Accounting for Regime and Parameter Uncertainty in Regime-Switching Models

Brian M. Hartman, PhD, ASA
University of Waterloo
Acknowledgments

- Joint work with Matthew J. Heaton, Duke University
- Special thanks to Mary Hardy and Emily Fox
Agenda

Introduction

Fixed Number of Regimes

Dynamic Number of Regimes

Simulation Study

Real Data Analysis

Conclusion
Regime-switching models are prominently used in actuarial science and risk management (e.g. Boyle and Draviam, 2007; Chen, 2008; Desmedt et al., 2004; Freeland et al., 2009; Hardy et al., 2006; Siu, 2008).
Regime-switching Models

In a regime-switching model, the current state, $x_t$, only depends upon the previous state, $x_{t-1}$. Additionally, the observation, $y_t$, depends only upon the current state, $x_t$. 

\[ X_{t-1} \rightarrow X_t \rightarrow X_{t+1} \]
\[ Y_{t-1} \rightarrow Y_t \rightarrow Y_{t+1} \]
RSLN Likelihood

- Given the state vector, the observations are simply draws from a lognormal distribution.
- For ease of computation, I exponentiate the values so they are draws from a normal distribution with the following likelihood

\[
p(y_i|\mu, \sigma^2, x_i = r) = \frac{1}{\sqrt{2\pi\sigma_r^2}} \exp \left\{ -\frac{(y_i - \mu_r)^2}{2\sigma_r^2} \right\}.
\]
Problem Description

With increasingly complicated products, it is common to simply simulate the future asset prices from an RSLN model using the following steps:

1. Fit RSLN models of various dimensions using maximum likelihood
2. Perform a model selection technique to find the “best” model
3. Obtain many simulated asset streams using the optimized number of regimes and parameter values
Problem Description

Because the asset streams are generated from the optimized number of regimes and parameter values, there are two implicit assumptions:

▶ The number of regimes is known
▶ The parameter values are known

Additionally, estimation using maximum likelihood can be very unstable in this situation (large number of parameters with identifiability concerns).
Why Bayesian?

- By using a Bayesian methodology, I am able to account for the uncertainty in the parameters and the number of regimes improving my pricing ability.
- Additionally, data augmentation makes the parameter estimates much more dependable.
Priors on $(\mu, \sigma^2)$

There are two options for the prior on $(\mu, \sigma^2)$:

\[
p(\mu, \sigma^2) = N(\mu|\mu_0, n_0\sigma^2)IG(\sigma^2|\alpha_0, \beta_0)
\]

\[
p(\mu, \sigma^2) \propto N(\mu|\mu_0, n_0\sigma^2)IG(\sigma^2|\alpha_0, \beta_0)1_{\mu_1<\mu_2<\ldots<\mu_R}
\]
Priors on \((\pi, x_i)\)

Each row of the transition matrix is given a Dirichlet prior.

\[
p(\pi_r) = \text{Dir}(1, 1, \ldots, 1) \quad \forall \quad r \in \{1, 2, \ldots, R\}
\]

All of the states are given a uniform prior.

\[
p(x_i = r) \propto 1 \quad \forall \quad r \in \{1, 2, \ldots, R\}, i \in \{1, 2, \ldots, n\}
\]
Posterior Distributions of \((\mu, \sigma^2, \pi)\)

Given the state vector, the full conditional posterior distributions for \(\mu\), \(\sigma^2\), and \(\pi\) are tractable.

\[
p(\mu_r|\cdot) = \text{St} \left( \mu_{\text{post}}, (n_r + n_0) \left( \alpha_0 + \frac{n_r}{2} \right) \beta_{\text{post}}^{-1}, 2\alpha_0 + n_r \right)
\]

\[
p(\sigma_r^2|\cdot) = \text{IG} \left( \alpha_0 + n_r/2, \beta_{\text{post}} \right)
\]

\[
p(\pi_r|\cdot) = \text{Dir} \left( 1 + n_{r1}, 1 + n_{r2}, \ldots, 1 + n_{rR} \right)
\]

\[
\mu_{\text{post}} = \frac{(n_0 + n_r)^{-1}(n_0\mu_0 + \sum (y_i \mathbf{1}_{x_i=r}))}{(n_0 + n_r)}
\]

\[
\beta_{\text{post}} = \beta_0 + \sum_{i:x_i=r} \frac{(y_i - \bar{y}_r)^2}{2} + \frac{n_0 n_r (\mu_0 - \bar{y}_r)^2}{2(n_0 + n_r)}
\]
Posterior Distribution of $x_i$

Given all the other parameters, the individual elements of the state vector have two options for the posterior distribution

$$f(x_i|\cdot) \propto N(y_i|x_i = r, \mu_r, \sigma_r^2)p(x_i) \quad \text{(Standard method)}$$

$$f(x_i|\cdot) \propto \int_{\mu, \sigma^2} N(y_i|y_{-i}, x, \mu, \sigma^2)p(x_i)d\mu d\sigma^2 \quad \text{(To improve mixing)}$$

$$\propto St(x_i|\mu_n, (n + n_0)(n + n_0 + 1)^{-1}(\alpha_0 + \frac{1}{2}n)\beta_n^{-1}, 2\alpha_0 + n)$$

$$\mu_n = (n_0 + n)^{-1}(n_0\mu_0 + n\bar{x})$$

$$\beta_n = \beta_0 + \frac{1}{2}ns^2 + \frac{1}{2}(n_o + n)^{-1}n_0n(\mu_0 - \bar{x})^2$$
Single-site Updating

With a given number of states, the most direct approach is to estimate each state \((x_i)\) individually given the other states and the model parameters.

\[
p(x_i|\theta, y_{1:n}, x_{1:i-1}, x_{i+1:n})
\]

reduces to

\[
p(x_i|\theta, y_{1:n}, x_{i-1}, x_{i+1}) \propto p(x_i)q(x_{i-1}, x_i)q(x_i, x_{i+1})p(x_i|y_{1:n})
\]

- When \(i = 1\), \(q(x_0, x_1)\) is replaced by \(\nu(x_1)\).
- When \(i = n\), \(q(x_n, x_{n+1})\) is replaced by 1.
Global Updating

Alternatively, the states could all be updated simultaneously through a forward filtering, backward sampling algorithm (Rabiner 1989).
Step 1: Forward Filtering

\[ \phi_{i+1|0}(j) = \nu(j) \quad \forall \quad j \in \{1, \ldots, R\} \]

For \( i \in \{1, \ldots, n\} \)

\[ c_i = \sum_{r=1}^{R} \phi_{i|i-1}(r)p(x_i = r|y_i) \]

\[ \phi_{i}(j) = \frac{\phi_{i|i-1}(j)p(x_i = j|y_i)}{c_i} \quad \forall \quad j \in \{1, \ldots, R\} \]

\[ \phi_{i+1|i}(j) = \sum_{r=1}^{R} \phi_{i}(r)q_{rj} \quad \forall \quad j \in \{1, \ldots, R\} \]
Step 2: Backward Sampling

After the $\phi_i$ values are generated in the previous slide, the states are drawn with the following probabilities:

$$p(X_n = x) = \phi_n(x)$$

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

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

- Robert et al. (2000) and Cappé et al. (2003) both looked at implementing reversible jump MCMC (Green, 1995) in RSLN models, but they needed to strongly restrict the parameter space.

- In my experience, it is very difficult to properly define the proposal distributions to obtain acceptable mixing.
Dirichlet Processes

Beal et al. (2002) was the first paper to use the Dirichlet process to create a regime-switching model with a dynamic number of states.
What is a Dirichlet Process?

It can be compared to a Chinese restaurant with an infinite number of tables and the capacity to make an infinite number of unique meals.
Each patron enters one at a time.
\begin{align*}
P(\text{person i sits at an empty table}) & \propto \alpha_0 \\
P(\text{person i sits at table j}) & \propto n_j
\end{align*}
\[ P(\text{person } i \text{ sits at an empty table}) \propto \alpha_0 \]
\[ P(\text{person } i \text{ sits at table } j) \propto n_j \]
\[ P(\text{person } i \text{ sits at an empty table}) \propto \alpha_0 \]
\[ P(\text{person } i \text{ sits at table } j) \propto n_j \]
\[ P(\text{person } i \text{ sits at an empty table}) \propto \alpha_0 \]
\[ P(\text{person } i \text{ sits at table } j) \propto n_j \]
\[ P(\text{person } i \text{ sits at an empty table}) \propto \alpha_0 \]
\[ P(\text{person } i \text{ sits at table } j) \propto n_j \]
\( P(\text{person } i \text{ sits at an empty table}) \propto \alpha_0 \)
\( P(\text{person } i \text{ sits at table } j) \propto n_j \)
DP Shortcoming

- The state of the current observation does not depend upon the state of the previous observation.
- Beal et al. (2002) overcame that by adjusting the assignment probabilities.
Assignment Probabilities

\[
P(\text{Self Transition}) = \frac{n_{ii} + \alpha}{\sum_i n_{ij} + \beta + \alpha}
\]

\[
P(\text{Existing Transition } j \neq i) = \frac{n_{ij}}{\sum_i n_{ij} + \beta + \alpha}
\]

\[
P(\text{Oracle}) = \frac{\beta}{\sum_i n_{ij} + \beta + \alpha}
\]
Oracle Probabilities

If the oracle is chosen, the state is chosen from the following probabilities:

\[ P(\text{Existing State } j) = \frac{n_j^o}{\sum_j n_j^o + \gamma} \]
\[ P(\text{New State}) = \frac{\gamma}{\sum_j n_j^o + \gamma} \]
iHMM Overview

Advantages

▶ Easy to explain and understand
▶ Highly customizable
▶ Hyperparameters have practical explanations

Disadvantages

▶ Unable to be efficiently estimated
Hierarchical Dirichlet Process

Teh et al. (2006) introduced the concept of a hierarchical Dirichlet Process. For illustration, it can be compared to a Chinese restaurant franchise (CRF) with:

- an infinite number of restaurants
- each with an infinite number of tables
- a global menu with an infinite number of meals
Hierarchical Dirichlet Process
Hierarchical Dirichlet Process
Hierarchical Dirichlet Process
Hierarchical Dirichlet Process
Hierarchical Dirichlet Process
HDP-HMM

In most regime-switching models, state persistence is expected. The HDP-HMM allows for state persistence, but does not expressly prefer it.
Fox et al. (2009) introduced the sticky HDP-HMM which can be thought of as a CRF with loyal customers.

- Each restaurant has a specialty dish with the same index as the restaurant.

- While the dish is made elsewhere, it is preferred in the namesake restaurant.
Sticky HDP-HMM

- The parent will choose a dish.
- The child will then enter the namesake restaurant for that dish.
- That will increase the probability that the child will have the same dish.
- This develops family loyalty to a given restaurant in the franchise.
Simulation Setup

To examine the effect of accounting for the regime and parameter uncertainty I performed the following simulation study:

▶ Simulate 100 datasets of 100 observations in each of 6 cases
▶ For each dataset, fit a RSLN model using ML, SSU, and Sticky HDP-HMM
▶ For each fit, generate asset streams
▶ Compare the estimated (mean) price, and 0.025 and 0.975 quantiles for a simple European put option with a strike prices of $S_0$ and $10S_0$ over various term lengths.
### Case Description

<table>
<thead>
<tr>
<th>Case</th>
<th>R1: $\mu, \sigma$</th>
<th>R2: $\mu, \sigma$</th>
<th>R3: $\mu, \sigma$</th>
<th>Transition Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.012 0.035</td>
<td>-0.016 0.078</td>
<td>-</td>
<td>$(0.96 \ 0.04)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$(0.21 \ 0.79)$</td>
</tr>
<tr>
<td>2</td>
<td>0.012 0.035</td>
<td>-0.016 0.078</td>
<td>-</td>
<td>$(0.50 \ 0.50)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$(0.50 \ 0.50)$</td>
</tr>
<tr>
<td>3</td>
<td>0.000 0.035</td>
<td>0.000 0.078</td>
<td>-</td>
<td>$(0.96 \ 0.04)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$(0.21 \ 0.79)$</td>
</tr>
<tr>
<td>4</td>
<td>0.025 0.035</td>
<td>-0.031 0.078</td>
<td>-</td>
<td>$(0.96 \ 0.04)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$(0.21 \ 0.79)$</td>
</tr>
<tr>
<td>5</td>
<td>0.014 0.050</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>0.012 0.035</td>
<td>-0.016 0.078</td>
<td>0.040 0.010</td>
<td>$(0.95 \ 0.04 \ 0.01)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$(0.21 \ 0.78 \ 0.01)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$(0.80 \ 0.19 \ 0.01)$</td>
</tr>
</tbody>
</table>
At-the-money ($S_0 = K = 1$) Put Option Prices

(a) Case 1

(b) Case 2

(c) Case 3

(d) Case 4

(e) Case 5

(f) Case 6
Deep-in-the-money $(10S_0 = K = 10)$ Put Option Prices

(a) Case 1 
(b) Case 2 
(c) Case 3 
(d) Case 4 
(e) Case 5 
(f) Case 6
Data Description

Date Range

- Daily Data: 02 Jan. 1981 - 12 Nov. 2010 (7,526 obs.)

Indices

- S&P 500
- TSX

Sources

- Standard and Poors (2010)
- Yahoo! Inc. (2010)
- Datastream International (2010)
- Shiller (2005)
## Posterior Probabilities for the Number of Regimes

### Monthly Data

<table>
<thead>
<tr>
<th>Index</th>
<th>Number of Regimes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>-</td>
</tr>
<tr>
<td>TSX</td>
<td>-</td>
</tr>
</tbody>
</table>

### Daily Data

<table>
<thead>
<tr>
<th>Index</th>
<th>Number of Regimes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>-</td>
</tr>
<tr>
<td>TSX</td>
<td>-</td>
</tr>
</tbody>
</table>
Parameter Estimates for S&P 500 monthly data

<table>
<thead>
<tr>
<th>Regime</th>
<th>$\mu_r$</th>
<th>$\sigma_r$</th>
<th>Transition Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0264</td>
<td>0.0683</td>
<td>$\begin{pmatrix} 0.876 &amp; 0.112 &amp; 0.013 \ 0.020 &amp; 0.963 &amp; 0.017 \ 0.002 &amp; 0.032 &amp; 0.957 \end{pmatrix}$</td>
</tr>
<tr>
<td>2</td>
<td>0.0114</td>
<td>0.0412</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0116</td>
<td>0.0239</td>
<td></td>
</tr>
</tbody>
</table>

Parameter estimates for TSX monthly data

<table>
<thead>
<tr>
<th>Regime</th>
<th>$\mu_r$</th>
<th>$\sigma_r$</th>
<th>Transition Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0181</td>
<td>0.0916</td>
<td>$\begin{pmatrix} 0.849 &amp; 0.151 \ 0.033 &amp; 0.967 \end{pmatrix}$</td>
</tr>
<tr>
<td>2</td>
<td>0.0124</td>
<td>0.0346</td>
<td></td>
</tr>
</tbody>
</table>
Probability of Positive Expected Value

![Graph showing probability of positive expected value over time with distinct peaks and valleys, labeled for US Recession, S&P 500, and TSX.]
Observation-wise Mean Estimates

Graph showing time series data with shaded areas indicating US recessions, and lines representing S&P 500 and TSX mean values.
Observation-wise Standard Deviation Estimates

- S&P 500 SD
- TSX SD

Description
Results
Conclusion

- Failing to account for regime or parameter uncertainty can drastically alter your predictions
- By using Bayesian methods, I am able to properly account for that uncertainty
- Contributions from the machine learning literature make the estimation of the number of regimes dependable and efficient
Accounting for Regime and Parameter Uncertainty in Regime-Switching Models

Brian M. Hartman, PhD, ASA
University of Waterloo