success of other methods can be measured.

6.3.2 The Dynamic Unconditional Volatility Approach

Instead of using the unconditional regime probabilities as the probability measure for deriving an unconditional volatility, one could instead use the data dependent regime probabilities. These probabilities take into account the movements of the data up to and including the current observation, resulting in more informed probability estimates of the underlying path of the series. As defined in section 2.4, the regime probabilities, conditional on the data and using the obtained estimates for the model parameters, are

$$p_k(t) = P[\rho_t = k | y_t, \ldots, y_1], \qquad k \in \{1, \ldots, K\}, \quad t \in \{1, \ldots, T\}$$

and are calculated recursively as before.

When creating a replicating portfolio for an option at time t, one is not so much concerned with the volatility of the process over the previous period $t - 1$ to t as one is with the coming period t to $t + 1$. It is this time period for which the re-balanced hedged portfolio is designed to move with the stock. Since the underlying regimes, ρ_t, represent the state of the underlying process for the log-return observation y_t, it is actually the distribution of ρ_{t+1} that is more important to the forward-looking hedger. The analyst still only has information from the data up to time t, however, meaning the regime probabilities conditional on the

data that the hedger is looking for are

$$P[\rho_{t+1} = k | y_t, \ldots, y_1], \qquad k \in \{1, \ldots, K\}, \quad t \in \{1, \ldots, T\}$$

These probabilities can be easily calculated from the probabilities $p_k(t)$. Let $p_k^+(t)$ represent the probability that, conditional on the data up to and including the observation at time t, the underlying state for the data observation at time $t+1$ will be regime k. The $p_k^+(t)$'s can then be calculated by

$$p_k^+(t) = \sum_{n=1}^{N} p_n(t) \cdot \hat{p_{n,k}} \qquad k \in \{1, \ldots, K\}, \quad t \in \{1, \ldots, T\}$$

or, descriptively, the sum of the probability of being in each of the states for the time t observation times the probability of transitioning from those states to state k under the estimated model.

Conditional on these data-dependent regime probabilities, the unconditional volatility of the process to be used for Black-Scholes portfolio replication can then be calculated:

$$\sigma_t^{DYN} = \sqrt{\sum_{k=1}^{n} \hat{\mu_k}^2 \cdot p_k^+(t) - \left(\sum_{k=1}^{n} \hat{\mu_k} \cdot p_k^+(t)\right)^2 + \sum_{k=1}^{n} \hat{\sigma_k}^2 \cdot p_k^+(t)}$$

This approach is called the Dynamic Unconditional Approach, as the volatility estimate of the upcoming observation will change based on the data observations and what underlying state path has been estimated.

Note this approach is still a weighted average approach, meaning the volatility estimates σ_t^{DYN} will fall somewhere between the regime-specific volatility parameters. This approach still estimates an overall volatility for the process, but now one that incorporates data movements in addition to model mechanics. The objective of this approach is to provide a volatility estimate for the hedger that will perform reasonably well under all feasible market conditions, whether the market behaves relatively stable or more volatile.

6.3.3 The Indicator Volatility Approach

An alternative approach to incorporating the data dependent future regime probabilities $p_k^+(t)$'s into the calculation of a Black-Scholes volatility input parameter is to use an indicator approach similar to that defined for residuals in Section 3.2.3. This approach selects the state-specific volatility parameter, σ_k, for which $p_k^+(t)$ is highest. More formally,

$$\sigma_t^{IND} = \hat{\sigma}_k, \qquad \text{where } p_k^+(t) = \max(p_1^+(t), \ldots, p_K^+(t))$$

This approach is quite different from the unconditional approaches in the sense that it does not try to perform reasonably well under all market conditions. It instead tries to perform the best it can under what the model estimates as the most likely set of market conditions, ignoring the potential downside of incorrectly estimating the market 'state'.

The potential upside of this approach is that if the different hidden Markov states of the model are adequately representative of the different market conditions that affect equity returns, and if within those states the model accurate captures market movements, then the majority of the time this approach will produce very effective hedges. It will generally not overestimate market volatility during stable times in an effort to account for possible shifting market conditions to more extreme returns the same way the dynamic unconditional approach does. Nor will it generally underestimate the volatility when the market is experiencing more drastic price movements.

The downside, however, is that this method will produce the least effective hedges when the state it selects is not representative of market conditions. If the approach selects a stable market state for a very volatile period, then the volatility used for hedging might be drastically underestimated. It can be likewise significantly overestimated for a relatively calm period of market activity.

The overall success of the indicator and dynamic unconditional approaches will be a a focus of the remainder of this chapter.

6.3.4 Additional Type of Hedging Error Introduced by the Dynamic Methods

One of the potential risks of using a dynamic approach to the volatility process under Black-Scholes hedging is that additional hedging error is introduced through that volatility. The Black-Scholes value of the option, and consequently the positions in stock and bond held by the hedger are dependent on the assumed volatility, and if that assumed volatility changes, so too will the calculated option value and hedged positions. This change alone can produce significant hedging error (current option price less accumulated hedge value), in addition to the hedging error from movements in the stock price itself.

Let $BSP_t(\sigma)$ and $HE_t(\sigma)$ be the respective Black-Scholes option price and hedging error at time t dependent on volatility input parameter σ. Normally under Black-Scholes hedging, this volatility parameter doesn't change, and thus the hedging error at time $t + 1$ can be represented as

$$HE_{t+1}(\sigma) \;=\; BSP_{t+1}(\sigma) - (H_t(\sigma) \cdot S_{t+1} + B_t(\sigma) \cdot e^r)$$

where $H_t(\sigma)$ and $B_t(\sigma)$ are the hedge position and the bond position entered into at time t respectively, given input volatility parameter σ.

However, if the assumed volatility were to change with time, letting σ_t be the calculated process volatility at time t, then the hedging error under the new process would be presented as

$$HE_{t+1}(\sigma_{t+1}) \;=\; BSP_{t+1}(\sigma_{t+1}) - (H_t(\sigma_t) \cdot S_{t+1} + B_t(\sigma_t) \cdot e^r)$$
$$HE_{t+1}(\sigma_{t+1}) \;=\; BSP_{t+1}(\sigma_{t+1}) - BSP_{t+1}(\sigma_t) + BSP_{t+1}(\sigma_t) - (H_t(\sigma_t) \cdot S_{t+1} + B_t(\sigma_t) \cdot e^r)$$

The hedging error can be de-constructed into two parts:

- The hedging error resultant from changing the processes volatility: $BSP_{t+1}(\sigma_{t+1}) - BSP_{t+1}(\sigma_t)$;

- The hedging error that normally results from the process due to gamma risk and using discrete time hedging: $BSP_{t+1}(\sigma_t) - (H_t(\sigma_t) \cdot S_{t+1} + B_t(\sigma_t) \cdot e^r)$

Transaction Costs

This extra hedging error can play a significant role due to the nature of transaction costs. Typically, transaction costs for trading are proportional to the size of the position change. A portfolio strategy that experiences smaller changes in its stock position will incur less transaction costs than a portfolio that experiences many large changes, even if the portfolios have the same average stock position.

The regime-switching portfolio replication methods are expected to experience different impacts from transaction costs. The static unconditional method has a constant volatility input parameter, and thus its transaction costs will operate similar to traditional portfolio replication: no transaction costs will result from changing volatility parameters.

The indicator replication method selects one of the volatility parameters associated with each of the regimes, and hedges using that parameter as input. For regime-switching models for equity data, where regime transitions are infrequent, transaction costs will for the most part behave similarly to the static unconditional method: the process spends a good deal of time remaining in its current regime, and thus the volatility parameter associated with that regime gets selected for many consecutive periods. However, when the underlying process is calculated to have transitioned to another regime, assuming the regimes have dissimilar distributions, then very large changes in the option value and thus large transaction costs associated with volatility parameter change can be expected.

For the dynamic unconditional replication approach, the volatility of the process is constantly changing, and thus transaction costs associated with volatility parameter change will be experienced at every re-balance. However, because the volatility parameter is a weighted average of the regime-specific volatility parameters, the dynamic volatility will always be

Regime	μ	σ	Transition Parameters
One	0.01024	0.03384	$p_{1,2} = 0.0337$
Two	-0.01448	0.06486	$p_{2,1} = 0.1517$

Table 6.1: The RSLN-2 MLE Parameters Fitted to the S&P 500 Monthly Log-Return from Jan 1950 to Oct 2010

calculated somewhere in the middle of the extreme volatility parameters associated with the regimes. While transaction costs will be associated with every portfolio re-balance, they will be smaller in magnitude than the large jumps associated with the indicator replication approach.

An important aspect of the analysis of each of these methods is the relative trade-off of potentially less hedging error due to gamma risk because of more responsive volatilities with larger transaction costs due to more volatile portfolios.

6.4 A Portfolio Replication Put Example using the RSLN-2 Model for the S&P 500

To investigate the effectiveness of each of the proposed hidden Markov portfolio replication methods, an example under ideal circumstances will be presented. Equity returns for these examples will be simulated from the RSLN-2 model estimated from the S&P 500 monthly data using maximum likelihood. Moreover, the model is assumed known to the investor, eliminating any possibility of parameter uncertainty or model error. The model parameters, as calculated in Section 2.4.1, are displayed again here in Table 6.1 for quick reference.

The example used will be a 10-year put option on the simulated index. The index value at time 0 will be $100, and the strike price $100. The put option example emulates a 10-year GMMB contract on the index, without the management expense ratio (the guarantee value is assumed paid for upfront in full). Transaction costs are assumed to be 0.02% of the change in the stock position for each transaction. The risk-free force of interest, from which investors can borrow or lend at no cost, is assumed at 5% per annum.

Volatility	Static	Dynamic	Indicator
EPV[Total Hedging Costs]	2.9681	2.5420	2.2518
90% CTE[Total Hedging Costs]	5.8092	5.6619	6.4691

Table 6.2: Total Hedging Cost Averages and 90% CTE's for the Hedging Strategies for the S&P 500 Put Option Example

The seller of the contracts will fund the liabilities of the contracts with a Black-Scholes hedge under the static unconditional, dynamic unconditional and indicator strategies defined above. The replicating portfolios will be re-balanced monthly. The discounted total cost of hedging the contract (initial hedge value plus monthly re-balancing costs) will be calculated, and used to measure the effectiveness of each of the hedging strategies. 100,000 simulations of 10-year index movements and resultant portfolio hedges were performed.

6.4.1 Maximum Likelihood Results

Since hedging is a trade of higher expected liability costs for lower risk, the effectiveness of the strategies will be evaluated on a risk/reward basis. The reward will be the expected total hedging costs of the contract, for which smaller values are obviously better, and the risk measure will be the 90% conditional tail expectation of the total hedging cost distribution, again for which smaller values are better.

The expected total hedging costs and 90% CTE's of the hedging cost distributions are displayed in Table 6.2. At a first glance, the dynamic replication method outperforms the static replication for both average total cost (14.4% less) and in the right tail of the hedging cost distribution (2.6% less). The dynamic hidden Markov strategy has already proven itself significantly better than a static hedging strategy, under a hidden Markov model. The indicator method is riskier than both the static and dynamic hedging strategies, but offers a much lower expected total cost in return for this risk.

The distributions of the total hedging costs are displayed in Figure 6.1. The indicator hidden Markov replication strategy clearly has on average lower total hedging costs, but also has the thickest tails for its hedging cost distribution. The dynamic replication strategy has the thinner total hedging cost right tail of the three distributions.
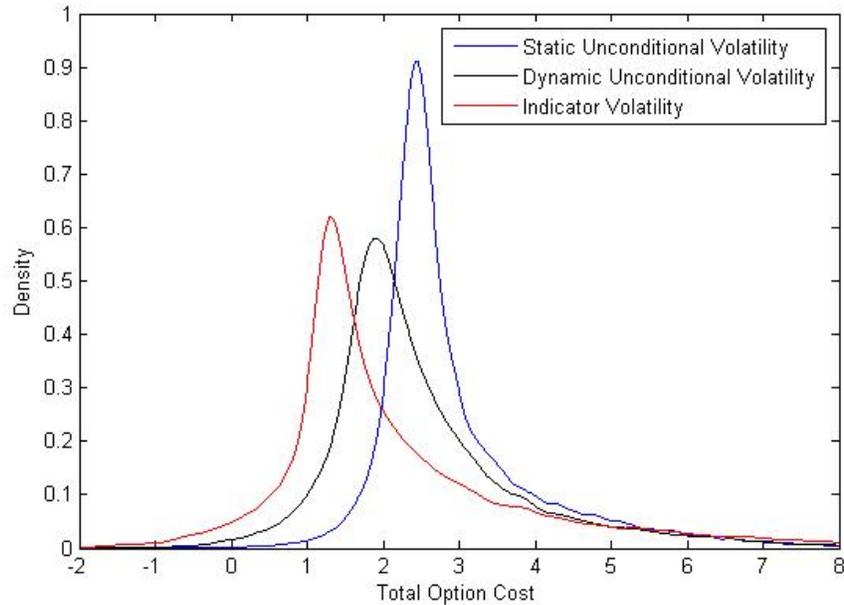
Figure 6.1: Total Hedging Cost Distributions for the Hedging Strategies for the S&P 500 Put Option Example

The reason for the relative success of the indicator and dynamic methods over the static strategy is because of the impact of inaccuracy in the volatility parameters used. Under the model, the static strategy either significantly underestimates the volatility or significantly overestimates the volatility. When the market volatility is higher than the static prediction, the model is in state two which is associated with negative returns. Since the short position of the static strategy is not as large as it should be, when the negative returns are experienced it does not gain as much money as it should. Conversely, when the market is in state one, the static method overestimates the volatility, which means its short position is larger than it should be. Since state one is associated with positive returns, the static method is losing more money on those positive returns due to its larger short position. So no matter what the underlying state, the static method portfolio performs worse than it should due to volatility mis-estimation. The indicator and dynamic strategies infer from the data observations more accurate information about current market volatility, and therefore result in better predictions of volatility and thus better hedges.

In order to quantify the effectiveness of the indicator and hidden Markov strategies over a

static strategy, however, it will be necessary to compare their expected total hedging costs against static strategies with similar 90% total hedging cost CTE's, or compare their 90% hedging cost CTE's with static strategies of similar expected hedging costs.

As stated earlier, the larger one's hedge position, the larger the expected hedging costs associated with that portfolio, but the smaller the portfolio risk. Static strategies at different hedge sizes will therefore have different associated total hedge cost distributions. The hedge size of a static strategy can be controlled by the level of the volatility parameter. The larger the volatility parameter, the more the process is expected to experience the deeper tails, and therefore the larger the hedge size, and visa versa.

The black line in Figure 6.2 shows the relationship of total expected hedging costs to the 90% CTE of hedging costs for static replication strategies. The dynamic and indicator hidden Markov replication strategies are also plotted, represented by the blue and magenta dots respectively. Both hidden Markov strategies are significantly below the static strategy frontier.

Relative to the static strategy of the same expected total hedging costs, the dynamic hidden Markov strategy offers a 16.7% percent reduction in the 90% CTE hedging cost risk. Similarly, the indicator strategy also reduces the right 90% CTE tail risk by 17.0% relative to a static strategy with the same expected total hedging cost. From a different angle, the dynamic strategy also has an average hedging cost 16.7% less than the corresponding static strategy with an identical total hedging cost 90% CTE. The expected total hedging cost of the indicator strategy is 14.5% less than the expected cost for a static strategy with the same hedging cost 90% CTE.

Both the indicator and dynamic hidden Markov strategies have proven very effective against static strategies for delta hedging an deep out-of-the-money put option for the hidden Markov model fitted to the S&P 500. The indicator strategy is somewhat riskier than the dynamic strategy, but offers a smaller expected total hedging cost in compensation. Comparisons between the indicator and dynamic strategies would require a cost/benefit assumption of investor preferences for risk and reward, similar to a utility function, which will be subjective from investor to investor. The two methods provide investors using hidden Markov models with effective choices.
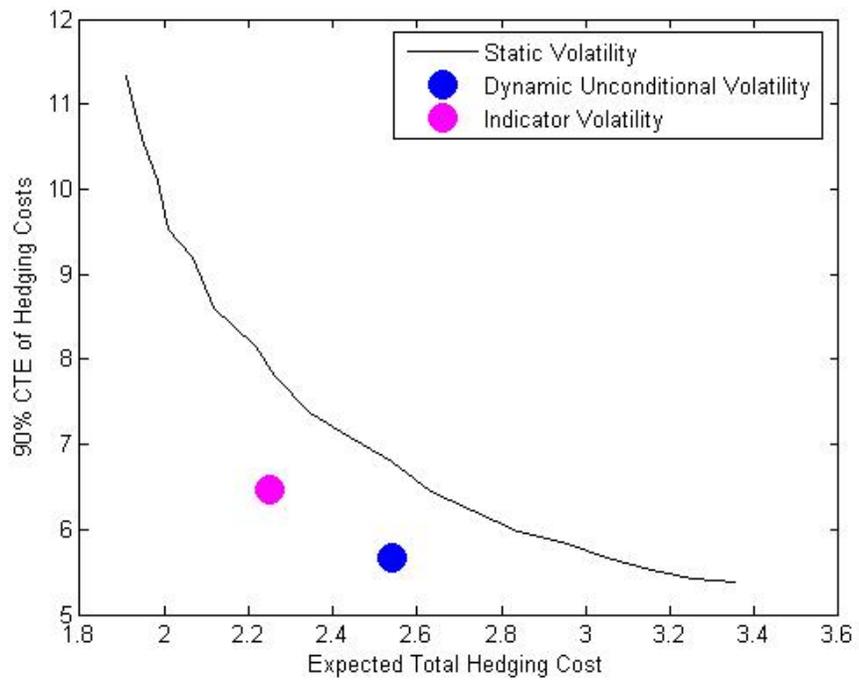
Figure 6.2: Comparison of Expected Hedging Costs and 90% Hedging Cost CTE's for Static and hidden Markov Hedging Strategies for the S&P 500 Put Option Example

This example was constructed under the hypothetical circumstance that the hidden Markov model is perfectly known to the investor and that the model is perfectly representative of market data. The next section will present the same example, except now the market will be constructed to perform differently than the investor expects, encapsulating the effect that parameter uncertainty can have on the hidden Markov replication strategies.

6.4.2   Evaluating the Effects of Parameter Uncertainty

One method of assessing the robustness of the performance of each of the hedging strategies is to evaluate their performance when faced with parameter uncertainty. Given the different approaches of the dynamic and indicator hedging strategies, parameter uncertainty may play a different role under each of the strategies. The indicator approach, for example, selects a regime-specific volatility as its hedging volatility which does not incorporate any parameters not associated with that regime. This may result in parameter uncertainty in the volatility parameters having a relatively larger effect on the indicator strategy. Likewise, uncertainty in the transition probability parameters, or even the regime-specific log-return means, may have a larger impact on the performance of the dynamic strategy, which incorporates all model parameters in each volatility calculation.

Parameter uncertainty was a central focus of Chapter 4 of this thesis, where a Bayesian approach was used to evaluate parameter uncertainty (and the resultant implications) of long-term equity models. Since the put example above used estimates for the monthly S&P 500 data, the Bayesian results for the S&P 500 models can be applied to the example to investigate the impact of parameter uncertainty on the hidden Markov hedging strategies.

The posterior distributions of the parameters of the RSLN-2 model fitted to the S&P 500 data, as found in Chapter 4, are displayed again here for quick reference (Figure 6.3). The regime 2 mean posterior distribution has a thicker left tail than a right one, and the regime 2 volatility parameter has a thicker right tail than a left one, meaning the uncertainty of the state extends towards it being even more volatile and less rewarding than the other direction. This could have especially severe implications on hedging strategies for low strike price put, given that under the model for the put to be in the money, the state path will almost

139
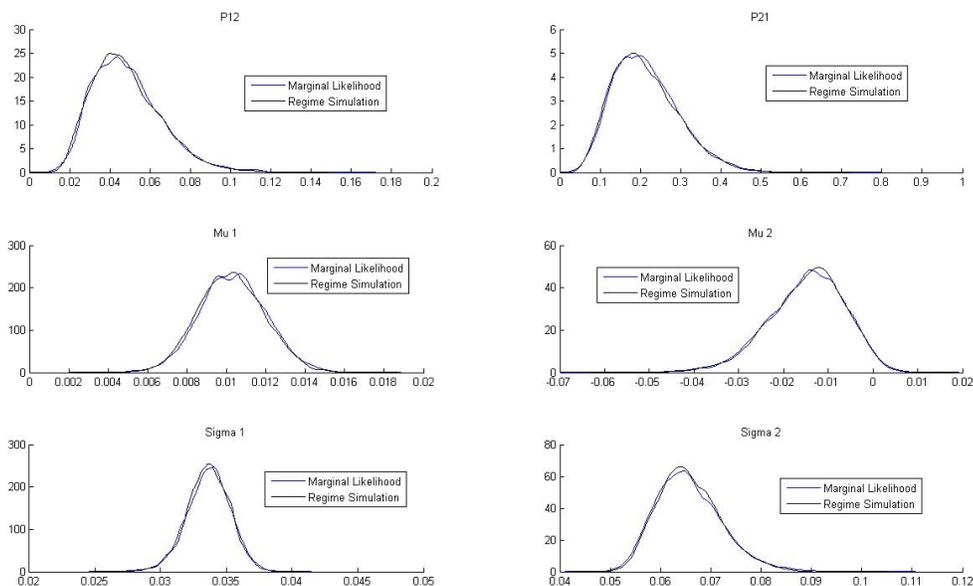
Figure 6.3: Distributions of the RSLN-2 parameters under Regime-Switching Bayesian Estimation

certainly have to spend a significant amount of time in this state. The relative uncertainty of the transition probability parameter $p_{2,1}$ may also have significant implications on the hedging strategies.

As a means of assessing the impact of parameter uncertainty the following adjustments to the example were made:

- Instead of the maximum likelihood parameter estimates, the Bayesian posterior parameter distributions obtained in Chapter 4 are used to generate the simulated 'real-world' data. An equal number of market simulations were performed for each parameter set iteration of the converged Markov chain.

- The hedger still uses the maximum likelihood RSLN-2 parameter estimates to perform her calculations for regime probabilities and portfolio weights.

The objective of this example is to evaluate the performance of the same hedging strategies and same model estimation given the Bayesian estimates of parameter uncertainty and the

140

Volatility	Static	Dynamic	Indicator
EPV[Total Hedging Costs]	3.0082	2.6823	2.4142
90% CTE[Total Hedging Costs]	5.9424	5.9493	6.7918

Table 6.3: Total Hedging Cost Averages and Standard Errors for the Hedging Strategies for the S&P 500 Put Option Example Under Bayesian Data Simulation

corresponding impacts on the model.

Table 6.3 displays the expected hedging costs and the 90% CTE of the hedging cost distribution for each of the three replication strategies. All three methods were affected by the market parameter uncertainty not accounted for, though the indicator and dynamic methods were more affected than the static strategy. The hedging costs 90% CTE for the dynamic and indicator methods both increased by approximately 5% due to parameter uncertainty, whereas it only increased 2.3% under the indicator strategy. The general relationship between the three methods is still the same however, which is further depicted in Figure 6.4.

In terms of the effectiveness of the indicator and dynamic strategies over static strategies with the same level of risk or reward, both hidden Markov strategies still offer significant improvements over the static method, but not to the same degree as when there was no assumed parameter uncertainty (see Figure 6.5). The dynamic hidden Markov strategy offers now a 10.9% percent reduction in the 90% CTE hedging cost risk over the static strategy with the same expected costs, down from 16.7% in the previous section. Similarly, the reduction in the indicator method over the static strategy with the same average cost was 11.1%, down from 17.0%.

Overall, the results from the Bayesian example showed that the hidden Markov replication strategies are more sensitive to parameter risk than static replication strategies for a deep out-of-the-money put option on a model Bayesian-fitted to the S&P 500. However, both methods still offer a significant reduction in hedging costs and/or risk, and the indicator strategy was still shown to be cheaper and more risky than the dynamic strategy.
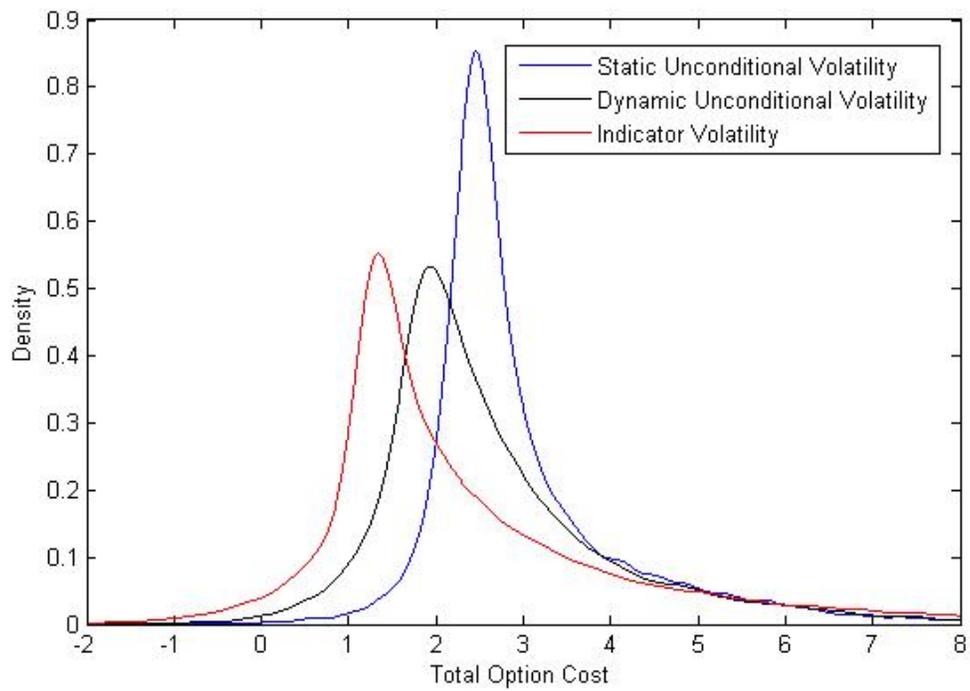
Figure 6.4: Total Hedging Cost Distributions for the Hedging Strategies for the S&P 500 Put Option Example Under Bayesian Data Simulation
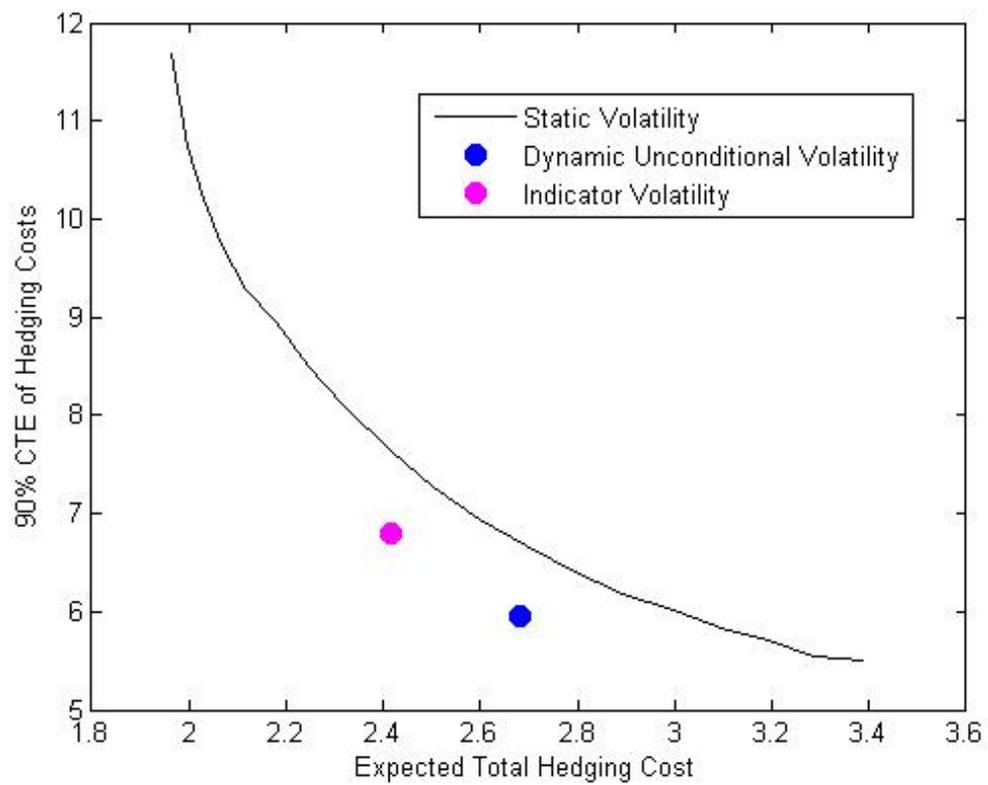
Figure 6.5: Total Hedging Cost Distributions for the Hedging Strategies for the S&P 500 Put Option Example Under Bayesian Data Simulation

6.5 Conclusion

This chapter presented two different delta-hedged portfolio replication strategies for put options, where the underlying equity price follows a hidden Markov distribution. The weighted average strategy averages state volatilities for input into the Black-Scholes framework using the data dependent underlying state probabilities of the estimated model. The indicator strategy uses the state specific volatilities, with transitions immediately when the data under the model indicate a probable state shift.

The strategies were tested for a long-term out-of-the-money put option, which resembles a long-term guaranteed minimum accumulation benefits guarantee, using the RSLN-2 model for the S&P 500 monthly data. Both options significantly outperformed static delta-hedging strategies. The weighted average strategy offered a very low risk option, while the indicator strategy was somewhat riskier, but offered an expected discount on total hedging costs. Incorporating parameter uncertainty into the analysis did not alter these overall results.

Chapter 7

Future Work

This Chapter will identify a few areas for further research that are natural extensions of the analysis presented in this thesis.

7.1 Relaxing the Constant Interest Rate Assumption

A focus of this thesis has been to obtain results and methods that would be of practical use for long-term equity analysis, in particular for insurance applications. Real world data was used for method and model comparisons. However, interest rates were held constant for the examples and markets calculations. For the GMAB example in Chapter 2, the risk free rate was held at 5%. The bond in Chapter 5 that was available for investment was assumed to yield a constant 6% per annum. And the risk free force of interest for the dynamic hedging example in Chapter 6 was also 5%.

A constant interest rate assumption is unrealistic. Moreover, since the general focus of this thesis was the long-term left tail of equities, strung together periods of negative returns are of concern. Such periods are often experienced at the same time as downturns in the economy, which also often prompt governments to lower interest rates for investment stimulation purposes.

A natural extension of the portfolio optimization and portfolio replication analyses would be to include floating interest rates. Perhaps variations of the Vasicek model (Vasicek, 1977) or the Cox-Ingersoll-Ross model (Cox, Ingersoll and Ross, 1985) for short term interest rates could be incorporated into the hidden Markov model framework for more realistic results for hidden Markov portfolio replication and optimization.

7.2   Developing further the Dynamic Hedging Analysis

Chapter 6 investigated dynamic hedging under a hidden Markov framework, and presented results for a deep out-of-the-money put option. The analysis in Chapter 6 could be easily adapted to study the effects of different types of options, such as calls, spreads, exotic options and lookbacks. In addition, moving from monthly data to higher frequency data could be considered to further investigate what the most beneficial hedging strategies are under a hidden Markov setting. In this sense, a mix of short-term and long-term hidden Markov dynamics could be integrated together, such that potential models could capture both short term and long term market states.

7.3   Bayesian Model Validation

The focus of Chapter 3 was validation of the maximum likelihood estimated hidden Markov models described in Chapter 2. A natural extension of this work would be a similar analysis for the Bayesian estimates of the models from Chapter 4.

Bayesian residual analysis (see Gelman, 2003, for example) even for single state models, encounters the same difficulties as hidden Markov analysis, in that parameter values are treated as random variables and thus residuals can not be read off as in the single state frequentist model case. Theorem 1 from Chapter 3 could be extended to broader classes of latent parameters.

Other forms of Bayesian model validation include posterior predictive checking (presented in

Rubin, 1981 and Rubin, 1984), involves replicating data multiple times under the model, and comparing the distribution of a statistic of interest from the simulations to that obtained from the model. For long-term equity models, the long-term left tail is often of primary concern. The oversampling technique used in Hardy, Freeland and Till (2006), which gains inference about the long-term left tail using a modified bootstrap technique, could be used for this type of analysis.

7.4   Extensions to other Hidden Markov Models

The hierarchical hidden Markov (HHM) class of models consists of models for which there are hidden Markov probabilistic state models within other probabilistic state models. When a state in the upper hierarchy is experienced under the model, it activates an embedded underlying state process, dependent on the upper hierarchy state. These models have only recently been proposed in a financial setting (Troiano and Kriplani, 2010), but have longer histories in the fields of speech recognition and computational molecular biology.

Theorem 1 from Chapter 3 could be adapted to the hierarchical hidden Markov model case to enable proper goodness-of-fit testing of HHM models through residual analysis. Bayesian analysis for single state hierarchical models and meta-analyses has been developed (De-Mouchel, 1990, for example), and could be adapted for a HHM setting. The potential gain in model goodness-of-fit for long-term equity data such as the S&P 500 over the simpler hidden Markov models used in this thesis would be an interesting analysis.

Theorem 1 from Chapter 3 could also be adapted to non-Gaussian hidden Markov models (for example, hidden Markov models with random innovations that are distributed according to a Student's t-distribution). Non-Gaussian innovations can represent a different way of capturing the data observations that don't fit within the simpler hidden Markov models, rather than adding additional model structure. The ability to adapt tests for single state models with the same respective random innovation processes to hidden Markov models could be prove valuable.

References

Akaike, H. (1974). "A new look at statistical model identification." *IEEE Trans Aut Control*, **19**, 716-723.

Alexander, S., T. Coleman and Y. Li. (2006). "Minimizing CVaR and VaR for a Portfolio of Derivatives." *Journal of Banking and Finance*, **20(2)**, 583-605.

American Academy of Actuaries (AAA), (2005). "Recommended Approach for Setting Regulatory Risk-Based Capital Requirements for Variable Annuities and Similar Products, Presented by the American Academy of Actuaries Life Capital Adequacy Subcommittee to the National Association of Insurance Commissioners Capital Adequacy Task Force (C3 Phase 2 Report)."
http://www.actuary.org/pdf/life/c3june05.pdf

Ang A., and G. Bekaert. (2004). "How Regime Affect Asset Allocation", *Financial Analysts Journal.* **60(2)**, 86-99.

Anscombe, F. J. (1961). "Examination of residuals", *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.* Volume 1, University of California Press, Berkeley.

Artzner, P., F. Delbaen, J.-M. Eber and D. Heath. (1999). "Coherent measures of risk." *Mathematical Finance*, **9(3)**, 203-228.

Bauwens, L., A. Preminger and J.V. Rombouts. (2006). "Regime switching GARCH models." CORE Discussion Paper, **11**, Louvain-La-Neuve.

Black, F. and M. Scholes. (1973). "The pricing of options and corporate liabilities." *Journal of Political Economy*, **80**, 637-654.

Bollerslev, T. (1986). "Generalized autoregressive condition heteroscedasticity." *Journal of Econometrics*, **31**, 307-327.

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

Boyle, P. and S.S. Liew. (2008). "Asset allocation with hedge funds on the menu." *North American Actuarial Journal*, **11(4)**, 1-22.

Campbell, J.Y., A.W. Lo and A.C. Mackinlay. (1997). *The Econometrics of Financial Markets.* Princeton: Princeton University Press.

Canadian Institute of Actuaries (CIA). (2001). "Report of the Task Force on Segregated Fund Investment Guarantees Canadian Institute of Actuaries."
http://www.actuaries.ca/publications/2002/202012e.pdf

Chib, S. (1996). "Calculating posterior distributions and modal estimates in Markov mixture models." *Journal of Econometrics*, **75**, 79-97.

Cox, J.C., J.E. Ingersoll and S.A. Ross. (1985). "A theory of the term structure of interest rates." *Econometrica*, **53**, 385-407.

DeMouchel, W.M. (1990). "Bayesian meta-analysis." In *Statistical Methodology in the Pharmaceutical Sciences*, ed. D.A. Barry. 509-529. Marcel Dekker. New York.

Diebold, F.X. (1986). "Modeling persistence in conditional variance: a comment." *Economic Reviews*, (5), 51-56.

Duan, Jun Chuan (1995) "The GARCH Option Pricing Model." *Mathematical Finance*, **5**, 13-32.

Engle, R.F. (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." *Econometrica*, **50**, 987-1006.

Engle, R.F. (Ed) (1995). *ARCH: Selected Readings*. Oxford University Press.

Gelman, A. (2003). "A Bayesian formulation of exploratory data analysis and goodness-of-fit testing." *International Statistical Review*. **71**, 369382.

Gelman, A., J.B. Carlin, H.S. Stern and D.B. Rubin. (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC. 2nd Edition.

Geman S., and D. Geman. (1984). "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

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

Hamilton, J.D. (1989), "A new approach to the analysis of non-stationary time series." *Econometrica*, **57**, 357-384.

Hamilton, J.D. and R. Susmel. (1994). "Autoregressive conditional heteroscedasticity and changes in regime." *Journal of Econometrics*, **64**, 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*, **3**, 185-211.

Hardy, M.R. (2003). *Investment Guarantees; Modeling And Risk Management For Equity Linked Life Insurance.* Wiley, New Jersey.

Hardy, M. R., R.K. Freeland and M. Till. (2006). "Validation of Long-Term Equity Return Models for Equity-Linked Guarantees." *North American Actuarial Journal*, **10**, 28-47.

Hastings, W.K. (1970). "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika*, **57**, 97-109.

Hull, J.C. (1989). *Options, futures and other derivative securities.* Prentice Hall, New Jersey.

Jarque, C.M. and A.K. Bera. (1980) "Efficient tests for normality, homoscedasticity and serial independence of regression residuals." *Economics Letters*, **6(3)**, 255-259.

Markowitz, H. (1952). "Portfolio selection." *Journal of Finance*, **7**, 77-91.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investment.* John Wiley & Sons, New York.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller. (1953). "Equation of state calculations by fast computing machines." *Journal of Chemical Physics*, **21**, 1087-1092.

Metropolis, N. and S. Ulam. (1949). "The Monte Carlo Method." *Journal of the American Statistical Association*, **44**, 335-341.

Morgan, J.N. (1954). "Analysis of Residuals from Normal Regressions." *Contributions of Survey Methods to Economics.* Columbia University Press.

Panjer, H. et al. (1998). *Financial Economics: With Applications to Investments, Insurance and Pensions.* Schaumburg, IL: Actuarial Foundation.

Panneton, C.M. (2004). "Mean-Reversion in Equity Models in the Context of Actuarial Provisions for Segregated Fund Investment Guarantees." *Segregated Funds Symposium Proceedings*, Canadian Institute of Actuaries. (Revised)

Robert, C.P., G. Celeux, and J. Diebolt. (1993). "Bayesian estimation of Hidden Markov Models: a stochastic implementation." *Statistics and Probability Letters*, **16(1)**, 77-83.

Robert, C.P., T. Ryden and D.M. Titterington. (1999). "Convergence controls for MCMC algorithms, with applications to hidden Markov chains." *Journal of Statistical Computation and Simulation*, **64**, 327-355.

Robert, C.P. and D.M. Titterington. (1998). "Reparameterization strategies for hidden Markov models and Bayesian approaches to maximum likelihood estimation." *Statistics and Computing*, **8**, 145-158.

Rubin, D.B. (1981). "Estimation in parallel randomized experiments." *Journal of Educational Statistics*, **6**, 377-401.

Rubin, D.B. (1984). "Bayesianly justifiable and relevant frequency calculations for the applied statistician." *Annals of Statistics*, **12**, 1151-1172.

Schwarz, G. (1978). "Estimating the dimension of a model." *Annals of Statistics*, **6**, 461-464.

Shephard, N. (1994). "Partial non-Gaussian state space." *Biometrika*, **81**, 115-131.

Troiano, L. and P. Kriplani. (2010). "Predicting trend in the next-day market by Hierarchical Hidden Markov Model." Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on. 199-204.

Vasicek, O. (1977). "An equilibrium characterisation of the term structure." *Journal of Financial Economics*, **5**, 177-188.

Wong, A.C.S. and W.S. Chan. (2005). "Mixture Gaussian Time Series Modeling of Long-Term Market Returns." *North American Actuarial Journal*, 9(4).

Zhou, X.Y. and G. Yin. (2003), "Markowitz's Mean-Variance Portfolio Selection with Regime Switching: A Continuous-Time Model", *Siam Journal on Control and Optimization*, **42**, 1466-1482.

Yin, G. and X.Y. Zhou (2004), "Markowitz's mean-variance portfolio selection with regime switching: from discrete-time models to their Continuous Limits", *IEEE Transactions on Automatic Control.*, **49(3)**, 349-360.

Appendix A

Posterior Convergence Assessment

For the Bayesian estimation of the hidden-Markov models in Chapter 4, a stationarity test (Robert et. al, 1999) was used to assess convergence of the posterior distributions. To perform the test, one first splits the chain

$$\Theta^{(b+1)}, \ldots, \Theta^{(S)}$$

into two sub-chains of equal size,

$$\Theta^{(b+1)}, \ldots, \Theta^{(\frac{S-b}{2})} \qquad \text{and} \qquad \Theta^{(\frac{S-b}{2}+1)}, \ldots, \Theta(S)$$

From both sub-chains, further sub-samples of size $G = G(S)$ are drawn. For each parameter θ_d, the test compares the empirical cumulative distribution functions of the sub-samples of size G with a Kolmogorov-Smirnov two-distribution test. The test statistic for each parameter is defined by

$$KS_d = \frac{1}{G} \max_{\eta} \left| \sum_{g=1}^{G} \left(I_{\theta_d^{(g)} \leq \eta} - I_{\theta_{2d}^{(g)} \leq \eta} \right) \right|$$

Parameter	K-G Statistic	P-value
$p_{1,2}$	0.0390	0.0932
$p_{2,1}$	0.0250	0.5545
μ_1	0.0375	0.1175
σ_1	0.0305	0.3056
μ_2	0.0270	0.4545
σ_2	0.0440	0.0404

Table A.1: Posterior Stationarity Test Results for the RSLN-2 Model

The p-values of the test statistics $KS_d, d \in \{1, \ldots, D\}$ are the p-values of the standard Kolmogorov-Smirnov test if $G(S) \in o(S)$, or more explicitly,

$$\lim_{S \to \infty} \frac{G(S)}{S} = 0$$

The term 'stationarity test' here can be misleading. There is some literature that disputes whether stationarity can be assessed from a single Markov chain (see, for example, Gelman and Rubin (1992)). The main area of danger is that there may be local areas the chain experiences for extended periods of time before moving to its stationary distribution. If the chain moves quickly to one such local area of attraction and remains in it until stopping time S, the chain would pass the stationarity test despite not yet reaching the stationary distribution of the posterior. The test therefore must be used with caution. If the test fails, then an acceptable conclusion is that the chain needs to continue to run. If the test passes, then one can conclude that the chain has at least reached a node of *near*-stationarity.

The number of simulations performed for each of the models was $21,000$. After the burn-in period of $1,000$ was removed, the remaining $20,000$ observations were split into two sub-chains. From each of these, $2,000$ observations for each parameter were simulated at random, without replacement, to form the two sets that were compared with the Kolmogorov-Smirnov test. The stationarity test results from the popular long-term equity hidden Markov models are displayed in the tables that follow.

Parameter	K-G Statistic	P-value
$p_{1,2}$	0.0512	0.1011
$p_{2,1}$	0.0460	0.2350
μ_1	0.0356	0.4512
φ_1	0.0290	0.7888
σ_1	0.0477	0.1956
μ_2	0.0400	0.3935
φ_2	0.0430	0.3072
σ_2	0.0660	0.0244

Table A.2: Posterior Stationarity Test Results for the RSDD-2 Model

Parameter	K-G Statistic	P-value
$p_{1,2}$	0.0349	0.5723
$p_{2,1}$	0.0630	0.0361
μ_1	0.0380	0.4586
β_1	0.0554	0.0921
μ_2	0.04919	0.1683
β_2	0.0310	0.7161

Table A.3: Posterior Stationarity Test Results for the RSGARCH Model

Parameter	K-G Statistic	P-value
P	0.0380	0.1089
μ_1	0.0560	0.0036
α_1	0.0165	0.9466
α_2	0.0160	0.9586
α_3	0.0225	0.6873
μ_2	0.0365	0.1364
σ_2	0.0225	0.6873

Table A.4: Posterior Stationarity Test Results for the MARCH Model