Bayesian Nonparametrics for Volatility Clustering and Prediction with Application to Bitcoin Prices

Asher A. Hensley
Department of Electrical and Computer Engineering
Stony Brook University
Stony Brook, NY, United States
asher.hensley@stonybrook.edu

Petar M. Djuric
Department of Electrical and Computer Engineering
Stony Brook University
Stony Brook, NY, United States
petar.djuric@stonybrook.edu

***Abstract*—Interest in digital currencies has intensified over the past year by investors, researchers, and governments. For the case of Bitcoin in particular, the potential for rapid wealth creation is coupled with highly volatile price behavior. Researchers have begun to address the volatility forecasting problem using traditional methods such as GARCH and its variants, however there are numerous alternatives for gaining insight into this new erratic market as a result of the recent surge of machine learning research. In this paper, we propose a Bayesian nonparametric approach based on a latent regime switching model governed by preferential attachment processes. Such processes provide a compelling explanation for the observed volatility clustering dynamics of actual Bitcoin prices. We present a Gibbs sampling procedure for inference and derive a Value-at-Risk (VaR) predictor for future closing prices. We then evaluate the performance of Bitcoin VaR predictions through backtesting experiments which we show to closely match theoretical values.**

***Index Terms*—Yule-Simon Process, Chinese Restaurant Process, Markov Chain Monte Carlo, Gibbs Sampling, Value at Risk, Time Series, Log Returns, Cryptocurrency, Digital Currency**

I. Introduction

With the introduction of digital currencies such as Bitcoin, the old system of a centralized monetary policy is being called into question. Aside from the advertised transparency of the blockchain technology for transactions, one of the main differences between Bitcoin and traditional currencies is the highly volatile behavior of the exchange rates. While volatility forecasting has always been a challenging problem, the dynamics of digital currencies seem to be an enigma with their potential for large gains or losses at any moment. Several researchers have begun to study these dynamics using traditional Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) models and their variants [1]–[4]. However, the time is right for a fresh look with state-of-the-art machine learning methods.

In this paper, we propose a latent regime switching generative process based on our previous work on Hidden Markov Models with preferential attachment dynamics in [5]. Here we use the term "regime" to denote segments of the exchange rate time series with the same return statistics. The idea behind our approach is that price fluctuations are a "bursty" process where long runs of stable behavior are punctuated by abrupt jumps. We describe a Markov Chain Monte Carlo inference procedure to fit our model to observed data and then derive and Value-at-Risk (VaR) predictor based on the measurement predictive posterior distribution which we evaluate through backtesting experiments on the Bitcoin return sequence. In doing so, we show that our VaR exceedance frequencies closely match theoretical values which we verify through unconditional and conditional coverage hypothesis tests.

II. Background

This work is loosely based on infinite state Hidden Markov Models (HMMs) first studied by Beal using Hierarchical Dirichlet Processes (HDPs) in [6]. The infinite state HMM was later refined by Teh in [7], however the use of HDPs created rapid state changes which led to excessive state learning. This problem was addressed by Fox who modified the HDP-HMM to prefer self transitions in what was called the "sticky" HDP-HMM [8]. More recent work in this area has focused on real-time segmentation [9], human dynamics anomaly detection [10], and vehicle behavior modeling [11]. Our approach is a departure from these models in that the switching process is non-Markovian. In particular, the switching probabilities are proportional to the time the process has spent in its current state. This results in sequences with volatility clustering in time with power law distributed regime durations.

Volatility clustering is an old idea, first noted by Mandelbrot in the 1960s [12]. This market characteristic has been studied by numerous other authors [13]–[16] and is one of many regularities that have been reported across several financial markets, instruments, and time periods [17]. Other reported regularities of market returns include fat tails, spectral whiteness, long range dependence, and heavy tailed volatility cluster durations [18]–[22]. Although the behavior of the Bitcoin exchange is a departure from traditional markets, our research indicates that the return sequence still adheres to these general properties, just at a more extreme level.

The key to our model is to control regime durations using a linear preferential attachment process called the Yule-Simon process. Although the term "preferential attachment" is commonly associated with link formation in complex networks due to the Barabasi-Albert model [23], it can refer to any "rich-get-richer" process. The original concept can be traced back

to the work of Markov in 1906 on urn models with dependent trials [24]. Markov studied urn processes where each time a ball was drawn, the ball would be replaced with an additional ball of the same color before the next trial. Some readers may recognize this as Pólya's urn scheme, which was rediscovered some years later by Eggenberger and Pólya in 1923 [25].

Around the same time, Yule used a similar idea to model the distribution of biological genera after observing their population sizes tended to follow a power law [26]. The basic idea behind his model was that the number of generic mutations within a genus was proportional to the number of species within the genus [25]. Thus, as a genus grows larger, the more mutations it accumulates, which causes it to grow even larger. A similar phenomenon was also observed in linguistics where word frequency distributions were found to follow power laws [27]. In the 1950s, Simon reached the same result as Yule, albeit by a different approach [28]. Today they share the name of the Yule-Simon distribution in honor of their findings.

The connection between Pólya urn schemes and Simon's model was subsequently pointed out by Price in 1975 while studying the distributions of bibliometric citations [29]. An important generalization of the Pólya urn model was the extension to an infinite number of ball colors by Blackwell and MacQueen [30]. They show the limiting distribution of this model is the so called "Ferguson" distribution [31], which we know today as the Dirichlet process. A closely related process important for this work is the Chinese Restaurant Process (CRP) [32] which results from the Rao-Blackwellized [33] Dirichlet Process Mixture Model [34]. Posterior sampling techniques for these models were studied extensively through out 1990s for both conjugate [35] and non-conjugate [36] prior distributions.

III. Methods

A. Generative Model

The generative model we propose can be written as the following stochastic process:

$$\lambda_j | c, d \sim \text{Gamma}(c, d) \tag{1}$$

$$z_k | z_1, ..., z_{k-1}, \gamma \sim \text{CRP}(\gamma) \tag{2}$$

$$s_t | \mathbf{x}, \alpha \sim \text{Bernoulli}(\alpha(n_{x_{t-1}} + \alpha)^{-1}) \tag{3}$$

$$x_t = x_{t-1} + s_t \tag{4}$$

$$y_t | x_t, \mathbf{z}, \boldsymbol{\lambda} \sim \mathcal{N}(0, \lambda_{z_{x_t}}^{-1}) \tag{5}$$

The process proceeds by creating volatility regimes according to equations (3) and (4) which together comprise a Yule-Simon process with a regime indicator variable x_t for each measurement y_t. The first value x_1 is always initialized to 1 and subsequent values of x_t are incremented each time s_t is equal to 1 which will occur with probability $p_t = \alpha(n_{x_{t-1}} + \alpha)^{-1}$. Here, the quantity $n_{x_{t-1}}$ is the number of samples in the x_{t-1}^{th} regime. As this number grows large, the switching probability decreases due to the preferential attachment to the current regime. The distribution of regime run-lengths n_k can be shown to follow the Yule-Simon distribution $p(n_k = \ell | \alpha) = \alpha B(\ell, \alpha + 1)$ which follows a power law in the tail with an exponent of $-(\alpha + 1)$.

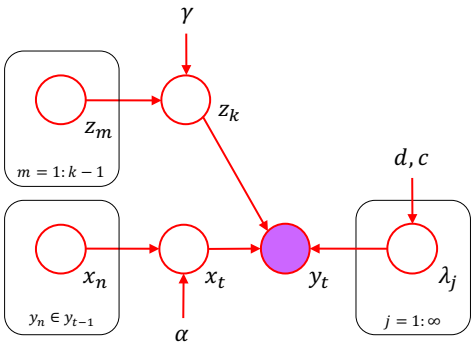

Fig. 1: Generating process graphical model representation. Each node represents a random variable, with shaded nodes indicating that the variable has been observed. Variables without nodes are hyperparameters. Arrows denote dependencies and boxes denote repetitions. For example, the box containing z_m is shorthand for $z_1, z_2, ..., z_{k-1}$.

For each regime, a table assignment is drawn from a Chinese Restaurant Process (CRP) based on the current "seating" arrangement. Each table is associated with a volatility precision λ which is the inverse of the variance for the Gaussian distribution. Each time a new table is created in the CRP, a new precision is randomly drawn from a gamma prior with shape factor c and rate factor d.

Recall the CRP is a generalization of the Pólya urn scheme to an infinite number of ball colors. Instead of balls, the CRP uses a restaurant metaphor where each trial consists of a customer selecting a table in a restaurant with infinite capacity. Each customer selects a table with a probability proportional to the number of customers already seated at the table. Customers will select an unoccupied table with probability proportional to the hyperparameter γ.

The observed log returns are drawn from a zero-mean normal distribution $\mathcal{N}(0, \lambda_{z_{x_t}}^{-1})$ with variance $\lambda_{z_{x_t}}^{-1}$ where z_k is the volatility precision indicator for regime k. A graphical representation of this process for the instantaneous variable dependencies is shown in **Figure 1**.

B. Inference

Inference for the generative process proceeds by using a Gibbs sampling strategy. The basic idea behind Gibbs sampling is to construct a Markov chain with the same target distribution as the model posterior and iterate until convergence. Once converged we draw samples from the posterior and perform data analysis. The Gibbs sampling algorithm we use is shown in Algorithm 1 where $\mathbf{x}_{-t}$ and $\mathbf{z}_{-k}$ are used to denote the set of all values of $\mathbf{x}$ and $\mathbf{z}$ with the t^{th} and k^{th} values removed (respectively). Sampling the regime indicator sequence $\mathbf{x} = [x_1, ..., x_T]$ requires some care as it is only locally exchangeable. Therefore, the approach we take is to only sample those x_t lying on the boundary of the regime switch points in the current state of the Markov Chain.

□ **Algorithm 1:** Gibbs Sampler

Initialize: $\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}$

Repeat:

Sample $x_t \sim p(x_t|\mathbf{x}_{-t}, y_t, \mathbf{z}, \boldsymbol{\lambda}, \alpha) \quad \forall\ t$

Sample $z_k \sim p(z_k|\mathbf{z}_{-k}, \mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}, \gamma) \quad \forall\ k$

Sample $\lambda_j \sim p(\lambda|\mathbf{x}, \mathbf{y}, \mathbf{z}, c, d) \quad \forall\ j$

There are 7 cases to be considered when sampling $\mathbf{x}$ which are shown in **Table I**. To simplify notation we first make the following definitions:

$$F_k = \mathcal{N}(y_t|0, \lambda_{z_k}^{-1})\delta_{x_t,k} \tag{6}$$

$$H_k = \frac{\alpha}{1+\alpha}\text{St}(y_t|0, c/d, 2c)\delta_{x_t,k} \tag{7}$$

where $\text{St}(y_t|\mu, \tau, \nu)$ denotes the Student's-t distribution with mean μ, precision τ, and degrees of freedom ν, and $\delta_{x_t,k} = 1$ if $x_t = k$ and 0 else. We first identify the case C and then sample regime indicator sequence using the corresponding distribution:

$$\begin{aligned}
x_1|C_1 &\sim \frac{n_1 - 1}{n_1 + \alpha}F_1 + H_2 \\
x_1|C_2 &\sim \frac{n_2}{n_2 + \alpha + 1}F_2 + H_1 \\
x_t|C_3 &\sim \frac{n_j - 1}{n_j + \alpha}F_j + \frac{n_{j-1}}{n_{j-1} + \alpha + 1}F_{j-1} + H_{\bar{k}} \\
x_t|C_4 &\sim \frac{n_{j-1}}{n_{j-1} + \alpha + 1}F_{j-1} + \frac{n_{j+1}}{n_{j+1} + \alpha + 1}F_{j+1} + H_j \\
x_t|C_5 &\sim \frac{n_j - 1}{n_j + \alpha}F_j + \frac{n_{j+1}}{n_{j+1} + \alpha + 1}F_{j+1} + H_{\bar{k}} \\
x_T|C_6 &\sim \frac{n_L - 1}{n_L + \alpha}F_L + H_{L+1} \\
x_T|C_7 &\sim \frac{n_{L-1}}{n_{L-1} + \alpha + 1}F_{L-1} + H_L
\end{aligned} \tag{8}$$

Here we denote the current value of x_t before sampling as j, the number of observations T, n_ℓ as the number of observations assigned to the ℓ^{th} regime partition, and the number volatility regime partitions as L.

Case	Example
C_1: no boundary ($t = 1$)	$\mathbf{x} = [(1), 1, 2, ...]$
C_2: right boundary ($t = 1$)	$\mathbf{x} = [(1), 2, 2, ...]$
C_3: left boundary	$\mathbf{x} = [..., j-1, (j), j, ...]$
C_4: double boundary	$\mathbf{x} = [..., j-1, (j), j+1, ...]$
C_5: right boundary	$\mathbf{x} = [..., j, (j), j+1, ...]$
C_6: no boundary ($t = T$)	$\mathbf{x} = [..., L-1, L, (L)]$
C_7: left boundary ($t = T$)	$\mathbf{x} = [..., L-1, L-1, (L)]$

TABLE I: Volatility regime sampling cases with examples. The regime indicator to be sampled is shown in parenthesis; T denotes the last data point in the observations and L denotes the last regime partition. The above examples are used to determine which PDF to use when resampling the regime boundaries. Note that only samples on the partition boundaries are resampled, all other samples are skipped.

To sample the regime volatility assignment sequence $\mathbf{z} = [z_1, ..., z_L]$, we use an approach similar to that of Neal [36]. The main difference is that instead of each customer representing one data point, now each customer is associated with multiple data points (one for each return in the volatility regime). That is, each customer represents a different *regime* instead of a different data point. Thus, to sample z_k we need to take the product of the likelihoods as follows:

$$p(z_k = j|\mathbf{z}_{-k}, \mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}, \gamma) \propto \frac{r_{-k,j}}{C}\prod_{t \in P_k}\mathcal{N}(y_t|0, \lambda_j) \tag{9}$$

where $r_{-k,j}$ is the number of customers sitting at the j^{th} table with the k^{th} customer removed, P_k is the set of measurements assigned to customer k, and $C = L - 1 + \gamma$. A customer will choose to sit at a new table with probability:

$$p(z_k = K + 1|\mathbf{z}_{-k}, \mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}, \gamma) \propto \frac{\gamma}{C}\prod_{t \in P_k}\text{St}(y_t|0, c/d, 2c) \tag{10}$$

where K is the number of occupied tables under the current seating arrangement.

Sampling each λ_j proceeds using the conjugacy property of the gamma and normal distributions. For each table j we assign all data indices associated with each customer (regime) sitting at table j to the set $\mathcal{Y}_j$. We then sample the λ_j gamma posterior conditioned on $\mathcal{Y}_j$

After each pass through the latent variables, there's the option to resample the hyperparameters. Manually setting the hyperparameters beforehand can be challenging because this adds an additional model selection problem to the inference process. In order to sample the hyperparameters, we must endow each with a hyperprior distribution and sample the hyperposterior after each Gibbs iteration. For this work we use the following hyperpriors:

$$\alpha \sim \text{Gamma}(a_\alpha, b_\alpha) \tag{11}$$

$$\gamma \sim \text{Discrete}(\boldsymbol{\beta}_\gamma) \tag{12}$$

$$c \sim \text{Discrete}(\boldsymbol{\beta}_c) \tag{13}$$

$$d \sim \text{Gamma}(a_d, b_d) \tag{14}$$

where a is the shape factor, b is the rate factor, and $\boldsymbol{\beta}$ is the set of discrete bins. For c, d, and γ, sampling proceeds by randomly drawing a value from hyperposterior computed directly using the conjugacy properties of the priors and likelihoods. To sample α we use an additional Gibbs sampling algorithm proposed by Leisen [37].

C. Prediction

We make predictions about the next log return using the measurement posterior predictive distribution which can be written as follows:

$$\begin{aligned}
p(y_{t+1}|\boldsymbol{\Theta}) \propto n_{x_t} p(y_{t+1}|\boldsymbol{\Theta}, &x_{t+1} = x_t) + \\
&\alpha p(y_{t+1}|\boldsymbol{\Theta}, x_{t+1} \neq x_t)
\end{aligned} \tag{15}$$

where n_{x_t} is the number of samples assigned to regime x_t, and $\boldsymbol{\Theta}$ denotes the set of latent variables $\{\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}, \alpha, \gamma, c, d\}$

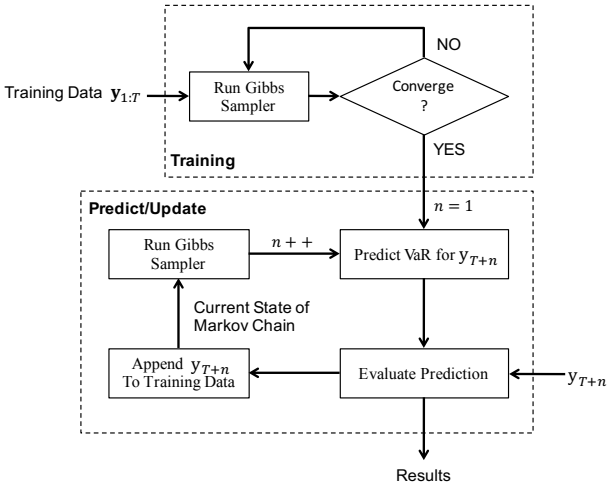

Fig. 2: Experiment procedure diagram.

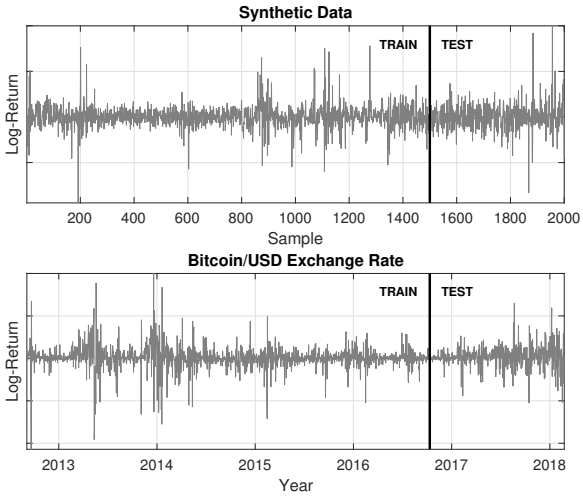

Fig. 3: Training/test splits for inference experiments.

up to time t. For the case of $x_{t+1} = x_t$ we have:

$$p(y_{t+1}|\mathbf{\Theta}, x_{t+1} = x_t) = \mathcal{N}(y_{t+1}|0, \lambda_{z_{x_t}}^{-1}) \tag{16}$$

For the case of $x_{t+1} \neq x_t$ we have:

$$p(y_{t+1}|\mathbf{\Theta}, x_{t+1} \neq x_t) = \sum_{j=1}^{K} \frac{r_j}{x_t + \gamma} \mathcal{N}(y_{t+1}|0, \lambda_j^{-1}) + \frac{\gamma}{x_t + \gamma} \mathrm{St}(y_{t+1}|0, c/d, 2c) \tag{17}$$

where the number of customers in $\mathbf{z}$ sitting at table j is denoted by r_j.

Using this distribution, we make Value-at-Risk (VaR) predictions which determine a loss threshold θ which will only be exceeded $q\%$ of the time. Formally, the VaR can be computed by solving the following equation for θ:

$$\int_{-\infty}^{\theta} p(y_{t+1}|\mathbf{\Theta}) dy_{t+1} = \frac{n_{x_t} \Phi(\lambda_{z_{x_t}}^{1/2}\theta) + \alpha G(\theta)}{n_{x_t} + \alpha} = q \tag{18}$$

where,

$$G(\theta) = \sum_{j=1}^{K} \frac{r_j}{x_t + \gamma} \Phi(\lambda_j^{1/2}\theta) + \frac{\gamma}{x_t + \gamma} T_{2c}(\theta\sqrt{c/d}) \tag{19}$$

$\Phi(\cdot)$ is the unit normal CDF and $T_\nu(\cdot)$ is the unit Student's-t CDF with ν degrees of freedom. To test this predictor, we make hundreds of VaR threshold predictions conditioned on the past for a given time in the future and then count the number of times this threshold is exceeded by actual losses. Statistical tests are then done to evaluate the adequacy of the VaR model.

Complications arise because we are sampling the model posterior, which leads to a sampled measurement predictive posterior distribution. That is, each time we sample the model posterior, we are sampling an entirely different measurement posterior predictive distribution. Therefore, measurement predictions will vary from sample to sample. A reasonable approximation is to average the measurement posterior predictive distributions before computing the VaR threshold. This approach leads to the following "sample" predictive posterior:

$$\hat{p}(y_{t+1}|\mathbf{\Theta}) = \frac{1}{N_G} \sum_{i=1}^{N_G} p(y_{t+1}|\mathbf{\Theta}^{(i)}) \tag{20}$$

where $\mathbf{\Theta}^{(i)}$ denotes the i^{th} Gibbs sample of the latent variables and N_G denotes the total number of Gibbs samples used in the calculation.

IV. Experiments

The approach we take to analyze the inference algorithm is given in **Figure 2**. We begin by running the Gibbs sampling algorithm for an extended period on the training data, until the log-posterior has stabilized. We then make a VaR prediction for the next unobserved measurement, measure the result, and append new measurement to the training data. The state of the Markov Chain is then updated to account for the new measurement by rerunning the Gibbs sampler for additional iterations. We then proceed to make new VaR predictions for the next measurement and repeat the loop.

For this paper we ran inference experiments on (1) Synthetically generated log returns (SYN), and (2) Bitcoin/USD exchange rate log returns (BTC). Log returns are defined as $y_t = \log P_t - \log P_{t-1}$, where P_t is the closing price on day t. The synthetic data set was created by running the generative model in the forward direction with known parameters with the purpose of verifying the inference algorithm works as designed. The Bitcoin data set was downloaded from `http://www.coindesk.com/price/` and consists of daily log returns from from August 3rd 2012 to January 23rd 2018. Each data set consists of 2000 samples with a 1500/500 training/test split as shown in **Figure 3**.

VaR predictions were made based on the 1% and 5% percentiles of the measurement posterior predictive distribution using $N_G = 100$ Gibbs samples per prediction. VaR exceedances were recorded into binary "hit" sequences which

	1%		5%	
	SYN	BTC	SYN	BTC
VaR Exceedances	2/500	5/500	25/500	30/500
Exceedance %	0.4%	1.0%	5.0%	6.0%
UC Test p-Value	0.125	1.000	1.000	0.319
IND Test p-Value	0.899	0.750	0.842	0.878
CC Test p-Value	0.306	0.951	0.980	0.602

TABLE II: Value-at-Risk exceedance results.

were tested for frequency and independence using unconditional [38] and conditional [39] coverage hypothesis tests. For additional information on these tests, see Braione and Scholtes [40]. The VaR exceedance and hypothesis test results are given in **Table II**.

V. Conclusion

In this paper we have a taken a fresh look at volatility clustering and prediction for the Bitcoin/USD exchange rate using a new regime switching model governed by Yule-Simon and Chinese Restaurant Processes. The preferential attachment properties of the model offer a compelling explanation of the observed dynamics of actual Bitcoin log returns, which we verified through backtesting experiments. Future work will need to investigate additional markets and datasets to validate the general robustness of the model for deployment.

References

[1] Viviane Y Naimy and Marianne R Hayek. Modelling and predicting the bitcoin volatility using garch models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 8(3):197–215, 2018.

[2] Jeffrey Chu, Stephen Chan, Saralees Nadarajah, and Joerg Osterrieder. Garch modelling of cryptocurrencies. *Journal of Risk and Financial Management*, 10(4):17, 2017.

[3] Paraskevi Katsiampa. Volatility estimation for bitcoin: A comparison of garch models. *Economics Letters*, 158:3–6, 2017.

[4] Ruiping Liu, Zhichao Shao, Guodong Wei, and Wei Wang. Garch model with fat-tailed distributions and bitcoin exchange rate returns. *Journal of Accounting, Business and Finance Research*, 1(1):71–75, 2017.

[5] A.A. Hensley and P.M. Djurić. Nonparametric learning for hidden markov models with preferential attachment dynamics. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pages 3854–3858. IEEE, 2017.

[6] M.J. Beal, Z. Ghahramani, and C.E. Rasmussen. The infinite hidden Markov model. In *Advances in Neural Information Processing Systems*, pages 577–584, 2001.

[7] Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 2012.

[8] E.B. Fox, E.B. Sudderth, M.I. Jordan, and A.S. Willsky. Bayesian nonparametric methods for learning Markov switching processes. *IEEE Signal Processing Magazine*, 27(6):43–54, 2010.

[9] Ava Bargi, Richard Yi Da Xu, and Massimo Piccardi. Adon hdp-hmm: An adaptive online model for segmentation and classification of sequential data. *IEEE transactions on neural networks and learning systems*, 2017.

[10] Takashi Fuse and Keita Kamiya. Statistical anomaly detection in human dynamics monitoring using a hierarchical dirichlet process hidden markov model. *IEEE Transactions on Intelligent Transportation Systems*, 18(11):3083–3092, 2017.

[11] Donghao Xu, Xu He, Huijing Zhao, Jinshi Cui, Hongbin Zha, Franck Guillemard, Stephane Geronimi, and François Aioun. Ego-centric traffic behavior understanding through multi-level vehicle trajectory analysis. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pages 211–218. IEEE, 2017.

[12] B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963.

[13] S. Thurner, J.D. Farmer, and J. Geanakoplos. Leverage causes fat tails and clustered volatility. *Quantitative Finance*, 12(5):695–707, 2012.

[14] A. Gaunersdorfer and C. Hommes. A nonlinear structural model for volatility clustering. *Long Memory in Economics*, pages 265–288, 2007.

[15] S.R. Bentes, R. Menezes, and D.A. Mendes. Long memory and volatility clustering: Is the empirical evidence consistent across stock markets? *Physica A: Statistical Mechanics and its Applications*, 387(15):3826–3830, 2008.

[16] S.T. Rachev, Y.S. Kim, M.L. Bianchi, and F.J. Fabozzi. *Financial models with Lévy processes and volatility clustering*, volume 187. John Wiley & Sons, 2011.

[17] R. Cont. Volatility clustering in financial markets: empirical facts and agent-based models. *Long Memory in Economics*, 2:289–309, 2007.

[18] A. Chakraborti, I.M. Toke, M. Patriarca, and F. Abergel. Econophysics review: I. empirical facts. *Quantitative Finance*, 11(7):991–1012, 2011.

[19] D.M. Guillaume, M.M. Dacorogna, R.R. Davé, U.A. Müller, R.B. Olsen, and O.V. Pictet. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, 1(2):95–129, 1997.

[20] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. 2001.

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

[22] M. Liu. Modeling long memory in stock market volatility. *Journal of Econometrics*, 99(1):139–171, 2000.

[23] A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[24] A.A. Markov. Extension of the law of large numbers to dependent quantities. *Izv. Fiz.-Matem. Obsch. Kazan Univ.(2nd Ser)*, 15:135–156, 1906.

[25] M.V. Simkin and V.P. Roychowdhury. Re-inventing Willis. *Physics Reports*, 502(1):1–35, 2011.

[26] G.U. Yule. A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London. Series B*, 213:21–87, 1925.

[27] G.K. Zipf. The psycho-biology of language. 1935.

[28] H.A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

[29] D.S. Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American society for Information science*, 27(5):292–306, 1976.

[30] D. Blackwell and J.B. MacQueen. Ferguson distributions via Pólya urn schemes. *The annals of statistics*, pages 353–355, 1973.

[31] T.S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pages 209–230, 1973.

[32] D.J. Aldous. Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XIII 1983*, pages 1–198. Springer, 1985.

[33] A.E. Gelfand and A.F.M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

[34] C.E. Antoniak. Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The annals of statistics*, pages 1152–1174, 1974.

[35] S.N. MacEachern. Estimating normal means with a conjugate style dirichlet process prior. *Communications in Statistics-Simulation and Computation*, 23(3):727–741, 1994.

[36] Radford M Neal. Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265, 2000.

[37] F. Leisen, L. Rossini, and C. Villa. A note on the posterior inference for the Yule-Simon distribution. *arXiv preprint*, 2016.

[38] Paul H Kupiec. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2):73–84, 1995.

[39] Peter F Christoffersen. Evaluating interval forecasts. *International economic review*, pages 841–862, 1998.

[40] Manuela Braione and Nicolas K Scholtes. Forecasting value-at-risk under different distributional assumptions. *Econometrics*, 4(1):3, 2016.