Accounting for Regime and Parameter Uncertainty in Regime-Switching Models

Brian M. Hartman*

University of Waterloo, Department of Statistics and Actuarial Science, 200 University Avenue West, Waterloo, ON, Canada N2L 3G1, 519.888.4567 ext. 35506, Fax: 519.746.1875

Matthew J. Heaton

Duke University, Department of Statistical Science, Box 90251, Durham, NC, United States 27708-0251

Abstract

As investment guarantees become increasingly complex, realistic simulation of the price becomes more critical. Currently, regime-switching models are commonly used to simulate asset returns. Under a regime switching model, simulating random asset streams involves three steps: (i) estimate the model parameters given the number of regimes using maximum likelihood, (ii) choose the number of regimes using a model selection criteria, and (iii) simulate the streams using the optimal number of regimes and parameter values. This method, however, does not properly incorporate regime or parameter uncertainty into the generated asset streams and therefore into the price of the guarantee. To remedy this, this article adopts a Bayesian approach to properly account for those two sources of uncertainty and improve pricing.

Keywords:
JEL: G17, C11
IB10, IM22, Asset Price Simulation, Bayesian Modeling, Dirichlet process

1. Introduction

In pricing increasingly complex insurance guarantees, closed-form analytical solutions are either difficult or impossible to find. Monte Carlo simulation methods enable the individual pieces of the guarantee (asset prices, mortality, interest rates, etc.) to be simulated. Using these simulations, the price of the guarantee can be calculated. However, price calculations are only as accurate as

*Corresponding author
Email addresses: bhartman@uwaterloo.ca (Brian M. Hartman), matt@stat.duke.edu (Matthew J. Heaton)

the models and algorithms which generate the individual pieces. In this paper, we will focus on simulating the asset price.

When deciding on a model for the asset price, a standard starting point is geometric Brownian motion with a constant variance. Geometric Brownian motion, however, is often too simplistic for asset prices. For instance, Bakshi et al. (1997) show that incorporating stochastic volatility and jumps is important for pricing and internal consistency. Regime-switching models are a simple and intuitive way to enable stochastic volatility. Regime-switching models in economic series were first introduced by Hamilton (1989) and further analyzed in Hamilton and Susmel (1994). These models gained prominence in the actuarial literature when Hardy (2001) compared many different models and found that the two regime lognormal model maximized the Schwarz-Bayes Criterion (Schwarz, 1978) for monthly S&P 500 total return data and monthly TSE 300 total return data (On May 1, 2002 Standard and Poor's began managing the TSE 300 and changed the name to S&P/TSX Composite Index. To prevent confusion, we will refer to the index as TSX for the remainder of the paper). According to the Akaike Information Criterion (Akaike, 1974), the two regime model was found to be optimal in the TSX data, but the three regime model performed the best on the S&P data. Hardy (2003) notes that this shows that model selection is not often clear cut and, while the results should inform the decision, there is room for judgement as well.

The primary contribution of this paper is to present a Bayesian method to quantify the uncertainty associated with the number of regimes in regime-switching (RS) models. While we focus on the regime-switching lognormal (RSLN) model, the methods described herein can be easily adapted to other RS models. Hardy (2002) offered a first look at how to estimate the RSLN model in a Bayesian framework using the Metropolis-Hastings algorithm. Specifically, Hardy (2002) analyzed the effect of parameter uncertainty on equity-linked insurance contracts. We build on that work by incorporating regime uncertainty. Additionally, we propose methods to incorporate parameter uncertainty which achieve superior mixing properties and flexibility in the specification of the prior distributions.

By operating in a Bayesian framework, we are also able to obtain posterior probabilities for each model. A manager can use those probabilities to make an informed decision on which model to use. Furthermore, the probabilities allow for model averaging when producing the simulated asset streams. For example, if the posterior probabilities were 0.05, 0.80, and 0.15 for one, two, and three regimes, respectively, to make 100 simulated streams incorporating the model uncertainty 5, 80, and 15 streams would be generated from the one, two, and three regime models, respectively.

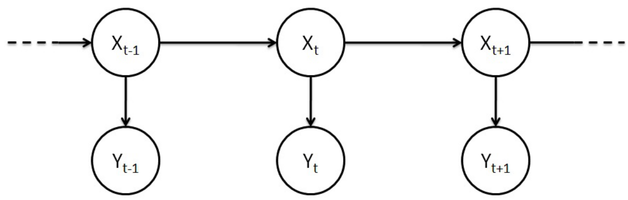
By accounting for both the uncertainty in the parameters and model specification, we are able to more clearly understand the asset distribution, especially the tails of this distribution. We find that properly incorporating parameter uncertainty has a large effect on the asset distribution while incorporating the regime uncertainty has a smaller, yet significant, effect. Additionally, our results indicate a more definitive decision rule between a two- and three-regime model for monthly S&P return data. We find that a large number of regimes is required to properly fit daily data. Finally, our methods allow us to thoroughly investigate moments of the asset distribution for each individual time period and compare them to generally accepted economic events (National Bureau of Economic Research, 2010).

The rest of the paper is organized as follows. Section 2 describes the RSLN model. Section 3 describes two methods for when the number of regimes is assumed fixed and known. Section 4 relaxes that assumption and estimates both the number of regimes and the model parameters. In section 5, we describe a simulation study which illustrates the advantages of our method. Section 6 describes the effect of the method on a few total return series from the S&P 500 and TSX. Finally, some conclusions are available in section 7.

2. Regime-switching lognormal (RSLN) model

In a Markov regime-switching model with R regimes, the current regime, $x_t \in \{1, \dots, R\}$, only depends upon the previous regime, x_{t-1}. This is known as the Markov property. Additionally, the regime is often unobservable. These models are also known as hidden Markov models (HMMs). The observation, y_t, depends only upon the current regime, x_t (Figure 1).

Figure 1: Graphical representation of a regime-switching model



There are three main parameters of interest in an RSLN model with R regimes. $\boldsymbol{\mu} = (\mu_1, \dots, \mu_R)$, the vector of means (one for each regime), $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_R^2)$, the vector of variances, and $\boldsymbol{\theta}$, an R by R matrix where θ_{ij} is the probability of moving from regime i to regime j. To enable efficient estimation, we augment the data by explicitly estimating the regime (x_i) for each observation. For ease of notation and computation, we exponentiate all the observations so the model is now a regime-switching normal model. The estimation of the posterior distributions of the $\boldsymbol{\mu}$, $\boldsymbol{\sigma}^2$, and $\boldsymbol{\theta}$ parameters are relatively straightforward because given the x_i vector, the transformed data values are independent and normally distributed with the following density function

$$p(y_i|\mu_r, \sigma_r^2, x_i = r) = \frac{1}{\sqrt{2\pi\sigma_r^2}} \exp\left\{-\frac{(y_i - \mu_r)^2}{2\sigma_r^2}\right\}. \tag{1}$$

3. Fixed number of regimes

We first assume that the number of regimes is known a priori. While this is not very likely in practice, we are able to investigate the benefit of accounting for only the parameter uncertainty.

3.1. Regime-specific parameters

Because the observations are independent and normally distributed, a natural choice for the prior distributions of μ and σ^2 is normal-gamma.

$$p(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \prod_{r=1}^{R} N(\mu_r|\mu_0, \sigma_r^2/n_0) Ga(1/\sigma_r^2|\alpha_0, \beta_0). \tag{2}$$

Alternatively, to assist in the analysis of the posterior simulations, we can constrain the $\boldsymbol{\mu}$ vector to be ordered.

$$p(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) \propto \prod_{r=1}^{R} N(\mu_r|\mu_0, \sigma_r^2/n_0) Ga(1/\sigma_r^2|\alpha_0, \beta_0) 1_{\mu_1 < \mu_2 < \ldots < \mu_R} \tag{3}$$

Constraining the $\boldsymbol{\mu}$ vector to be ordered will prevent label switching. If exchangeable prior distributions are placed on the parameters, the posterior distribution will be invariant to permutations in the labeling of the parameters (Jasra et al., 2005). Each row of the transition matrix is given a Dirichlet prior.

$$p(\theta_r) = Dir(1, 1, \ldots, 1) \quad \forall \quad r \in \{1, 2, \ldots, R\} \tag{4}$$

Finally, the prior on the state vector (x_i) will give equal weight to each state.

Given the state vector, the full conditional posterior distributions for $\boldsymbol{\mu}$, $\boldsymbol{\sigma}^2$, and $\boldsymbol{\theta}$ are tractable.

$$p(\mu_r|\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y}) = St\left(\mu_{post}, (n_r + n_0)(\alpha_0 + n_r/2)\beta_{post}^{-1}, 2\alpha_0 + n_r\right) \tag{5}$$

$$\mu_{post} = (n_0 + n_r)^{-1}(n_0\mu_0 + \sum(y_i 1_{x_i=r})) \tag{6}$$

$$\beta_{post} = \beta_0 + \sum_{i:x_i=r} (y_i - \bar{y}_r)^2/2 + (n_0 + n_r)^{-1} n_0 N_r (\mu_0 - \bar{y}_r)^2/2 \tag{7}$$

$$p(1/\sigma_r^2|\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y}) = Ga(\alpha_0 + n_r/2, \beta_{post}) \tag{8}$$

$$p(\theta_r|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{x}, \boldsymbol{y}) = Dir(1 + n_{r1}, 1 + n_{r2}, \ldots, 1 + n_{rR}) \tag{9}$$

3.2. State vectors

There are two main methods for estimating the state vector (x_i) given the other parameters. They could be generated one observation at a time or all at once. Estimating the states simultaneously should allow the chain to mix more quickly requiring fewer iterations, but each iteration is more computationally intense.

3.2.1. Single-site updating

When estimating each observation individually, the full conditional posterior distribution of each state value is

$$p(x_i|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \boldsymbol{y}_{1:n}, \boldsymbol{x}_{1:i-1}, \boldsymbol{x}_{i+1:n}) \tag{10}$$

which reduces through the Markov property to

$$p(x_i|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, y_i, x_{i-1}, x_{i+1}) \propto p(x_i)q(x_{i-1}, x_i)q(x_i, x_{i+1})p(x_i|y_i) \tag{11}$$

where when $i = 1$, $q(x_0, x_1)$ is replaced by $\nu(x_1)$, the unconditional probability of regime x_1, and when $i = n$, $q(x_n, x_{n+1})$ is replaced by 1. For each iteration of the sampler, a state value is drawn for each observation, either in sequence or a random order.

3.2.2. Forward filtering, backward sampling

Alternatively, the entire state vector can be simultaneously updated through a forward filtering, backward sampling algorithm (Rabiner, 1989).
Step 1: Forward Filtering

$$\phi_{1|0}(j) = \nu(j) \quad \forall \quad j \in \{1, \ldots, R\} \tag{12}$$

For $i \in \{1, \ldots, n\}$

$$c_i = \sum_{r=1}^{R} \phi_{i|i-1}(r)p(x_i = r|y_i) \tag{13}$$

$$\phi_i(j) = \frac{\phi_{i|i-1}(j)p(x_i = j|y_i)}{c_i} \quad \forall \quad j \in \{1, \ldots, R\} \tag{14}$$

$$\phi_{i+1|i}(j) = \sum_{r=1}^{R} \phi_i(r)q_{rj} \quad \forall \quad j \in \{1, \ldots, R\} \tag{15}$$

Step 2: Backward Sampling After the ϕ_i values are generated, the states are drawn with the following probabilities:

$$p(X_n = x) = \phi_n(x) \tag{16}$$

For $i \in \{n-1, \ldots, 1\}$

$$p(X_i = x|X_{i+1} = y) = \frac{\phi_n(x)q_{xy}}{\sum_{r=1}^{R} \phi_n(r)q_{ry}} \tag{17}$$

4. Dynamic number of regimes

While performing the estimation with a fixed number of states allows for easier implementation and possibly more coherent estimation of the model parameters, the uncertainty about the true number of regimes is not included.

4.1. Reversible jump MCMC

Generalization of Markov chain Monte Carlo methods to parameter vectors of varying dimension was introduced through reversible jump MCMC (Green, 1995). Robert et al. (2000) applied the RJMCMC method to regime switching models, but was only able to allow the variances to vary, the means were forced to be set at 0. One large difficulty with the implementation of RJMCMC is the specification of the proposal distribution. To our knowledge, there is not yet an effective way to specify that distribution for general regime-switching models.

4.2. Dirichlet process

Dynamic estimation of the number of states in an HMM has received attention in the machine learning literature. One stream of work is based on the Dirichlet process.

The existence of a Dirichlet process was established by Ferguson (1973). Assume $(\Theta, \mathcal{B})$ is a measurable space with probability measure G_0 and α_0 is a positive real number. A Dirichlet process $DP(\alpha_0, G_0)$ is defined to be the distribution of a random probability measure G over $(\Theta, \mathcal{B})$ such that, for any finite measurable partition $(A_1, A_2, \ldots, A_r)$ of Θ, the random vector $(G(A_1), \ldots, G(A_r))$ is distributed as a finite-dimensional Dirichlet distribution with parameters $(\alpha_0 G_0(A_1), \ldots, \alpha_0 G_0(A_r))$. If G is a random probability measure with distribution given by the Dirichlet process it is written as $G \sim DP(\alpha_0, G_0)$.

One perspective on the Dirichlet process provided by the Pólya urn scheme (Blackwell and MacQueen, 1973) showed that draws from the Dirichlet process are discrete and exhibit a clustering property. This perspective led to another similar metaphor called the Chinese restaurant process (Aldous, 1985).

Imagine a Chinese restaurant with an infinite number of tables (clusters or states) and the capacity to make an infinite number of distinct dishes (parameter values). Customers (data points) enter one at a time. The first customer sits at the first table. Subsequent customers will join an already-occupied table i with probability proportional to n_i, the number of customers sitting at table i. They could also sit at an empty table with probability proportional to α_0. All of the customers at a single table must eat the same meal (have the same parameter value).

After all the customers have entered the restaurant, they will each, in turn, be removed from their seat and reseated according to the probabilities above (proportional to n_i for an occupied table and proportional to α_0 for an unoccupied table). Also, the meals will be updated taking into account the preferences of the customers sitting at the table.

4.3. Infinite hidden Markov model

Beal et al. (2002) were the first to use the Dirichlet process in an HMM framework through an Infinite HMM (iHMM). At any given iteration of their sampler, let $\{1, \ldots, R\}$ be the current number of states occupied. The states are drawn from the following probability:

$$P(x_{t+1} = j | x_t = r, \beta) = \frac{n_{rj}}{\sum_{j'=1}^{R} n_{rj'} + \beta} \quad r \in \{1, \ldots, R\}. \tag{18}$$

Notice that the probabilities do not sum to one. With probability

$$\frac{\beta}{\sum_{j'=1}^{R} n_{ij'} + \beta} \tag{19}$$

an oracle state is chosen which has the ability to visit any state (not just those which have been visited from the current state before) or generate an entirely new state.

While this method is very intuitive, posterior inference is rather awkward and in the paper only a heuristic approximation to a Gibbs sampler is proposed.

4.4. HDP-HMM

Teh et al. (2006) defined a hierarchical Dirichlet process which is analogous to a Chinese restaurant franchise. The franchise has an infinite number of restaurants, each with an infinite number of tables, and a global menu to choose from. The first person sits at a table and chooses the meal. Having multiple restaurants allows the probability that a certain table/meal is chosen to vary depending on the restaurant. In a regime-switching context, observation i chooses a table. Observation $i+1$ will enter the restaurant with the same index as the table chosen by observation i. The options available to observation $i+1$ and their associated probabilities are dependent upon the choices of observation i (the Markov property). This implies a doubly infinite transition matrix where each row is a restaurant. They named the method the hierarchical Dirichlet process hidden Markov model, or HDP-HMM.

4.5. Sticky HDP-HMM

Many regime-switching models fit to real asset data, including those in Hardy (2001) and Hardy (2003), exhibit state persistence. In the HDP-HMM all moves are equally likely a priori. This can discourage parsimony. Fox et al. (2007) overcame this problem by explicitly specifying a parameter which defines the increase in prior probability for a self-transition over any other state. They named their method the sticky HDP-HMM.

This can be thought of as a Chinese restaurant franchise with loyal customers. The restaurant with the same index as a dish makes it best. When an observation enters a restaurant, the namesake dish is more popular there than anywhere else.

This also creates family loyalty. If a grandparent selects dish i, then the parent will go into restaurant i making it more likely that he will also try dish i. The details of the sampling algorithm are as follows with $\delta(j, k)$ denoting the number of transitions from j to k and $f_r(y_i) = N(y_i|\mu_r, \sigma_r^2)$:

Given the previous state assignments $x_{1:I}^{(n-1)}$, global transition distribution $\beta^{(n-1)}$, and hyperparameters α and κ:

1. Set $x_{1:I} = x_{1:I}^{(n-1)}$ and $\beta = \beta^{(n-1)}$. For each $i \in \{1, \dots, I\}$, sequentially
   (a) Decrement $n_{x_{i-1}x_t}$ and $n_{x_i x_{i+1}}$ and remove y_i from the cached statistics (mean $\hat{\mu}_r$, variance $\hat{\sigma}_r^2$, and sample size ν_r) for the current assignment $x_i = r$.
   (b) For each of the R currently instantiated states where $x_{i-1} \neq r$, determine

$$f_r(y_i) = (\alpha\beta_r + n_{x_{i-1}}r) \left(\frac{\alpha\beta_{x_{i+1}} + n_{rx_{i+1}} + \kappa\delta(r, x_{i+1})}{\alpha + n_{r\cdot} + \kappa} \right) t_{\nu_r}(y_i; \hat{\mu}_r, \hat{\sigma}_r^2) \quad (20)$$

   For $x_{i-1} = r$

$$f_r(y_i) = (\alpha\beta_r + n_{rr}^{-t} + \kappa\delta(r, r)) \left(\frac{\alpha\beta_r + n_{rr}^{-i} + \kappa\delta(r, y_{x_{i+1}}) + \delta(r, r)\delta(r, y_{x_{i+1}})}{\alpha + n_{r\cdot}^{-i} + \kappa + \delta(r, r)} \right) t_{\nu_r}(y_i; \hat{\mu}_r, \hat{\sigma}_r^2) \quad (21)$$

   Also, for a new state $R + 1$

$$f_{R+1}(y_i) = \frac{\alpha^2 \beta_{\bar{r}} \beta_{x_{i+1}}}{\alpha + \kappa} t_{\nu_{R+1}}(y_r; \hat{\mu}_{R+1}, \hat{\sigma}_{R+1}^2) \quad (22)$$

   (c) Sample the new state assignment x_i:

$$x_i \sim \sum_{r=1}^{R} f_r(y_i)\delta(x_i, r) + f_{R+1}(y_i)\delta(x_i, R + 1) \quad (23)$$

   If $x_i = R + 1$, then increment R and transform β as follows. Sample $b \sim \text{Beta}(1, \gamma)$ and assign $\beta_{R+1} \leftarrow b\beta_{\bar{r}}$ and $\beta_{\bar{r}} \leftarrow (1 - b)\beta_{\bar{r}}$, where $\beta_{\bar{r}} = \sum_{r=R+1}^{\infty} \beta_r$.
   (d) Increment $n_{x_{i-1}x_i}$ and $n_{x_i x_{i+1}}$ and add y_i to the cached statistics ($\hat{\mu}_r$, $\hat{\sigma}_r^2$, and ν_r) for the new assignment $x_i = r$.
2. Fix $x_{1:I}^{(n)} = x_{1:I}$. If there exists a j such that $n_{j\cdot} = 0$ and $n_{\cdot j} = 0$, remove j and decrement R.
3. Sample auxiliary variables m, w, and $\bar{m}$ as follows:
   (a) For each $(j, k) \in \{1, \dots, R\}^2$, set $m_{jk} = 0$ and $n = 0$. For each customer in restaurant j eating dish k, that is for $i = 1, \dots, n_{jk}$, sample

$$x \sim \text{Ber}\left(\frac{\alpha\beta_k + \kappa\delta(j, k)}{n + \alpha\beta_k + \kappa\delta(j, k)} \right) \quad (24)$$

   Increment n, and if $x = 1$ increment m_{jk}.

(b) For each $j \in \{1, \ldots, R\}$, sample the number of override variables in restaurant j:

$$w_{j\cdot} \sim \text{Binomial}\left(m_{jj}, \frac{\rho}{\rho + \beta_j(1-\rho)}\right), \tag{25}$$

set the number of informative tables in restaurant j considering dish r to:

$$\bar{m}_{jr} = \begin{cases} m_{jr}, & j \neq r; \\ m_{jj} - w_{j\cdot}, & j = r. \end{cases} \tag{26}$$

4. Sample the global transition distribution from

$$\beta^{(n)} \sim \text{Dir}(\bar{m}_{\cdot 1}, \ldots, \bar{m}_{\cdot R}, \gamma) \tag{27}$$

5. Simulation study

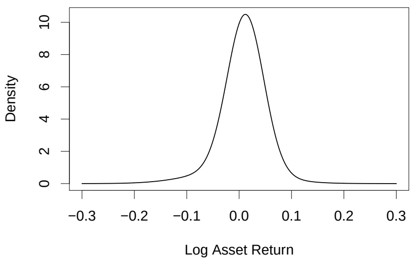
To exemplify the effect of properly accounting for the parameter and regime uncertainty, we performed a simulation study. We generated 100 datasets of 1,000 observations from each of six RSLN models drawn from the actuarial literature. The parameters are specified in table 1. The densities for each case are plotted in figure 2. Each case is described below.

1. Fitted values for the optimal model (RSLN-2) for the TSX data in Hardy (2001)
2. Similar to case 1, but with no Markov dependence
3. Similar to case 1, but the two regimes have equal mean
4. Similar to case 1, but with twice the difference between the means
5. Similar to case 1, but with only one regime
6. Similar to case 1, but with a rare, high expected value regime added, making three total regimes

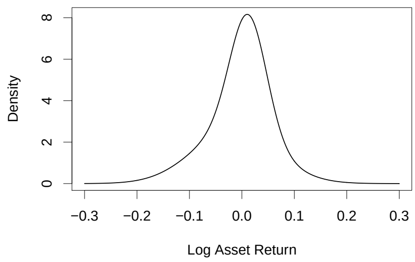
While there is some more information in the actual dataset (the order of the observations helps to inform the estimation) it appears from the plots of the densities that it will be very difficult to determine that more than one regime exists. There is some skewness and kurtosis, especially in cases 2 and 4, which may help. The posterior probabilities (estimated using the sticky HDP-HMM methodology) of each model is in Table 2.

In all cases, the posterior probability for the one-regime model is largest. Even so, there is still a large probability for the two regime model and non-negligible probabilities for three, four, and five regimes. As we would expect case 5, which actually has only one regime, has the highest probability for the one regime model. Case 4, which has the largest difference between the two regime means, has the highest probability for the remaining models. Notice also that the probabilities for case 6 are very similar to case 1. While case 6 does have a third regime, its stationary probability is only 0.01, having very

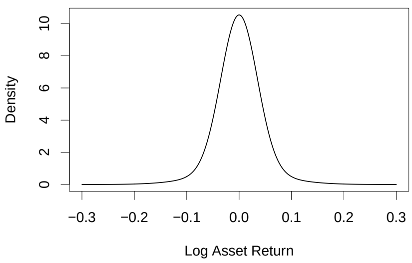
Figure 2: Simulation densities



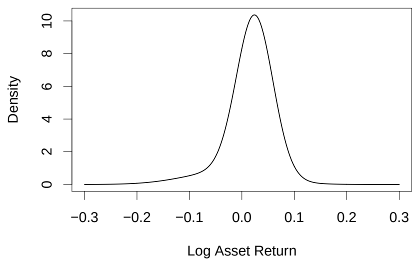
(a) Case 1



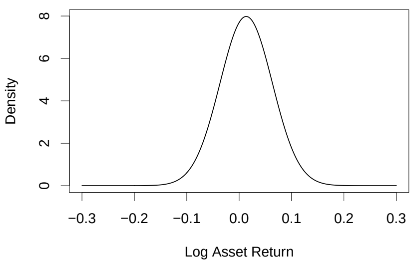
(b) Case 2



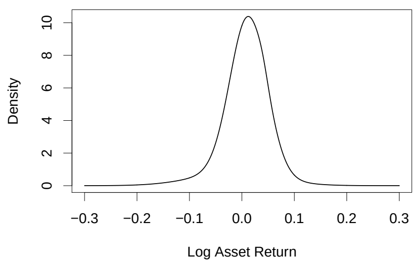
(c) Case 3



(d) Case 4



(e) Case 5



(f) Case 6

Table 1: Parameters for simulated datasets

Case	Regime 1 (μ_i, σ_i)	Regime 2 (μ_i, σ_i)	Regime 3 (μ_i, σ_i)	Transition Matrix
1	(0.012,0.035)	(-0.016,0.078)	-	$\begin{pmatrix} 0.963 & 0.037 \\ 0.210 & 0.790 \end{pmatrix}$
2	(0.012,0.035)	(-0.016,0.078)	-	$\begin{pmatrix} 0.500 & 0.500 \\ 0.500 & 0.500 \end{pmatrix}$
3	(0.000,0.035)	(0.000,0.078)	-	$\begin{pmatrix} 0.963 & 0.037 \\ 0.210 & 0.790 \end{pmatrix}$
4	(0.025,0.035)	(-0.031,0.078)	-	$\begin{pmatrix} 0.963 & 0.037 \\ 0.210 & 0.790 \end{pmatrix}$
5	(0.014,0.050)	-	-	-
6	(0.012,0.035)	(-0.016,0.078)	(0.040,0.010)	$\begin{pmatrix} 0.953 & 0.037 & 0.010 \\ 0.210 & 0.780 & 0.010 \\ 0.800 & 0.190 & 0.010 \end{pmatrix}$

Table 2: Posterior probabilities for the number of regimes

Case	Number of Regimes						
	1	2	3	4	5	6	7
1	0.647	0.214	0.088	0.034	0.013	0.004	0.001
2	0.789	0.157	0.038	0.011	0.003	0.001	0
3	0.629	0.221	0.098	0.036	0.012	0.003	0.001
4	0.518	0.289	0.122	0.047	0.017	0.005	0.001
5	0.864	0.109	0.02	0.004	0.002	0	0
6	0.641	0.203	0.094	0.039	0.014	0.005	0.002

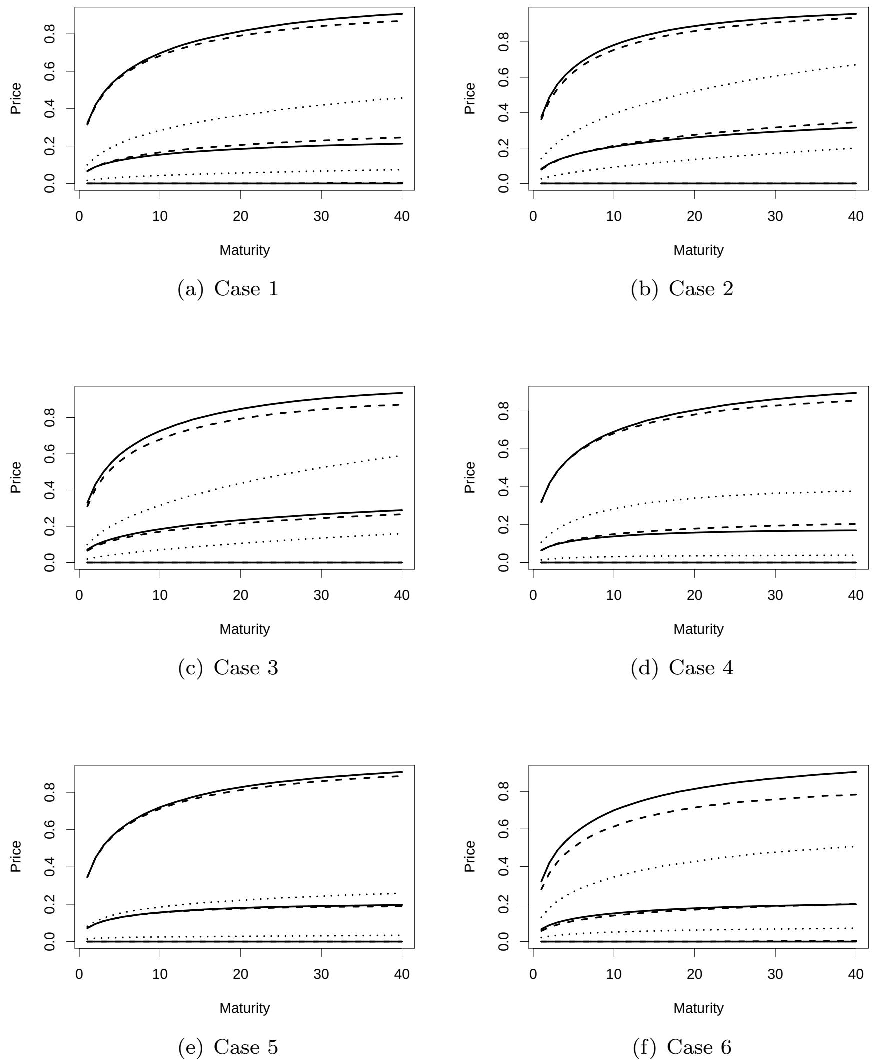
little effect on the overall distribution. Overall, there is a lot of uncertainty about the number of regimes to use when fitting this data.

For each of the simulated datasets, the model is fit using maximum likelihood (Method A), the single-site updating scheme of section 3 to incorporate the parameter uncertainty (Method B), and the sticky HDP-HMM method of section 4 to incorporate both the parameter and regime uncertainty (Method C). The results are then used to price a two European put options of various maturity lengths. Figure 3 shows the median price and 95% pointwise credible intervals for the price of an at-the-money put option ($S_0 = K = 1$). Figure 4 shows the same results for a put option which is deep in-the-money ($S_0 = 1; K = 10$). Method A is plotted with the dotted lines, method B with the dashed lines, and method C with the solid lines.

A few facts are readily apparent from figures 3 and 4. The largest discrepancy between the methodologies comes in the credible intervals. Method C provides

Figure 3: Median price and 95% pointwise credible intervals for at-the-money ($S_0 = K = 1$) put option prices

······	Method A	Maximum Likelihood
- - - -	Method B	Single-site Updating, parameter uncertainty only
———	Method C	Sticky HDP-HMM, regime and parameter uncertainty



(a) Case 1

(b) Case 2

(c) Case 3

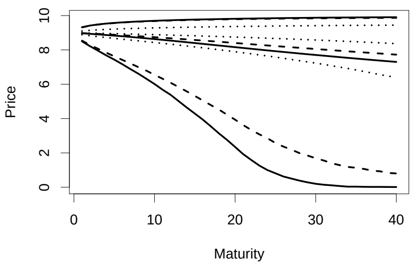
(d) Case 4

(e) Case 5

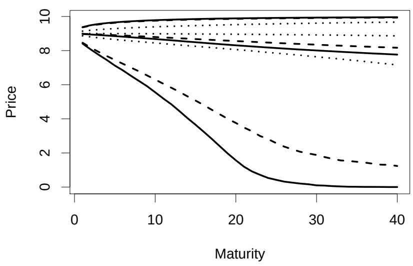
(f) Case 6

Figure 4: Median price and 95% pointwise credible intervals for deep-in-the-money ($S_0 = 1; K = 10$) put option prices

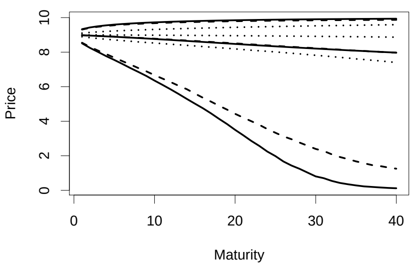
······	Method A	Maximum Likelihood
- - - -	Method B	Single-site Updating, parameter uncertainty only
———	Method C	Sticky HDP-HMM, regime and parameter uncertainty



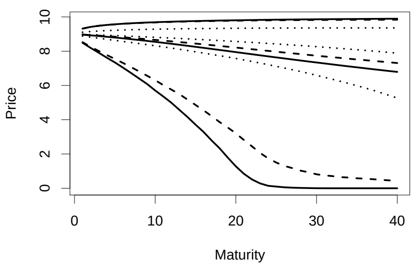
(a) Case 1



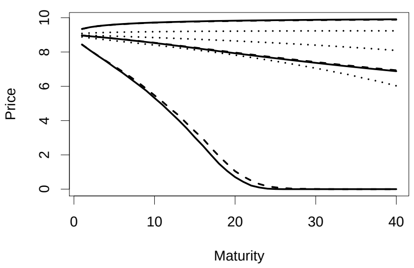
(b) Case 2



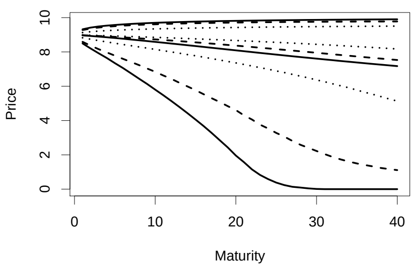
(c) Case 3



(d) Case 4



(e) Case 5



(f) Case 6

the widest intervals followed closely by Method B. It is interesting to note that the difference when accounting for parameter uncertainty is added (Method A to Method B) is much larger than the difference when the regime uncertainty is added (Method B to Method C). In all cases, the medians for methods B and C are very similar to each other and far from the median for method A.

6. Real data examples

In this section we describe the results of applying the sticky HDP-HMM to four datasets; monthly data from January 1956 through October 2010 (657 total observations) and daily data from 02 January 1981 through 12 November 2010 (7,526 total observations) for both the S&P 500 and the TSX. We obtained the data by comparing multiple sources (Standard and Poor's, 2010; Yahoo! Inc., 2010; Datastream International, 2010; Shiller, 2005).

6.1. Number of regimes

For each dataset, the sticky HDP-HMM methodology was run with 5,000 burn-in iterations and then the posterior distributions were estimated from 5,000 samples. The posterior probabilities for the order of the model are presented in table 3.

Table 3: Posterior probabilities for the number of regimes

Frequency	Index	Number of Regimes							
		1	2	3	4	5	6	7	8
Monthly	S&P 500	-	-	**0.56**	0.31	0.11	0.02	-	-
	TSX	-	**0.63**	0.29	0.08	0.01	-	-	-
Daily	S&P 500	-	-	-	-	-	0.38	**0.57**	0.05
	TSX	-	-	-	-	-	0.32	**0.63**	0.05

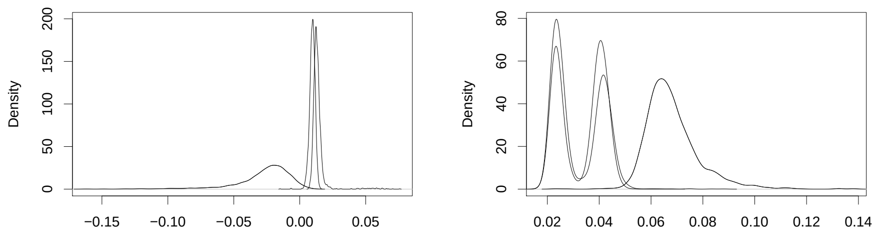
The two-regime model was selected for the monthly data of the TSX. That matches what was found in Hardy (2001). Even so, the probability of the two-regime model being the best is only 0.627 meaning there is still quite a bit of model uncertainty. For the monthly S&P 500 data, the three-regime model has the highest posterior probability. This also matches the AIC-based results of Hardy (2001). It is very interesting that the most recent recession did not necessitate more regimes. The two-regime model's probability is near zero. This is a much more definitive statement than can be made by likelihood-based approaches.

Far more complicated models were selected for the daily data. For both the S&P 500 and the TSX data, the seven-regime model was chosen as optimal. The six-regime model also has a large posterior probability.

6.2. Parameters of highest posterior probability model

For all of the draws in the model with the highest posterior probability, we have organized the estimated parameter values. To prevent label switching, we have multiple options. We could sort each set of regimes by their posterior means. The regime with the smallest posterior mean is regime 1, the next smallest is regime 2, and so on. Figure 5 presents kernel density estimates of the posterior distributions of the regime parameters.

Figure 5: S&P 500 monthly parameter posterior distributions ordered by the regime means



(a) Mean

(b) Standard Deviation

Regimes 2 and 3 appear to have very similar posterior distributions. The distributions for the mean are highly peaked and very close to each other, while the distributions for the standard deviations are bimodal. The bimodality leads us to believe that it may be better to sort these regimes by their standard deviations. Figure 6 shows the kernel density estimates with the adjusted ordering.

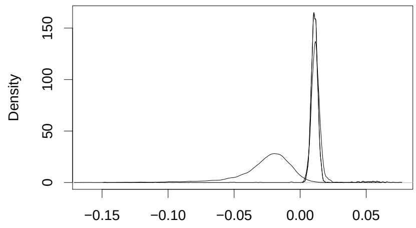
Now the mean distributions for regimes 2 and 3 overlap but they are unimodal and the distributions for the standard deviations are no longer bimodal. We think this more accurately portrays the results. This shows the dependence of summary results on seemingly small assumptions. Care should be taken when reporting or using the summaries. Table 4 presents the regime parameter estimates with the labels sorted by the standard deviations.

Table 4: Parameter Estimates for S&P 500 monthly data

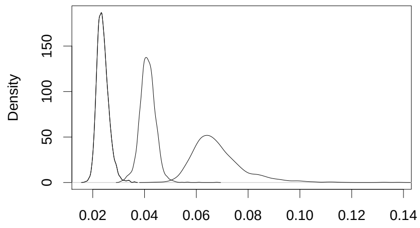
Regime	μ_i	σ_i	Transtion Matrix		
1	-0.0264	0.0683	0.876	0.112	0.013
2	0.0114	0.0412	0.020	0.963	0.017
3	0.0116	0.0239	0.002	0.032	0.957

All of the regimes show strong state persistence (witnessed by the diagonal

Figure 6: S&P 500 monthly parameter posterior distributions ordered by the regime standard deviations



(a) Mean



(b) Standard Deviation

elements of the transition matrix being close to one). There is one regime with a low expected value and high standard deviation and two regimes with positive expected values and lower standard deviations. If the series is currently in regime 1 and it moves to another regime, it will change to regime 2, 90% of the time. From regime 2, it is equally likely to move to either 1 or 3. From regime 3, it will move to regime 2, 94% of the time. That implies that there is some warning before a crash.

The parameter estimates for the monthly TSX data are presented in table 5.

Table 5: Parameter estimates for TSX monthly data

Regime	μ_i	σ_i	Transtion Matrix
1	-0.0181	0.0916	$\begin{pmatrix} 0.849 & 0.151 \\ 0.033 & 0.967 \end{pmatrix}$
2	0.0124	0.0346	

The parameter estimates for the TSX monthly data are very similar to the results in Hardy (2001) and Hardy (2002). There is one regime with a high mean and low standard deviation and another with a low mean and high standard deviation. Our estimated state persistence for the low mean regime is slightly higher than what is found in those papers (0.849 to 0.7899 or 0.7942). The parameter estimates for the daily datasets are presented in tables 6 and 7.

For both of the daily datasets, there is still strong state persistence. The S&P 500 estimates of all the self-transition probabilities are greater than 0.87. And removing the two regimes with the lowest expected value, all of the probabilities are above 0.96. The TSX data also have lower self-transition probabilities for the extreme states. It is also interesting to note that when the chain is in either

Table 6: Parameter estimates for S&P 500 daily data

Regime	μ_i	σ_i	Transtion Matrix						
1	-0.0134	0.0423	0.871	0.006	0.018	0.029	0.037	0.023	0.017
2	-0.0034	0.0196	0.001	0.872	0.007	0.026	0.053	0.025	0.016
3	-0.0008	0.0174	0.002	0.002	0.963	0.010	0.008	0.008	0.008
4	0.0001	0.0132	0.002	0.004	0.005	0.973	0.004	0.005	0.006
5	0.0006	0.0083	0.002	0.007	0.004	0.003	0.978	0.004	0.003
6	0.0008	0.0079	0.002	0.007	0.005	0.004	0.003	0.978	0.002
7	0.0014	0.0113	0.001	0.006	0.006	0.006	0.003	0.002	0.976

Table 7: Parameter estimates for TSX daily data

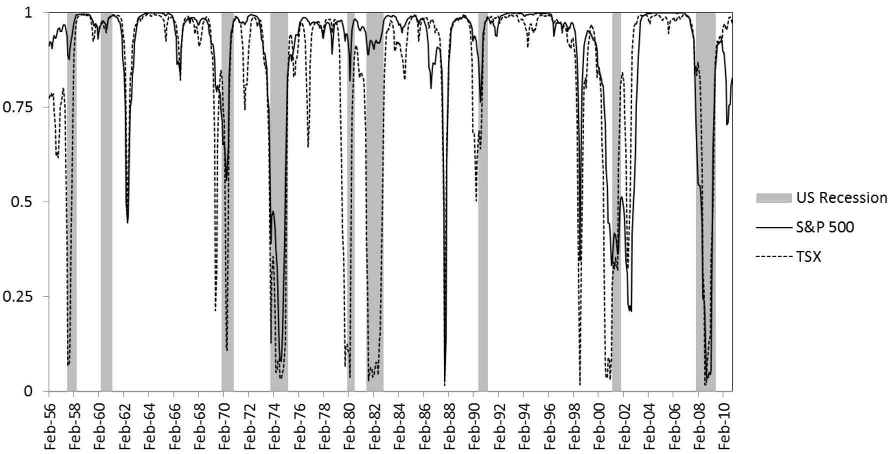
Regime	μ_i	σ_i	Transtion Matrix						
1	-0.0092	0.0299	0.671	0.009	0.018	0.055	0.047	0.031	0.169
2	-0.0054	0.0166	0.008	0.678	0.011	0.048	0.047	0.035	0.173
3	-0.0019	0.0149	0.010	0.013	0.875	0.009	0.013	0.038	0.042
4	-0.0003	0.0092	0.018	0.032	0.008	0.858	0.002	0.009	0.074
5	0.0003	0.0116	0.013	0.028	0.009	0.002	0.888	0.009	0.052
6	0.0016	0.0082	0.005	0.011	0.028	0.011	0.007	0.936	0.003
7	0.0052	0.0050	0.002	0.004	0.042	0.196	0.114	0.002	0.641

of the two lowest regimes and it switches to another regime, it is very likely that it will jump to the highest regime. From there it will likely move to one of the two middle regimes whose expected values are closest to zero.

6.3. Analysis of observations

In addition to increased insights into the parameter estimates, our methodology allows for individual observation analysis. For example, we can calculate the probability that the expected return at any given time period is positive (somewhat the probability of a bull market). That probability for the monthly data is plotted in figure 7. The solid line is the S&P data and the dashed line is the TSX data.

Figure 7: Probability the expected return is positive



By looking at the months when the probability of a positive mean is less than 0.5, we can find an alternate way to define recessions. Our results are similar to the generally accepted results (National Bureau of Economic Research, 2010). Table 8 outlines those periods for each dataset (Gable, 1959; Labonte, 2002; Hansen, 1960; Knoop, 2004). The month numbers which correspond to figure 7 are in parentheses.

There are three different periods where the TSX probabilities dropped below 0.5 and the S&P probabilities did not. The length of all the periods seem to be very similar between markets.

In addition to the probability that the expected value is positive, we can find estimates of both the expected value and the standard deviation for each observation (figure 8).

As expected, the expected value decreases and the standard deviation increases during the recessionary periods. The expected value of the TSX is generally higher (68% of the observations) especially during the really good times.

Figure 8: Parameter estimates by observation

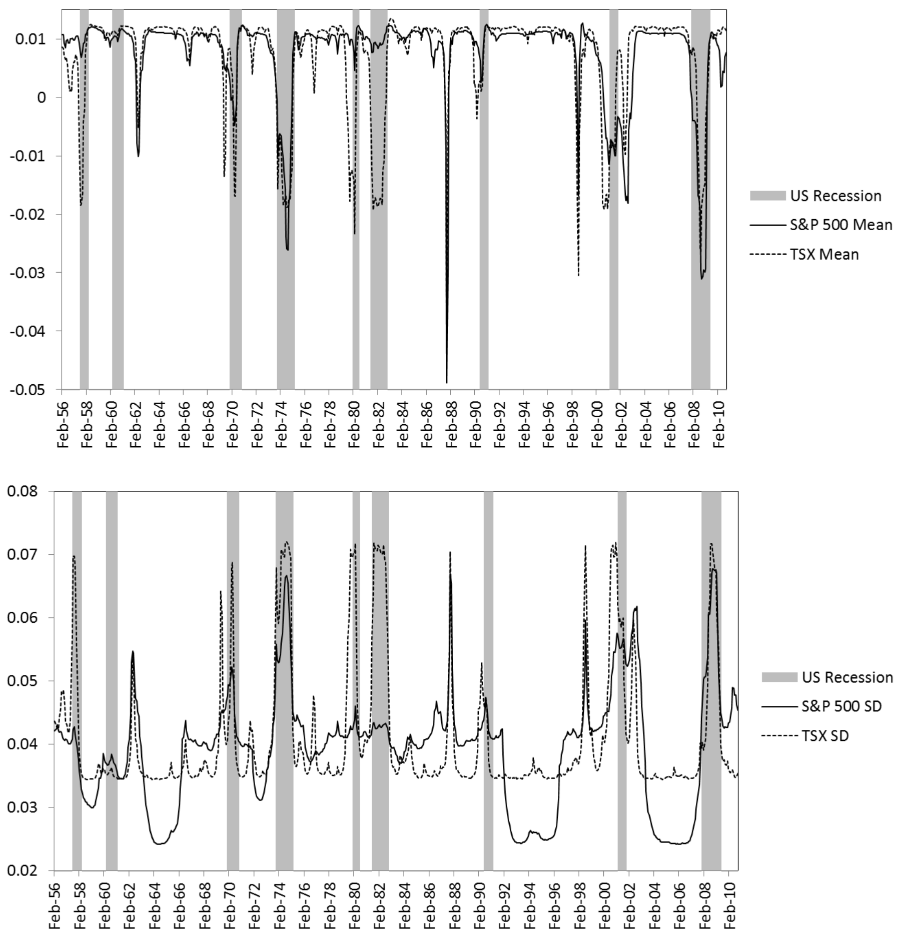


Table 8: Periods with high probabilities of negative expected values

TSX	S&P 500	Notes
07/57-11/57	-	
04/62	03/62-05/62	
06/69-07/69 04/70-05/70	-	Quebec Crisis
10/73-01/75	11/73-12/74	OPEC raised oil price
08/79-03/80 07/81-10/82	-	Iran Revolution
09/87-11/87	09/87-11/87	"Black Monday" crash
06/98-10/98	08/98	
07/00-08/01 05/02-08/02	11/00-10/01 02/02-10/02	Tech bubble burst, September 11th attacks
06/08-03/09	05/08-03/09	Subprime mortgage crisis

When the mean of the S&P is above 0.01, the TSX is higher 73% of the time. There are a few times when the mean of the TSX drops and the mean of the S&P remains high which correspond to the drops in the posterior probability presented in table 8.

7. Conclusion

When generating asset streams to use in pricing a complicated guarantee, it is essential that the variability in the model be properly accounted for. We discussed two methods in this paper. One accounts for the parameter uncertainty and the other accounts for both parameter and regime uncertainty. In our simulation study we showed that accounting for the uncertainty has a large effect on the price of simple put options, especially in the tails. The vast majority of the difference was attributed to the parameter uncertainty, but the regime uncertainty still had a significant effect.

We also confirmed some results in the current literature about the optimal number of regimes in monthly S&P and TSX data. The daily data required many more regimes. A seven-regime RSLN model is very difficult to fit using maximum likelihood. By using our algorithms, we are able to fit complicated RSLN models effectively.

For all of the real data, there is a large amount of model uncertainty. If that uncertainty is not accounted for in the model, prices on guarantees will be incorrect and the company will be exposed to more risk than assumed.

8. Acknowledgements

The authors would like to thank Mary Hardy and Emily Fox for their thoughtful guidance and advice. They would additionally like to thank the attendees at the 14th International Congress on Insurance: Mathematics and Economics, the 45th Actuarial Research Conference, and the University of Western Ontario Seminar Series for their comments and suggestions.

Akaike, H. (1974). A new look at the statistical identification model. *IEEE transactions on Automatic Control 19*(6), 716–723.

Aldous, D. (1985). Exchangeability and related topics. *École d'Été de Probabilités de Saint-Flour XIII1983*, 1–198.

Bakshi, G., C. Cao, and Z. Chen (1997). Empirical performance of alternative option pricing models. *Journal of Finance 52*(5), 2003–2049.

Beal, M., Z. Ghahramani, and C. Rasmussen (2002). The infinite hidden Markov model. *Advances in Neural Information Processing Systems 1*, 577–584.

Blackwell, D. and J. MacQueen (1973). Ferguson distributions via Pólya urn schemes. *The annals of statistics 1*(2), 353–355.

Datastream International (2010, December). Datastream.

Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *The Annals of Statistics 1*(2), 209–230.

Fox, E. B., E. B. Sudderth, M. I. Jordan, and A. S. Willsky (2007, November). The sticky hdp-hmm: Bayesian nonparametric hidden markov models with persistent states. Technical Report 2777, MIT Laboratory for Information and Decision Systems.

Gable, R. (1959). The Politics and Economics of the 1957-1958 Recession. *The Western Political Quarterly 12*(2), 557–559.

Green, P. J. (1995, December). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika 82*(4), 711–732.

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

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

Hansen, A. H. (1960). *Economic Issues of the 1960s*. New York: McGraw-Hill.

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

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

Hardy, M. (2003). *Investment Guarantees: Modeling and Risk Management for Equity Linked Life Insurance*. John Wiley and Sons.

Jasra, A., C. Holmes, and D. Stephens (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science 20*(1), 50–67.

Knoop, T. A. (2004). *Recessions and Derpressions: Understanding Business Cycles*. Praeger Publishers.

Labonte, M. (2002). The current economic recession. *Congressional Research Service*.

National Bureau of Economic Research (2010, December). Us business cycle expansions and contractions.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE 77*(2), 257–286.

Robert, C., T. Ryden, and D. Titterington (2000). Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62*(1), 57–75.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics 6*(2), 461–464.

Shiller, R. (2005). *Irrational Exuberance*, Volume 2. Broadway Books. http://www.econ.yale.edu/ shiller/data.htm.

Standard and Poor's (2010). Monthly s&p500 returns. `http://www2.standardandpoors.com/spf/xls/index/MONTHLY.xls`.

Teh, Y., M. Jordan, M. Beal, and D. Blei (2006). Hierarchical dirichlet processes. *Journal of the American Statistical Association 101*(476), 1566–1581.

Yahoo! Inc. (2010, December). Yahoo! finance.