Asset Returns from a Distance-Based Random Partition Model

Richard D. Payne   Brian M. Hartman   David B. Dahl

April 12, 2013

1 Introduction

Regime-switching models are important in finance and actuarial science because they are often successful in simulating asset returns. These models fit the complicated nature of market risk exposure more effectively and naturally than other models. Many factors affect market risk exposure in addition to total returns. Policyholder behavior (lapses, mortality, morbidity, etc.), policy riders (guarantees, look-backs, ratchets, etc.), industry and government forces (competition and regulation), among others can all depend on the asset returns and affect the total risk exposure from the market (Hardy, 2003, ch. 1; Booth et al., 2004, ch. 4). In these complicated settings, an analytical solution is often unavailable and the asset returns will need to be simulated.

Regime-switching models have been used for these simulations for a variety of reasons. Regime-switching models are very intuitive and, more importantly seem to fit the return data well. Many different papers in the finance and economics literature (e.g. Cont, 2001; Ding and Granger, 1996; Hommes, 2002; Lux, 2009; Rydn et al., 1998) discuss stylized facts apparent in many financial time series (differing in both market and instrument). Regime-switching models are flexible enough to incorporate these stylized facts in a very natural way.

Although regime-switching models have been successful, we propose a more flexible model that clusters data based on both the value of the data as well as its proximity to adjacent observations. Hereafter we discuss stylized facts of financial data, present our model, demonstrate its clustering abilities, forecast data from the S&P 500, discuss results, and state conclusions.

Stylized Facts of Financial Data

Regime-switching models and asset returns rarely have significant autocorrelations. Given the regime vector, observations from these models are independent, making it easy to remove autocorrelation in the return series. Currently, the regimes are defined by a Markov-switching process implying that the autocorrelation will decay geometrically.

The heavy tails inherent in asset returns are modeled well by regime-switching models. Even with lognormal component distributions, adding regime-switching makes the tails thicker (Hardy, 2001). Conditionally, even after removing the volatility clustering, the returns continue to show heavy tails. The tail thickness in both cases can be modified by adjusting the component distributions. The flexibility in tail thickness also lends itself well to the imbalance of returns – large negative returns are more common than large positive returns. Using a regime-switching model, heavier tails can be specified for poor returns and lighter tails for strong returns.

These models also describe volatility clustering and intermittency well. Periods of high variability and jumps are easily modeled as rare, high volatility regimes with extreme means within a regime-switching model. When volatility persists, volatility clusters are grouped as observations from the same regime.

In Hartman and Heaton (2011), we show that as the time scale becomes shorter, the number of regimes in a regime-switching lognormal distribution grows, making the distribution less Gaussian. The leverage effect is naturally handled in a two regime model where the first regime has a high mean and low variance and the second regime has a low mean and high variance (Hardy 2001). Finally, we notice that course-grained measures of volatility predict fine-scale volatility better than the other way around. When estimating the regime-switching models in a Bayesian paradigm, volatility estimates can be used from coarser data to inform the prior for models of fine-scale data.

These stylized facts which are handled well by regime switching models are also handled well by our model. In the ensuing document, we will compare our model with regime-switching models in their ability to cluster and forecast data.

2 Model

2.1 Sampling Model

Let S_t be the price of an asset (e.g., a stock index) at time t (for $t = 1, \ldots, n$) and y_t is the return on the asset, defined as:

$$y_t = \frac{S_t}{S_{t-1}}.$$

Under geometric Brownian motion (and, by extension, the Black-Scholes-Merton option pricing model), y_t is assumed to follow a lognormal distribution as follows:

$$\log(y_t) \sim \mathrm{N}(\mu, \lambda)$$

and $y_1, \ldots, y_n$ are independent given μ and precision λ. This independent lognormal model does not permit for variance persistence or jumps, both of which are visible in stock market data and generally accepted in the finance literature . Both of those features can be intuitively incorporated by allowing the μ and λ to vary between the observations:

$$\log(y_t) \sim N(\mu_t, \lambda_t)$$

As before, we assume independence among $y_1, \ldots, y_n$ given the parameters. For notational convenience, let $\theta_t = (\mu_t, \lambda_t)$.

We model the distribution from which the y_t are drawn as a mixture of distributions of the form $F(\theta)$, with the mixing distribution over θ being G. The prior for this mixing distribution will be a modified Dirichlet process with mass parameter α and base distribution G_0. This gives the following model:

$$\begin{aligned} y_t \mid \theta_t &\sim F(\theta_t) \\ \theta_t \mid G &\sim G \\ G &\sim DP(G_0, \alpha) \end{aligned}$$

2.2 Prior Distribution

The task is to estimate n couplets: $\theta_1, \ldots, \theta_n$. Given the limited data, efficiency can be obtained by assuming there is a latent clustering yielding ties among the n vectors. The clustering is encoded in a partition $\pi = \{S_1, \ldots, S_q\}$ of the set $S_0 = \{1, \ldots, n\}$ having the following properties: (i) $S_i \neq \emptyset$ for $i = 1, \ldots, q$ (non-empty subsets), (ii) $S_i \cap S_j = \emptyset$ for $i \neq j$ (mutually exclusive subsets), and (iii) $\cup_{j=1}^{q} S_j = S_0$ (exhaustive subsets). The number of subsets q for a partition π can range from 1 (i.e., all vectors belong to the same subset) to n (i.e., each vector is in a singleton subset). It is sometimes convenient to represent a partition π as a vector $\boldsymbol{c}$ of cluster labels, where $c_i = j$ if and only if $i \in S_j$, for $i = 1, \ldots, n$ and $j = 1, \ldots, q$.

It is convenient to reparameterize as follows: $\theta_i = \sum_{j=1}^{q} \phi_j \mathrm{I}\{i \in S_j\}$, where $\phi_j \sim G_0$ and $\pi \sim p(\pi)$, where G_0 is a prior distribution from which $\phi_1, \ldots, \phi_q$ are independently drawn and $p(\pi)$ is a prior distribution over partitions of the form $\pi = \{S_1, \ldots, S_q\}$.

The centering distribution for the Dirichlet process will be assumed to follow a normal-gamma distribution. The precision λ will be assumed to follow a gamma distribution, and mean μ a conditional normal distribution. Thus G_0 will be distributed as follows:

$$G_0 \sim N(\mu_0, c_0\lambda \mid \lambda) * Gam(a, b)$$

where μ_0, c_0, a (shape), and b (rate) are hyperparameters. The conjugacy of this model will allow us to sample more efficiently when forecasting.

When $p(\pi) \propto \prod_{S \in \pi} \alpha\Gamma(|S|)$, G_0 is known as the "centering" distribution of the Dirichlet process and the model is the popular Dirichlet process mixture model. Instead, assume:

$$p(\pi) \propto \prod_{S \in \pi} \alpha\Gamma\left(\frac{1}{|S|} \sum_{i \in S} h_i(S)\right), \tag{1}$$

where h_{ik} it the attraction of item i to item k where $h_{ii} = 1$ and $\sum_{k \neq i} h_{ik} = n - 1$. The scaling of h_{ik} is a key feature — it effectively leaves the total mass of the $n - 1$ items unchanged while permitting the pairwise information to shift mass away from distant items and towards proximate items.

Suppose, for example, that the information regarding clustering of n items can be expressed in terms of an $n \times n$ distance matrix of elements d_{ik} giving the distance between items i and k. Our framework can use any distance metric to define the distance matrix. Items i and k with a large distance d_{ik} should have a small attraction h_{ik} to induce a low clustering probability. Raising the reciprocal of distance to a power provides one potential transformation, i.e., for $i \neq k$, $h_{ik} \propto (d_{ik} + \epsilon)^{-t}$, where the temperature t is non-negative and ϵ is a small positive value to ensure that all attractions are strictly positive. When examining stock returns over time, a natural distance measure would be the temporal distance between y_i and y_k, or $d_{ik} = |i - k|$. Alternatively, the clustering may only use information from the adjacent observations:

$$d_{ik} = \begin{cases} |i - k| & \text{if } |i - k| \leq 1 \\ \infty & \text{if } |i - k| > 1 \end{cases}$$

This is similar to the regime-switching or hidden Markov models common in finance. We call t the temperature as it has the effect of dampening or accentuating the proximity measurements. If the pairwise information comes in the form of pairwise proximity p_{ik}, where large values are meant to increase the likelihood of clustering, a simple definition of the attractions may be $h_{ik} \propto (p_{ik})^t$.

Note that if there pairwise information is noncommittal, meaning the all distances or proximities are equal, then $h_{ik} = 1$ for all i, k. Also, the pairwise information can be extinguished by setting $t = 0$, leading to $h_{ik} = 1$ for all i, k. In this case, the partition distribution is the same as that from the Dirichlet process.

3 Clustering

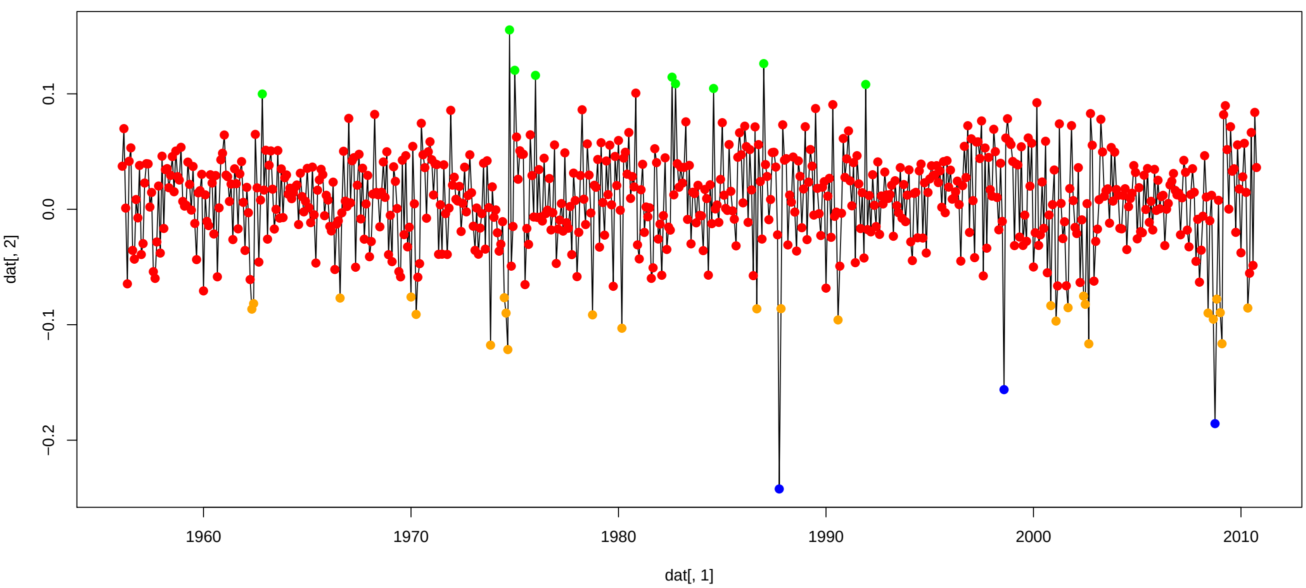
For the simulations and forecasts presented hereafter, for $i \neq k$ we use the following distance metric:

$$d_{ij} = \begin{cases} |i-j| & \text{if } |i-j| \leq 1 \\ 1000 & \text{if } |i-j| > 1 \end{cases}$$

This distance metric will place a much higher probability of an observation clustering with adjacent observations. The use of other distance matrices is discussed in the discussion and limitations section.

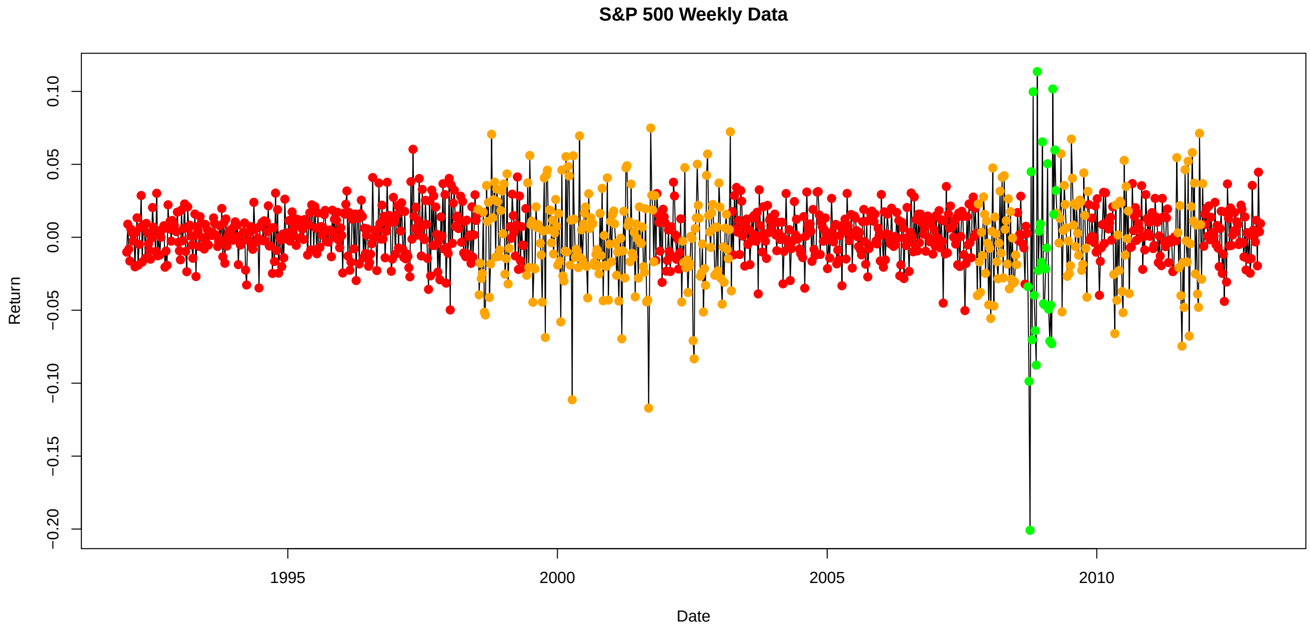
To demonstrate this model's ability to cluster data based on both the value of the asset return as well as its proximity to other returns, we use data from the S&P 500 Index. If the attraction matrix introduced in our model is not used, the clustering proceeds as the classic Dirichlet clustering model (Figure 1). If the attraction matrix from our model is used, we get the results displayed in Figure 2 (page 5).

Figure 1: Dirichlet clustering. Each color represents one cluster.



It is clear from Figures 1 & 2 that the attraction matrix incorporates the natural feel of clusters more intuitively. This is a useful feature in our model. A key difference between the classic Dirichlet process of clustering and our model are the means of the clusters. The Dirichlet process has clusters whose means differ significantly, while our cluster means are closely centered

Figure 2: Clustering with our model. Each color represents one cluster.



around 0. Both models incorporate differing spreads. Although our model accounts for the varying of both means and precisions, only the precisions seem to vary significantly when incorporating the attraction matrix.

4 Simulation Study

One practical application of this model is its ability to simulate asset returns. This simulation is based on algorithm 3 presented in Neal's (2000) paper. Using n data points $(y_1, \ldots, y_n)$, p data points $(y_{n+1}, \ldots, y_{n+p})$ will be forecast on each iteration. Forecasting proceeds as follows:

1. Set initial values for $y_{n+1}, \ldots, y_{n+p}$
2. Cluster $y_1, \ldots, y_{n+p}$
3. Draw new values for $y_{n+1}, \ldots, y_{n+p}$ based on the data in the cluster
4. Repeat steps 2 & 3

At any given iteration, each y_t will be associated with a cluster label c_t. During the clustering algorithm, the probability that y_t has a cluster label c is proportional to a t-distribution with degrees of freedom, location, and scale parameters ν, ω, and η respectively:

$$P(c_t = c \mid y_t, c_{-t}) \propto St(\nu, \omega, \eta)$$

where

$$\nu = n + 2a$$

$$\omega = \frac{\sum y_t + c_0\mu_0}{n + c_0}$$

$$\eta = \frac{(n + c_0)(n + 2a)}{2(n + c_0 + 1)\left(b + \frac{1}{2}\left(\sum y_t^2 + c_0\mu_0^2 - \frac{(\sum y_t + c_0\mu_0)^2}{n + c_0}\right)\right)}$$

and a, b, μ_0, c_0 are the prior hyperparameters.

This simulation provides posterior predictive distributions for $y_{n+1}, \ldots, y_{n+p}$. Using these simulations, this model can predict the probability that an asset return will have a given event (e.g. the probability that a return loses 6% in a 10 week period). Once these probabilities are obtained, logistic regression can be used to determine how successful the model was in prediction. The dependent variable is whether or not the event occurred, and the independent variable is the predicted probability of the event.

4.1 Analysis of S&P 500 Index

Weekly asset returns from the S&P 500 were obtained for the years 1992-2013. Hyperparameters were selected to be $a = 2,\ b = .01,\ \mu_0 = 0$ and $c_0 = 50$. The simulation was iterated 100,000 cycles and 50 observations were forecast. This simulation was done several times with mass parameters ranging between .5 and 50. The autocorrelation of the simulation was calculated by assessing the correlation between the y_{n+1} value and each consecutive forecasted value. As seen in Figure 3 (page 7), as the mass increased (and subsequently the number of clusters), short-term autocorrelation increased and long-term autocorrelation decreased.

This flexibility in autocorrelation simulation is useful particularly if one is performing short-term forecasting. Unfortunately, this model does not appear to simultaneously produce high short-term autocorrelation and a persistent long-term autocorrelation. Regime switching models still maintain a slight edge in this regard.

4.2 Analysis of S&P 500 Stocks

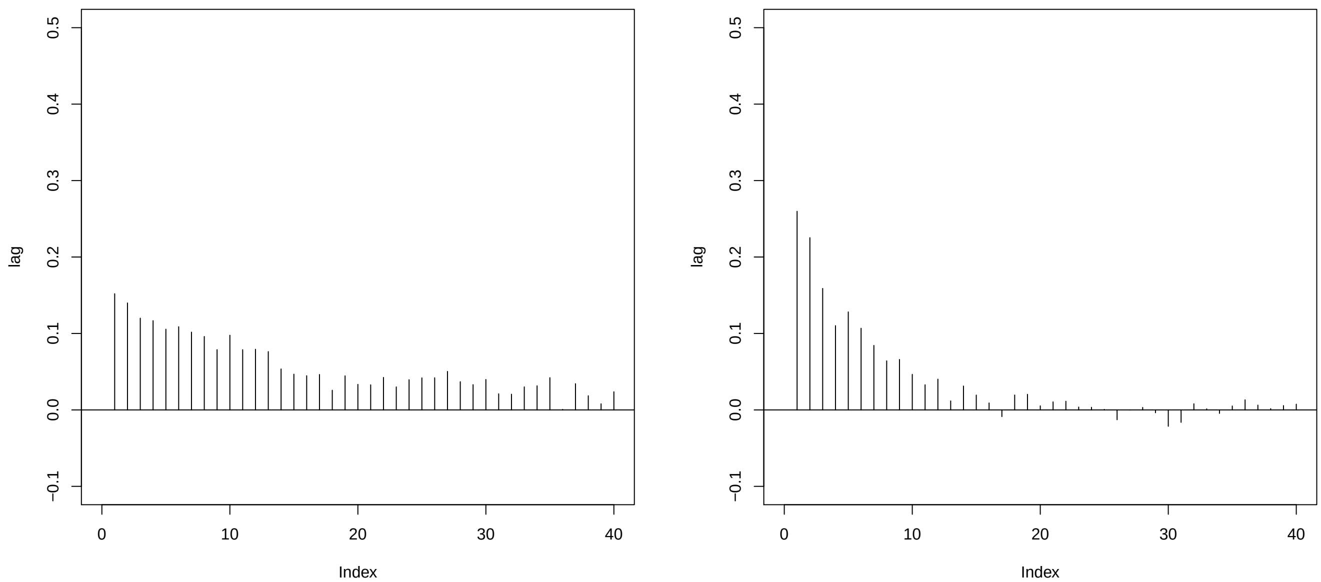
In order to evaluate the effectiveness of the predictions of the model, 100 stocks were randomly selected from the S&P 500 in March 2013. Weekly returns through March 4, 2013 were gathered for each stock. Stocks with fewer than 500 weeks of data were removed from the study (9 in total). The data was then adjusted for the S&P market as follows:

$$\log\left(\frac{y_t}{y_{t-1}}\right) - \log\left(\frac{S_t}{S_{t-1}}\right)$$

where y_t is the value of an individual stock and S_t is the value of the S&P 500 at time t. For each stock, a time interval of 100 weeks was randomly chosen and used to forecast 10 weeks of stock returns using our model. Hyperparameters were set at $a = 2,\ b = .01,\ \mu_0 = 0,\ c_0 = 50$, and a mass parameter $\alpha = 10$. Each stock iterated 10,000 times.

Using the simulation results, the probability that a stock would have a return worse than 6% in 10 weeks was calculated using the posterior probability distributions. Whether or not each stock

Figure 3: Average ACF of forecasted data with mass parameter of .5 (left) and 50 (right)



realized a return worse than 6% in 10 weeks was recorded and logistic regression was performed to determine if there was a relationship. The stock data was also run through a regime switching model with the same hyperparameters for a detailed comparison. The results are summarized in Table 1.

Table 1: Logistic regression and prediction of a 6% loss in 10 weeks

Our Model

	Estimate	Std. Error	z value	p value
Intercept	-1.7468	0.8903	-1.962	0.0498
Posterior Probabilities	2.8871	2.7749	1.040	0.2981

Regime Switching

	Estimate	Std. Error	z value	p value
Intercept	-2.060	1.192	-1.729	0.0839
Posterior Probabilities	3.762	3.641	1.033	0.3015

Predictions		**Actual**
Our Model	0.3022	0.2967
Regime Switching	0.3150	

Additionally, the probability that a stock would have two or more 3% losses in 10 weeks was also calculated, whether or not this occurred was recorded, and logistic regression was performed (Table 2).

Table 2: Logistic regression and prediction of two or more 3% losses in 10 weeks

Our Model

	Estimate	Std. Error	z value	p value
Intercept	-8.155	2.407	-3.387	0.000705
Posterior Probabilities	9.356	2.701	3.464	0.000532

Regime Switching

	Estimate	Std. Error	z value	p value
Intercept	-8.928	2.817	-3.169	0.00153
Posterior Probabilities	10.084	3.123	3.228	0.00124

Predictions		**Actual**
Our Model	.8732	0.5165
Regime Switching	.8874	

The results of this analysis indicate that our model produces equivalent results to the regime-switching models. However, these models are far from perfect. For example, when predicting a 6% loss in 10 weeks, both models do not have significant p-values, but their overall estimates are very precise. When predicting two 3% losses in 10 weeks, the opposite occurred; both models had high prediction, but inaccurate estimates. Despite these limitations in prediction, the models produce equivalent predictions.

5 Discussion & Limitations

When comparing autocorrelation between the regime-switching model and our model, it is important to note that autocorrelation is calculated differently for the two models. For the regime-switching models, one large simulation was performed and the ACF was calculated on this one simulation. Our model, which generates many simulations in order to provide a posterior predictive distribution, determined the correlation between the y_{n+1} and each subsequent observation in the simulation. In short, the regime-switching model calculated the ACF from one simulation, and our model's autocorrelation is an aggregate measure of many simulations. Therefore, the autocorrelation comparison may be limited.

This model does not perform as quickly as the regime switching models. Although we provide greater flexibility in autocorrelation decay (i.e. the ability to increase short-term autocorrelation by increasing the mass parameter), this does not provide more accurate results. Further research is needed to determine if there are other advantageous properties of this model over regime-switching models.

Another limitation of this analysis is the potential for survivor bias in our sample of companies from the S&P 500. A comprehensive set of data for all companies which have ever been part of the S&P 500 was unavailable at the time of analysis. Thus the results of our stock simulations, sampled from current members of the S&P 500, is limited since it does does not account for companies which failed. It is unknown how predictions would have changed if we had accounted for companies which either failed or are no longer part of the S&P 500.

As mentioned previously , the distance matrix for this model can be chosen arbitrarily. This is a useful feature of our model. While this feature of our model has been explored, preliminary results indicate this did not significantly affect the autocorrelation. Perhaps applying this model to different setting will make this feature more useful.

6 Conclusion

Overall, this model successfully clustered the data based both on the value of the asset return, as well as its proximity to neighboring returns. In particular, our model was successful in capturing clusters based on their precision. This model produced predictions comparable to current regime-switching models. The increased flexibility in autocorrelation decay may be useful in special cases performing short-term forecasting. However, this model is slower than regime-switching models and may therefore not be preferred in many instances.

References

[1] Booth, P., Chadburn, R., Haberman, S., James, D., Khorasanee, Z., Plumb, R. H., & Rickayzen, B. (2004). *Modern actuarial theory and practice.* Chapman & Hall/CRC.

[2] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance 1* (2), 223-236.

[3] Ding, Z. and C. Granger (1996). Modeling volatility persistence of speculative returns: a new approach. *Journal of Econometrics 73* (1), 185-215.

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

[5] Hardy, M. (2003). *Investment guarantees: modeling and risk management for equity-linked life insurance* (Vol. 168). Wiley.

[6] Hommes, C. (2002). Modeling the stylized facts in finance through simple nonlinear adaptive systems. *Proceedings of the National Academy of Sciences of the United States of America 99* (Suppl 3), 7221-7228.

[7] Lux, T. (2009). Stochastic behavioral asset pricing models and the stylized facts. *Handbook of Financial Markets: Dynamics and Evolution 1*, 161.

[8] Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2), 249-265.

[9] Rydn, T., T. Tersvirta, and S. sbrink (1998). Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics 13* (3), 217-244.