Arthur Tarso Rego

Dynamic Volatility Models for Market Risk and Portfolio Analysis

Belo Horizonte, Brasil

2020

Arthur Tarso Rego

Dynamic Volatility Models for Market Risk and Portfolio Analysis

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutor em Estatística.

Universidade Federal de Minas Gerais – UFMG

Instituto de Ciências Exatas

Programa de Pós-Graduação

Orientador: Prof. Dr. Thiago Rezende dos Santos

Belo Horizonte, Brasil

2020

Rego, Arthur Tarso

R343d Dynamic volatility models for market risk and portfolio analysis [manuscrito] / Arthur Tarso Rego. - 2020.
160 f. il.

Orientador: Thiago Rezende dos Santos
Tese (doutorado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento Estatística.
Referências: f.123-129
.
1. Estatística – Teses. 2. Analise de séries temporais – Teses. 3. Processo estocástico – Teses. 4. Modelo dinâmico – Teses. 5. Análise multivariada - Teses I. Santos,Thiago Rezende dos. II. - Universidade Federal de Minas Gerais;Instituto de Ciências Exatas, Departamento de Estatística. III.Título.

CDU 519.2(043)

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa
CRB 6ª Região nº 1510

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

UFMG

ATA DA DEFESA DE TESE DE DOUTORADO DO ALUNO NOME ARTHUR TARSO RÊGO, MATRICULADO NO PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA, DO INSTITUTO DE CIÊNCIAS EXATAS, DA UNIVERSIDADE FEDERAL DE MINAS GERAIS, REALIZADA NO DIA 28 DE AGOSTO DE 2020.

Aos Vinte e oito dias do mês de Agosto de 2020, às 13h00, em reunião pública virtual 64 (conforme orientações para a atividade de defesa de tese durante a vigência da Portaria PRPG nº 1819) OU na sala 2076 do Instituto de Ciências Exatas da UFMG, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pelo Colegiado do Programa de Pós-Graduação em Estatística, para julgar a defesa de tese do(a) aluno(a) ARTHUR TARSO RÊGO, intitulada: "*Dynamic Volatility Models for Market Risk and Portfolio Analysis*", requisito final para obtenção do Grau de doutor em Estatística. Abrindo a sessão, o Senhor Presidente da Comissão, Prof(a). Thiago Rezende dos Santos, passou a palavra ao aluno para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do aluno. Após a defesa, os membros da banca examinadora reuniram-se reservadamente sem a presença do aluno e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação:

(x) Aprovada.
( ) Reprovada com resubmissão do texto em ____ dias.
( ) Reprovada com resubmissão do texto e nova defesa em ____ dias.
( ) Reprovada.

Thiago Rezende dos Santos

Fabio Nogueira Demarqui

Cristiano de Carvalho Santos

Fernando Ferraz do Nascimento

João Batista de Morais Pereira

XXXXXXX

O resultado final foi comunicado publicamente ao(à) aluno(a) pelo(a) Senhor(a) Presidente da Comissão. Nada mais havendo a tratar, o(a) Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 28 de agosto de 2020.

Observações:
1. No caso de aprovação da tese, a banca pode solicitar modificações a serem feitas na versão final do texto. Neste caso, o texto final deve ser aprovado pelo orientador da tese. O pedido de expedição do diploma do candidato fica condicionado à submissão e aprovação, pelo orientador, da versão final do texto.
2. No caso de reprovação da tese com resubmissão do texto, o candidato deve submeter o novo texto dentro do prazo estipulado pela banca, que deve ser de no máximo 6 (seis) meses. O novo texto deve ser avaliado por todos os membros da banca que então decidirão pela aprovação ou reprovação da tese.
3. No caso de reprovação da tese com resubmissão do texto e nova defesa, o candidato deve submeter o novo texto com a antecedência à nova defesa que o orientador julgar adequada. A nova defesa, mediante todos os membros da banca, deve ser realizada dentro do prazo estipulado pela banca, que deve ser de no máximo 6 (seis) meses. O novo texto deve ser avaliado por todos os membros da banca. Baseada no novo texto e na nova defesa, a banca decidirá pela aprovação ou reprovação da tese.

Resumo

Lidar com séries temporais financeiras traz muitos desafios à modelagem de dados, dada a existência de caudas pesadas e valores extremos de retorno causados por eventos externos, como eventos políticos, econômicos, desastres naturais ou por especulação. Fornecer *insights* confiáveis e interpretáveis aos agentes do mercado, com base em modelos estatísticos, para basear decisões estratégicas na criação de portfólios de ativos, tomar decisões de arbitragem e gerenciar riscos de investimento é crucial para evitar perdas e precificar corretamente os ativos para o desenvolvimento de estratégias de investimento bem sucedidas.

Rego e Santos (2020) propuseram o Modelo Não-Gaussiano de Volatilidade Estocástica com Saltos (NGSVJ) para avaliação da volatilidade do mercado, que inclui procedimento de inferência automática, permitindo que o modelo seja rápido o suficiente para trazer resultados tangíveis para o usuário executar operações de negociação. A classe de Modelos Dinâmicos (DM), da qual o NGSVJ faz parte, possui uma estrutura flexível que permite a inclusão de novos recursos nos modelos e possui simplicidade de implementação pela ótica computacional. Essa classe de modelos ainda é inexplorada para aplicações financeiras quando comparada às outras classes de modelos comumente usadas na literatura, principalmente baseadas nas classes de Volatilidade Estocástica (SV) e Heterocedasticidade Condicional Autorregressiva Generalizada (GARCH).

Nesta tese, vários desenvolvimentos são feitos usando como base a classe DM e o modelo NGSVJ. No âmbito de séries temporais financeiras univariadas, desenvolvimentos são feitos ao NGSVJ para estimar o grau de liberdade do parâmetro da mistura *Gamma* e incluir uma estrutura de Markov Oculto (HMM) para fornecer ao modelo flexibilidade e interpretabilidade em operações de mercado *intraday* de arbitragem. No âmbito de séries temporais financeiras multivariadas, o Modelo Multivariado de Volatilidade Estocástica com Saltos (MSVJ) foi desenvolvido para permitir aos agentes financeiros estimar a volatilidade e correlação entre os ativos do portfólio simultaneamente e desenvolver estratégias efetivas de gerenciamento de riscos.

Esta tese fornece um conjunto amplo de modelos estatísticos, baseados na classe DM, que podem ser usados em finanças para tomar decisões de arbitragem e investimento, sejam para análise de um único ativo ou portfólio. São apresentados estudos de simulação, bem como aplicações no índice de mercado S&P 500, derivativos de commodities e taxas de câmbio, para ilustrar o desempenho do modelo. Os modelos propostos têm resultados altamente interpretáveis, trazendo grandes desenvolvimentos para a classe de modelos DM e suas aplicações em finanças.

Palavras-chave: Séries Temporais Financeiras. Modelos Dinâmicos. Modelos Estocásticos Multivariados.

Abstract

Dealing with financial time series brings many challenges to data modeling, given the existence of heavy tails and extreme return values caused by external events, such as politics, natural disasters, economical events, or even speculation. Providing reliable and interpretable insights to market agents, based on statistical models, to base strategic decisions on building assets portfolios, taking arbitrage decisions, and managing investment risks is crucial for avoiding losses and correctly pricing assets for developing successful investment strategies.

Rego and Santos (2020) proposed the Non-Gaussian Stochastic Volatility Model with Jumps (NGSVJ) for market volatility evaluation, which includes automatic inference procedure that allows the model to be fast enough to bring tangible results for the user, using an ordinary home computer, to perform trading operations. The Dynamic Models (DM) class, on which the NGSVJ is based, has a flexible structure that enables the inclusion of new features on the models and has implementation simplicity from the computational perspective. The DM class of models is still unexplored for financial applications when compared to the other classes of models commonly used on literature, mainly based on Stochastic Volatility (SV) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) classes of models.

In this thesis, several developments are made using as basis the DM class and the NGSVJ model. For dealing with a single asset, or univariate, financial time series, developments are made to the NGSVJ to be able to estimate the degree of freedom of the gamma mixture parameter and include a Hiden Markov (HMM) to give flexibility and interpretability to the model for applications on arbitrage intraday market operations. For dealing with multiple assets portfolio, or multivariate, financial time series the Multivariate Stochastic Volatility Model with Jumps (MSVJ) was developed, based on DM structure, to enable financial agents to estimate the volatility and correlation between portfolio assets and effectively develop a risk management strategy.

This thesis provides a wide set of statistical models, based on DM class, that can be used in finance for taking arbitrage and investment decisions, whether it is used for analyzing a single asset or a portfolio. Simulation studies are presented as well as applications on the S&P 500 market index, commodity derivatives, and exchange rates, to illustrate model performance. The proposed models have highly interpretable results, bringing major developments to the DM class of models and their applications on finance. The proposed models are robust in the sense to incorporate several stylized characteristics of return data, bringing major developments to the NGSVJ and their applications.

Keywords: Financial Time Series. Dynamic Models. Multivariate Stochastic Models.

List of Figures

List of Tables

List of abbreviations and acronyms

ARCH	Autoregressive Conditional Heteroscedasticity
BIC	Bayesian Information Criterion
BRN	Brent Crude Futures
BZUS	Brazil / U.S. Foreign Exchange Rate
CAUS	Canada / U.S. Foreign Exchange Rate
CHUS	China / U.S. Foreign Exchange Rate
DNUS	Denmark / U.S. Foreign Exchange Rate
HKUS	Hong Kong / U.S. Foreign Exchange Rate
INUS	India / U.S. Foreign Exchange Rate
JPUS	Japan / U.S. Foreign Exchange Rate
KOUS	South Korea / U.S. Foreign Exchange Rate
MAUS	Malaysia / U.S. Foreign Exchange Rate
DM	Dynamic Model
DLM	Dynamic Linear Model
EWMA	Exponentially Weigthed Moving Average
FFBS	Forward Filetring Backward Sampling
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
HMM	Hidden Markov Model
ICE	Intercontinental Exchange
log L	Log Likelihood
MCMC	Markov Chain Monte Carlo
MEWMC	Multivariate Exponentially Weighted Moving Covariance Matrix
MH	Metropolis Hastings

MSV	Multivariate Stochastic Volatility
MSVJB	Multivariate Stochastic Model with Jumps and Beta-Bartlett Evolution
MSVJM	Multivariate Stochastic Volatility Model With Jumps and Matrix-Beta Evolution
NGSVJ	Non-Gaussian Stochastic Volatility Model with Jumps
NGSVJ-HMM	Non-Gaussian Stochastic Volatility Model with Jumps and Hidden Markov Model
NGSVJ-MS	Non-Gaussian Stochastic Volatility Model with Jumps with Metropolis Steps
RMSE	Root Mean Square Error
SV	Stochastic Volatility
VaR	Value at Risk

Contents

1 Introduction

To understand the behavior of asset prices is essential for capital allocation decisions between the available investment options. Arbitrage, buying and selling to make gains of differing prices, catch the attention of several market players by the chance of getting rapid gains, over other investment options. Several assets can be negotiated in that way, including stocks, commodities, futures, currencies, etc. On the other hand, some investors prefer to build a portfolio, containing multiple assets and hold this position for some period of time to gain from dividends and asset valuation. For such investors, a diversified portfolio is essential in reducing risks and increase the overall performance of the investment portfolio.

For evaluating market volatility precisely, the model must take into account that returns usually follow a heavy-tailed distribution and are susceptible to market anomalies, such as the impact of speculative movements bringing abnormal changes to return of asset.

Concerning the inferential procedure, under the Bayesian perspective, the stochastic volatility models commonly used are mostly based on intensive computational methods, e.g., Markov Chain Monte Carlo (MCMC) methods using Metropolis-Hastings algorithms, that can fail on bringing tangible results on the required time frame, especially when dealing with high-dimensional data.

Rego and Santos (2020) proposed the Non-Gaussian Stochastic Volatility Model with Jumps (NGSVJ) for market volatility evaluation, that includes automatic inference procedure that allows the model to be fast enough to bring tangible results for the user, using an ordinary home computer, to perform trading operations. The NGSVJ allows returns to assume non-Gaussian distributions and includes jumps to catch the speculative movements of the market, preventing their negative impact on overestimating volatility.

Professional traders and investment funds usually rely on statistical models to take investment and arbitrage decisions, however, non-professional traders usually do not have access to complex models, and rely mostly on market reports, basic graphic analysis or even intuition in order to take such decisions. Nevertheless, even if they had access to those statistical models, most would lack the required knowledge to interpret the output correctly, since it involves statistical and financial concepts that are not common on the average user's daily environment.

With that in mind and the simplicity of the NGSVJ, this thesis will explore the incorporation of a Hidden Markov Model (HMM) to translate the volatility results from a mathematical language to a user-friendly language, so that it can be used on day-to-day operation for investment decisions. Furthermore, a natural extension of the NGSVJ is the multivariate case, to be able to evaluate the risk of an entire asset portfolio.

In the multivariate case, the main interest is on understanding how one asset affects the other inside the portfolio since diversification requires that the correlation between them is weak or even negative to mitigate risks from one specific asset. Thus, estimating the covariance matrix will provide the needed information to build an investment strategy. Prado and West (2010) present the Matrix-Beta evolution for a multivariate Dynamic Model that is effective to small dimensionality, due to strong constraints on both degrees of freedom from Wishart distribution and discount factor for the covariance matrix evolution. The Matrix-Beta evolution can be adapted on the NGSVJ to produce a multivariate model with jumps on returns and heavy tail distribution and take advantage of a retrospective analysis for sampling the covariance matrix from a smoothed distribution, but also some development must be made in order to solve the constrains of Matrix-Beta Evolution.

Another alternative is the use of a multivariate dynamic model with the Bartlett evolution equation. The evolution for the covariance matrix can flexible the constrains that exist on the Matrix-beta evolution, allowing a higher dimensionality to the model. Although, a retrospective analysis was still not proposed for this model, which is a big challenge that will be addressed by this doctoral dissertation.

In summary, the objective of this thesis is to provide a wide set of statistical models that can be used in finance for taking arbitrage and investment decisions, whether it is used by a professional or non-professional investor and with a single asset or a portfolio, the proposed model has highly interpretable results, bringing major developments to the DM class of models and their applications on finance. The proposed models are robust in the sense to incorporate several stylized characteristics of return data, bringing major developments to the NGSVJ and their applications.

1.1 The Data and Motivations

Dealing with financial data brings major challenges to data analysis. The first challenge, highlighted by Alexander *et al.* (2017) and Subrahmanyam (2019) is on data quality. According to Alexander *et al.* (2017), different financial firms report data differently, so that it is a challenge to integrate, aggregate, and analyze these data, *e.g.*, financial data can be reported by transactions or quotes on specific time frames. Also, because financial data are acquired by different data systems, particularities of those systems can interfere with data acquisition: specific countries' holidays, timezone, market open hours, etc, which is especially relevant when dealing with multivariate data analysis, since data integration is relevant on granting analysis quality.

In finance applications it is common to use as observations the log-returns time series. Log-returns are calculated by taking the natural log of the assets price at time t divided by the price at time $t-1$, that is, $log\left(\frac{P_t}{P_{t-1}}\right)$. By doing so it is assumed that returns

are compounded continuously rather than discrete, taking advantages of mathematical properties of its continuity for data modeling.

Furthermore, financial data also brings challenges to data modeling. The most common methods used for financial log-return modeling rely on statistical methods that assume a Gaussian distribution for asset price returns, however, empirical studies have shown that they are better described by heavy-tailed distributions. Jondeau and Rockinger (2003) investigates the existence and persistence of skewness and kurtosis of various financial time series taken at the daily frequency, finding that for many series they are persistent. Resnick (2007) defines a heavy-tail as a characteristic of phenomena where the probability of huge value is relatively big and exemplifies financial log-returns as heavy-tailed phenomena. Also, financial log-returns are subject to speculative movements of the market, which are an additional challenge in modeling its behavior.

In this chapter, we briefly present the financial data used in this thesis and a review of available models on literature that motivate the development of the NGSVJ, developed by Rego and Santos (2020).

1.1.1 Financial Data Sets

The first dataset contains S&P 500 stock index log-returns from January 2, 1980, to December 31, 1999. Excluding weekends and holidays, there are 5,054 daily observations for the S&P index. Data were obtained from the Yahoo Finance platform, and the purpose of using this specific dataset is to keep comparable results to Rego and Santos (2020). Figure 1 shows log-returns time series and histogram, and Table 1 shows summary statistics for this dataset.

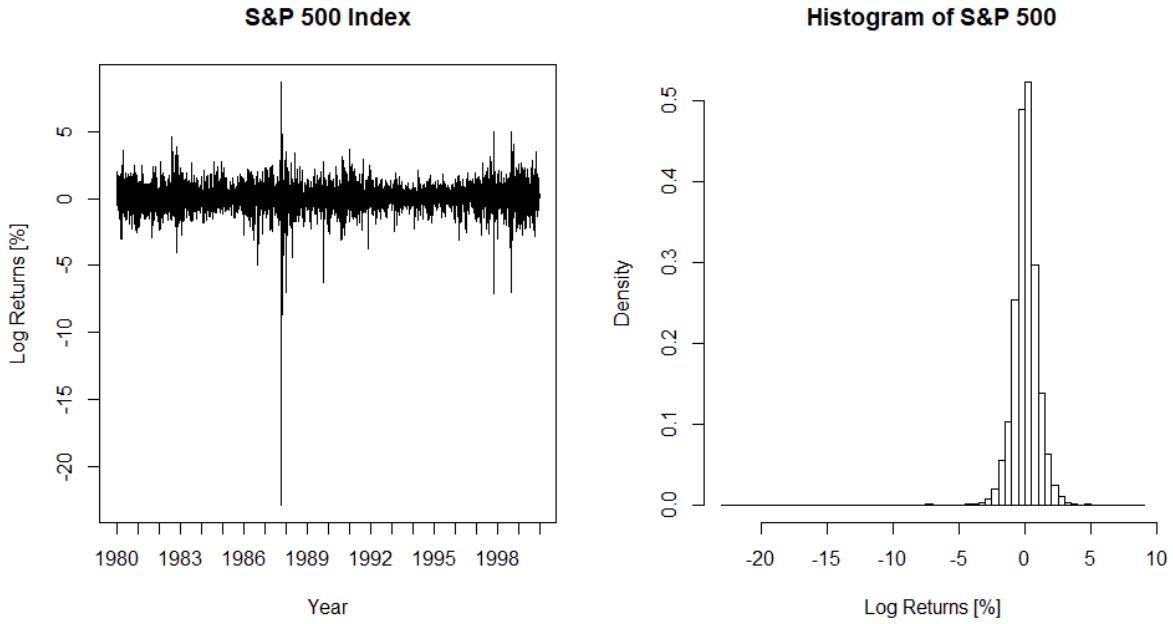


Figure 1 – S&P 500 dataset: Log-returns time series and histogram.

	Mean	Variance	Skewness	Kurtosis	Min	Max
S&P500	0.0521	0.9989	-2.6357	63.0710	-22.8997	8.7089

Table 1 – S&P 500 dataset: Summary Statistics

The second dataset contains Brent Crude futures, ICE:BRN, intraday log-returns time series, consisting of Brent Crude futures, ICE:BRN, log returns from August 15, 03:00, 2018 to August 16, 23:59, 2018, in a total of 2,472-minute observations, obtained from the Yahoo Finance. Figure 2 shows log-returns time series and histogram, and Table 2 shows summary statistics for this dataset.

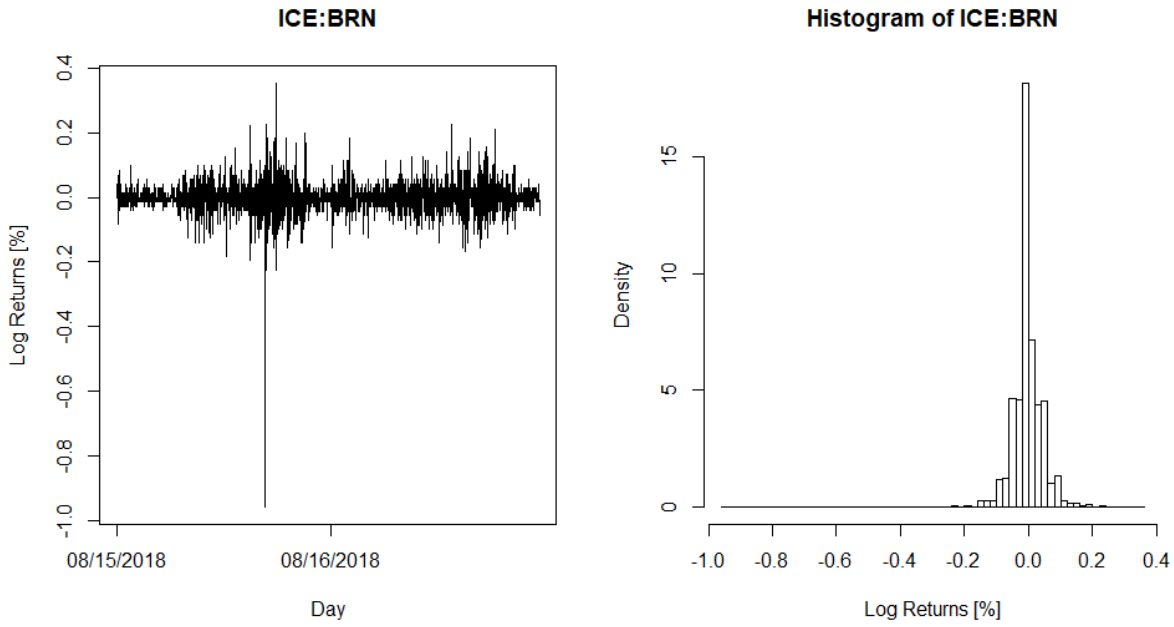


Figure 2 – ICE:BRN dataset: Log-returns time series and histogram.

	Mean	Variance	Skewness	Kurtosis	Min	Max
ICE:BRN	-0.0005	0.0490	-2.7660	64.5076	-0.9575	0.3536

Table 2 – ICE:BRN dataset: Summary statistics

The third dataset contains nine exchange rates against U.S. dollar: Brazilian Reals (BZUS), Canadian Dollar (CAUS), Chinese Yuan (CHUS), Danish Kroner (DNUS), Hong Kong Dollar (HKUS), Indian Rupees (INUS), Japanese Yen (JPUS), South Korean Won (KOUS), Malaysian Ringgit (MAUS), obtained from Federal Reserve of St. Louis website. The dataset contains 2,020 daily observations from December 2010 to January 2019. Figures **??** to **??** show log-returns time series and histogram, and Table 3 shows summary statistics for each time series of this dataset, individually.

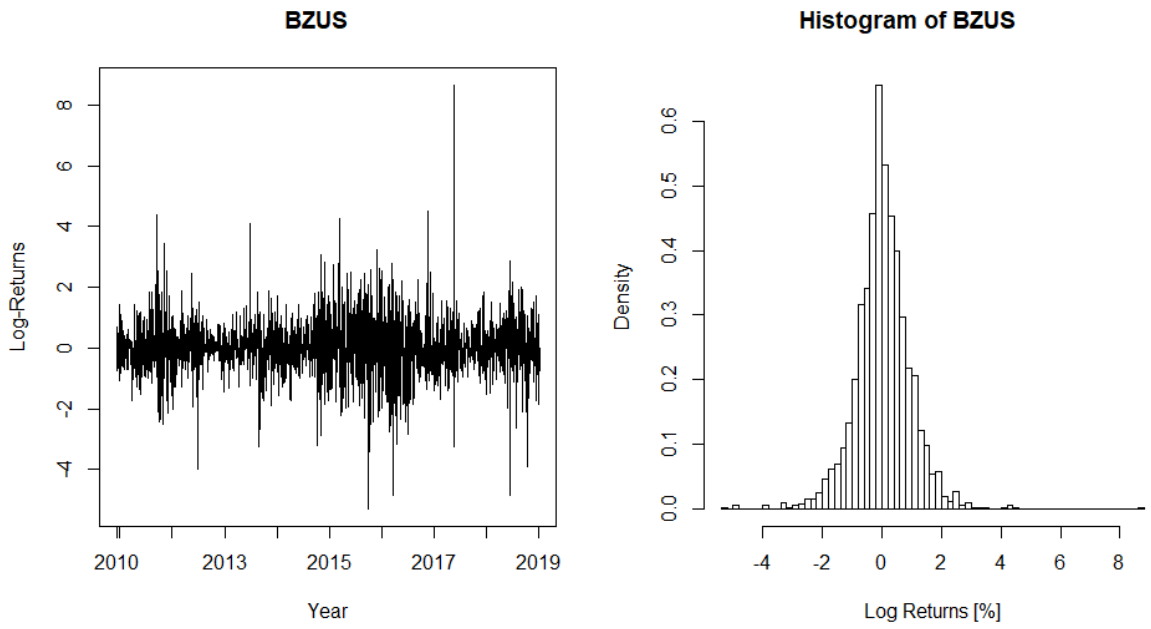


Figure 3 – Exchange rates dataset: Log-returns time series and histogram for BZUS.

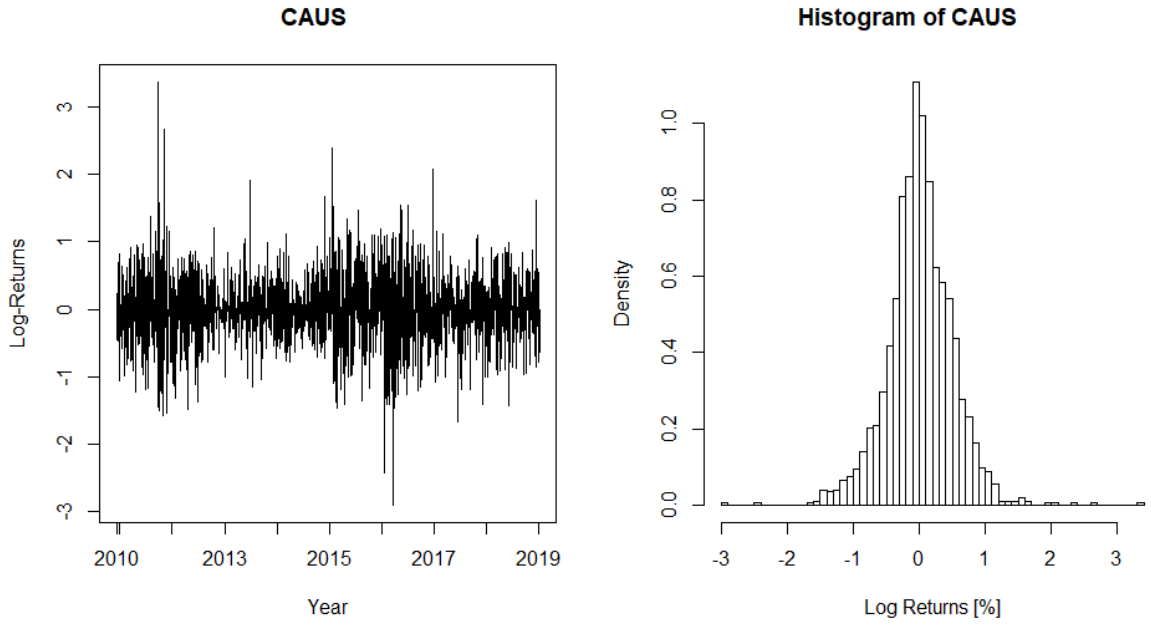


Figure 4 – Exchange rates dataset: Log-returns time series and histogram for CAUS.

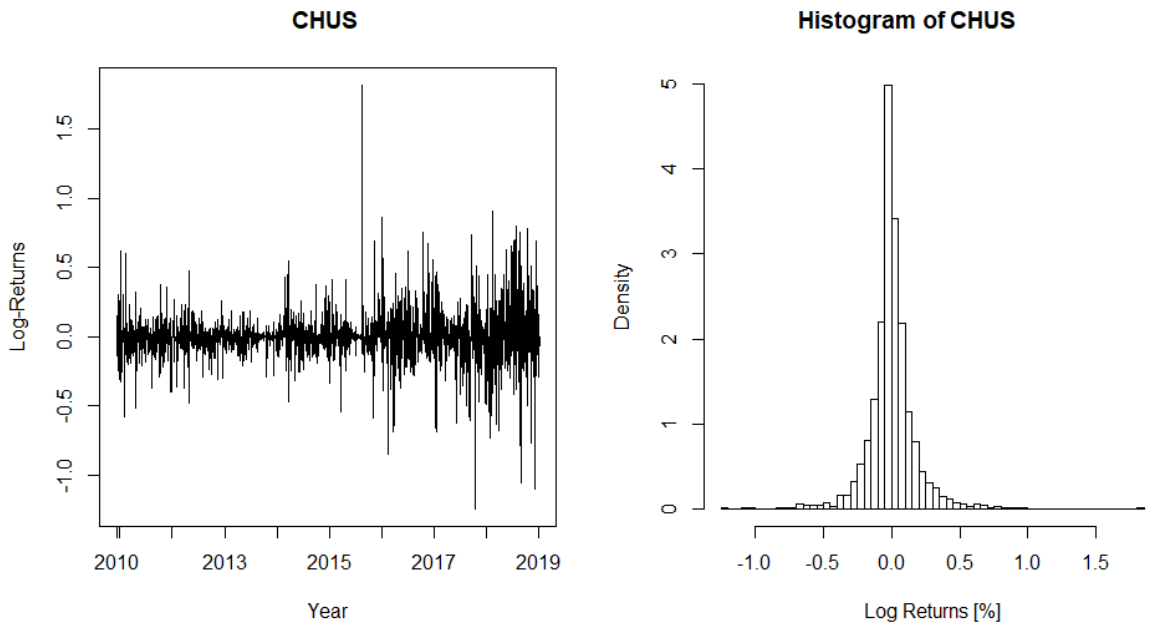


Figure 5 – Exchange rates dataset: Log-returns time series and histogram for CHUS.

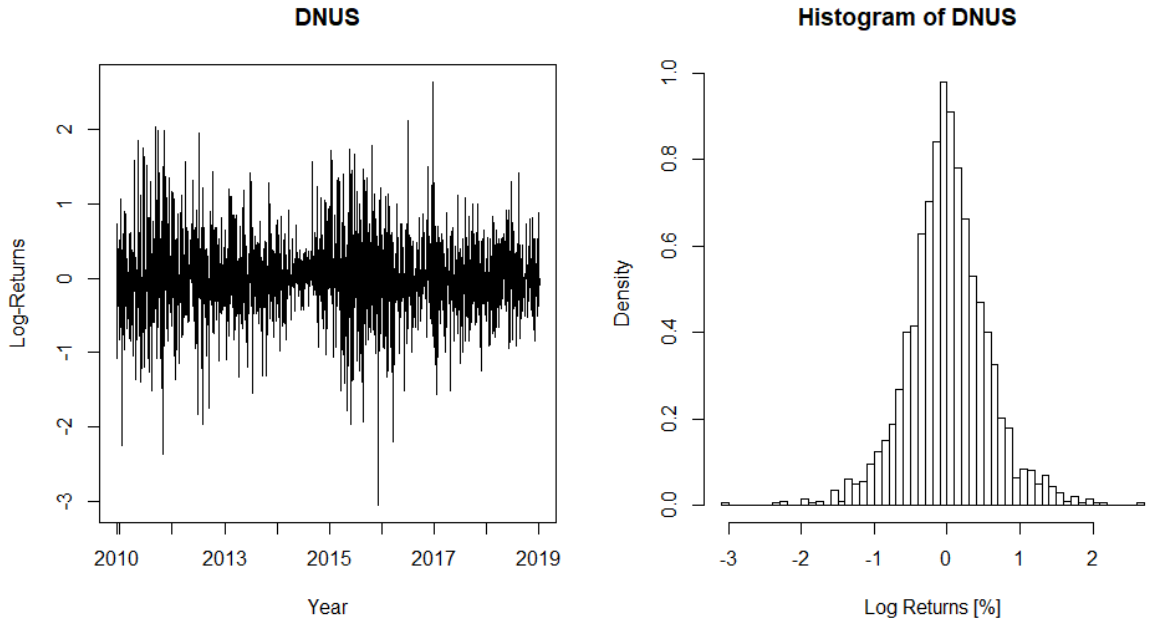


Figure 6 – Exchange rates dataset: Log-returns time series and histogram for DNUS.

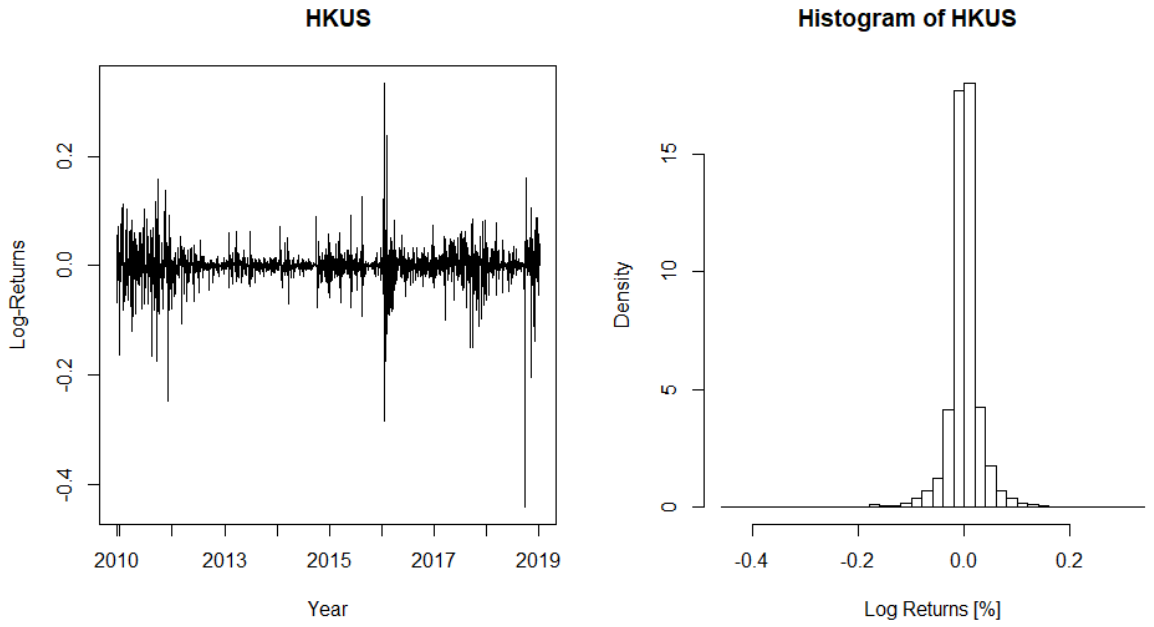


Figure 7 – Exchange rates dataset: Log-returns time series and histogram for HKUS.

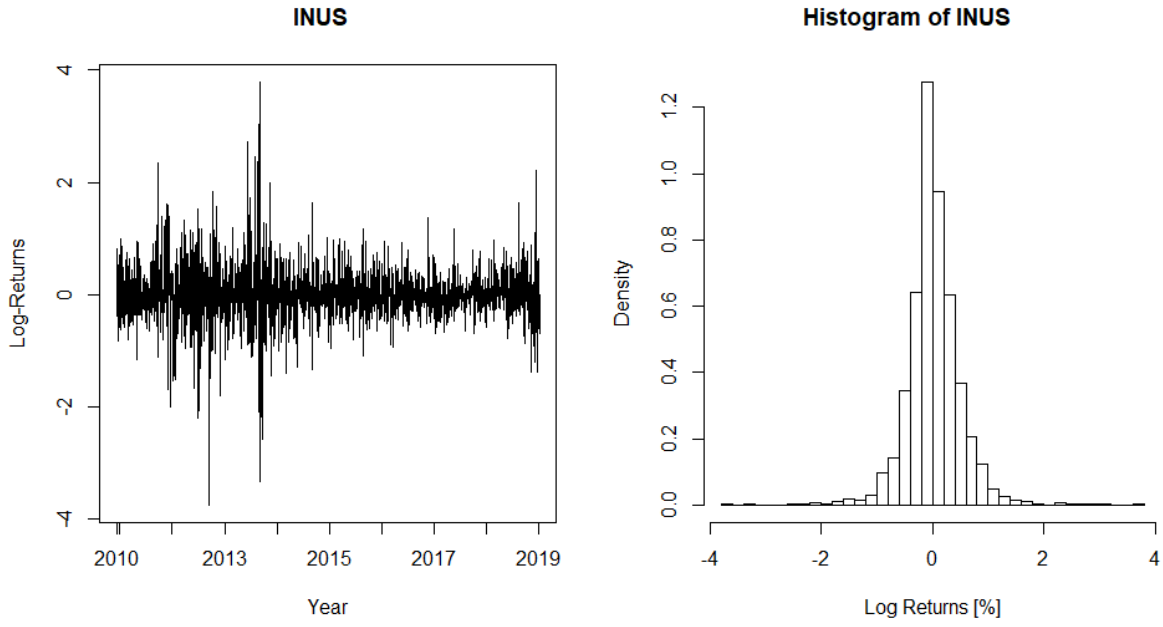


Figure 8 – Exchange rates dataset: Log-returns time series and histogram for INUS.

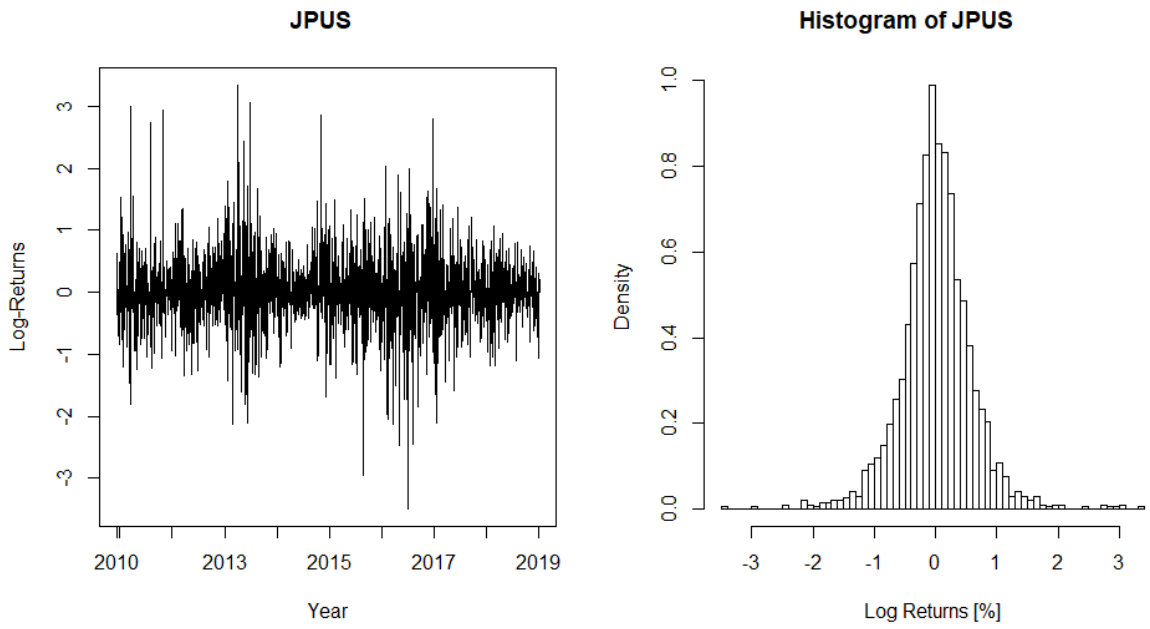


Figure 9 – Exchange rates dataset: Log-returns time series and histogram for JPUS.

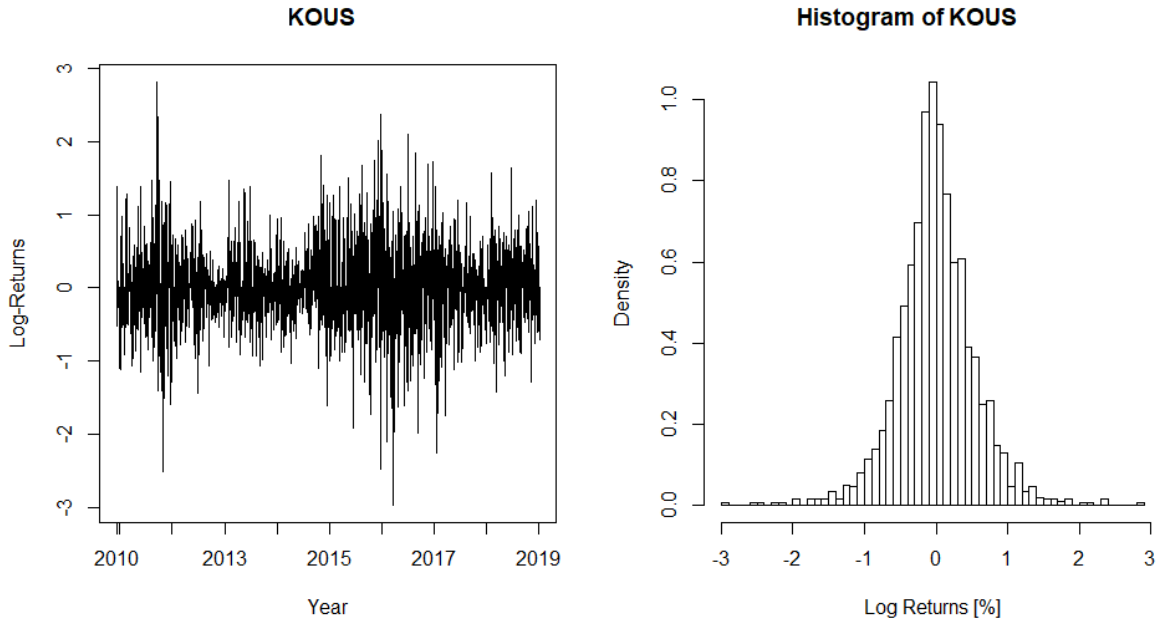


Figure 10 – Exchange rates dataset: Log-returns time series and histogram for KOUS.

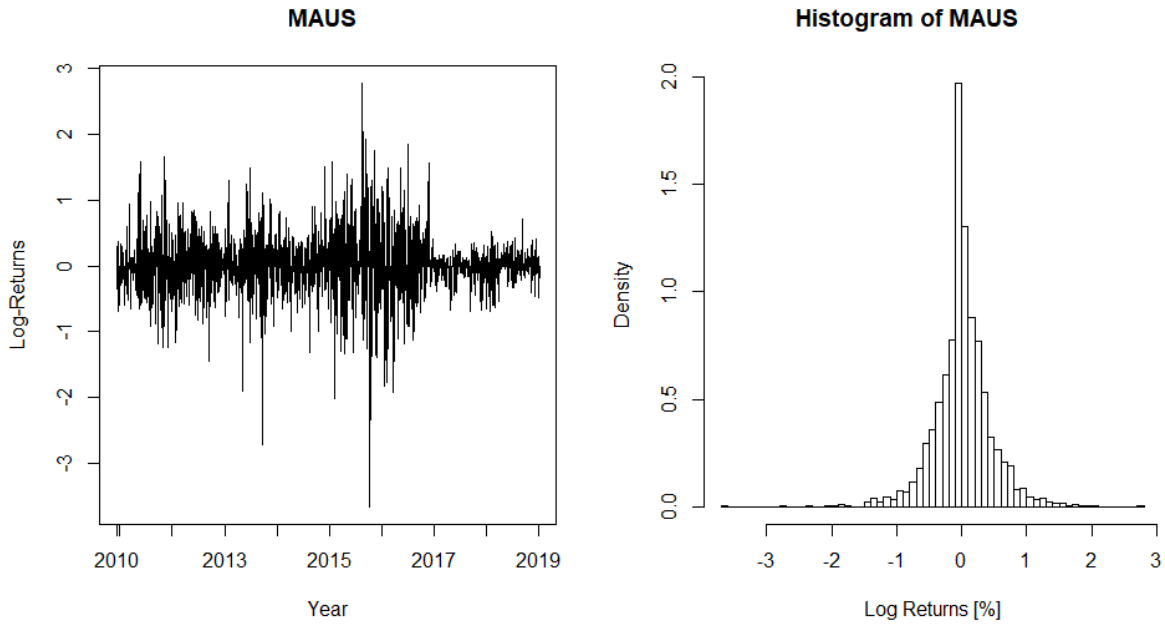


Figure 11 – Exchange rates dataset: Log-returns time series and histogram for MAUS.

	Mean	Variance	Skewness	Kurtosis	Min	Max
BZUS	0.0385	0.9397	0.2370	8.9720	-5.2991	8.6670
CAUS	0.0140	0.4967	0.0513	5.8700	-2.9001	3.3678
CHUS	0.0015	0.1805	0.4244	14.0396	-1.2417	1.8161
DNUS	0.0073	0.5534	0.0130	4.8866	-3.0555	2.6461
HKUS	0.0004	0.0338	-1.0342	32.4697	-0.4422	0.3345
INUS	0.0215	0.5007	0.1921	10.8799	-3.7560	3.7919
JPUS	0.0127	0.5885	0.1102	6.9096	-3.4977	3.3428
KOUS	-0.0010	0.5397	0.0488	5.5663	-2.9630	2.8113
MAUS	0.0136	0.4507	-0.3009	8.7820	-3.6571	2.7750

Table 3 – Exchange rates dataset: Summary Statistics

A remarkable characteristic of the datasets here presented is the presence of heavy tails, evidenced by the kurtosis statistical measure. For S&P500 and ICE:BRN log-returns kurtosis is over 60, which strongly indicates the presence of heavy tails. Exchange rates dataset time series also have kurtosis higher then what was expected to assume a Gaussian distribution. Also, some degree of asymmetry can be detected in some time series, such as S&P500, ICE:BRN and, HKUS, as evidenced by skewness statistical measure. Another characteristic of the financial data here presented is the existence of extreme values, e.g., the minimum log-return of -22.8997 on S&P 500 log-returns time series, which presents another challenge on data modeling.

1.1.2 Motivations for Further Developments on NGSVJ

According to Ozturk and Richard (2015), the volatility of asset returns has been a focus of financial econometrics for the last three decades, with the majority of available models developed on ARCH-GARCH and SV classes of models. Those models have a wide range of features for data modeling: The inclusion of jumps for capturing speculative movements of the market was introduced by Eraker, Johannes and Polson (2003) and further developed in earlier works as Gong and Zhuang (2016), Kirkby, Nguyen and Cui (2017) and Chaim and Laurini (2018); Omori *et al.* (2007) develop the leverage effect for the SV class of models, as Engle and Siriwardane (2018) for the GARCH class; The heavy-tail modeling is discussed by Nakajima and Omori (2009) and also Warty, Lopes and Polson (2018), with the inclusion of a Gamma process for volatility to capture the excess of returns, while Jondeau and Rockinger (2003) develop the GARCH model to capture skewness and kurtosis by modeling residuals as a generalized Student-t distribution; Feunou and Tédongap (2012) develop the SV model for dealing with the skewness by allowing current asset returns to be asymmetric conditional on current factors and past information, what we term contemporaneous asymmetry; Smith (2002) reviews the inclusion of Markov-switching structure on GARCH and SV classes of models for modeling volatility according to current market state.

Generalizations were also made for the ARCH-GARCH and stochastic volatility (SV) to extend their usage to the multivariate case for dealing with a portfolio risk management challenge. Harvey, Ruiz, and Shephard (1994), however, highlight that such generalizations to multivariate series can be difficult to estimate and interpret. According to Yu and Meyer (2006), the multivariate ARCH models have attracted a lot of attention in modern finance theory and enjoyed voluminous empirical applications as presented on a survey made by Bauwens et al. (2006) on multivariate GARCH models. Yet more limited due to the complexity of implementation and computational sampling issues, Yu and Meyer (2006) state that developments were also made for the SV class of model for extending to the multivariate case. Chib, Omori, and Asai (2009) present a collection of works in multivariate stochastic volatility (MSV) models available on literature that includes leverage effects, a mean factor MSV model and the inclusion of dynamic correlations trough a Wishart Process on the MSV model. Another common approach is the usage of copula methods, as explored by Dias and Embrechts (2004), Lee and Long (2009), Patton (2013), and Nasri and Rémillard (2019), which also has a complex implementation.

Another class of models that can be used for modeling volatility is the dynamic models (DM) in which the NGSVJ, developed by Rego and Santos (2020), resides. As discussed by Rego and Santos (2020) this approach has the advantage of a simpler implementation and a flexible model structure since the model is written in the space state form. Also, the DM class of models is relatively unexplored when compared to works on

ARCH and SV classes, rising the possibility of developing extensions and new applications to expand the usage of DM models.

Triantafyllopoulos (2008) develops a Bayesian procedure for estimation and forecasting of the volatility of multivariate time series using a matrix-variate DLM for volatility. This approach allows the estimation of the covariance matrix for an entire assets portfolio simultaneously, but lacks of structures to accommodate log-returns of financial time series, such as jumps and a heavy tail structure. In a latter work, Triantafyllopoulos (2014) proposes a particle filter algorithm for sequential estimation of volatility and cross-correlation of multivariate financial time series, including a structure so that innovations assume a Skew-T distribution to capture the heavy tails and asymmetry of financial returns, and the inverse of the volatility covariance matrix is modeled via a Wishart autoregressive process.

Lopes, McCulloch and Tsay (2016) present a normal dynamic linear model proposing flexible priors for applications on space-state models, so that posterior can be computed using the Gibbs sampler. However, this model uses univariate DLM for estimating states in each equation rather than using a multivariate structure to jointly draw the states, which increases computational complexity for the proposed algorithm. Nasri and Rémillard (2019) propose a time-dependent and time-independent copula model and univariate dynamic models to couple several log-returns time series for a multivariate approach.

West (2020) presents the Dynamic dependence network models (DDNM) that use the DLM structure together with a graphically structured space-state model, using hyper-inverse Wishart distributions for estimating the correlation between assets log-returns. This model however has its usage limited to lower-dimensional time series, since computational challenges arise as dimensionality grows.

The NGSVJ proposed by Rego and Santos (2020) has the notable advantage of its computational simplicity and a structure that grants an automatic sampling process for parameters, that allows sampling the volatility in a block via Gibbs sampler. This structure allows model parameters to achieve convergence with less MCMC iterations when compared to other SV models on literature, as stated by Rego and Santos (2020), providing fast and reliable estimates and allowing the model to be used in practical situations. Naturally, extensions and the implementation of new features arise to expand its applications and develop the model as a relevant new class on the literature for extensive use on the market.

The NGSVJ model is a Dynamic Model that considers the volatility as being stochastic. The Non-Gaussian comes from the mixture on volatility that leads to a heavy tailed distribution, with the inclusion of a jump component to capture extreme values from observations. Thus, the model is called NGSVJ, not to be confused to a model from the SV class, since it belongs to DM class.

The NGSVJ proposed by Rego and Santos (2020) includes jumps on returns to capture speculative movements, together with a Gamma mixture on volatility to capture the excess of kurtosis of returns, using an estimation procedure entire based on Gibbs Sampling, achieving convergence in fewer iterations, when compared to other methods. Yet, other possibilities were left unexplored, as the estimation of the degree of freedom parameter and the inclusion of a classifier for market states, which has a lot of practical appeal on financial markets, especially for the model interpretability by the final user. Another natural extension of the model is the multivariate case, to evaluate multiple time series jointly, as proposed by Rego and Santos (2020), since most applications on the financial market require a portfolio analysis, instead of a single asset, keeping the implementation and computational simplicity as a remarkable characteristic, in the same way in which the NGSVJ was proposed.

This thesis will address these challenges, which have an immediate impact on the the extensiveness of the model application and its usability, and is a larger step on developing the NGSVJ as a complete alternative to other models extensively explored in the literature, as the ARCH and SV classes models. On next chapter the NGSVJ model will be revisited to review the basis needed to understand further developments made on this thesis. On Chapter 3 the inclusion of a procedure for estimating degree of freedom will be addressed. Chapter 4 discusses the inclusion of the HMM structure on NGSVJ. Chapters 5 and 6 deal with the expansion to the multivariate case.

2 The NGSVJ Revisited

In this chapter, we revisit the Non-Gaussian Stochastic Volatility Model with Jumps on returns (NGSVJ), developed by Rego and Santos (2020) in order to build the theoretical basis needed to understand the models further developed in this thesis. The model has already presented advances, for example, it has a heavy tail with known degrees of freedom and jumps. Although, the model do not incorporate the HMM model and estimated degrees of freedom of Student-t distribution.

2.1 The Non-Gaussian stochastic volatility model with jumps on returns

The NGSVJ for log-return time series $\{y_t\}_{t=1}^{n}$ is given by:

$$
\begin{aligned}
y_t &= \mu_t + J_t^y + \upsilon_t, \qquad \upsilon_t|\gamma_t \sim N(0, \gamma_t^{-1}\lambda_t^{-1}), && (2.1)\\
\lambda_t &= \beta^{-1}\lambda_{t-1}\zeta_t, \qquad \zeta_t|\mathcal{D}_{t-1}, \varphi \sim Beta(\beta a_{t-1}, (1-\beta)a_{t-1}), && (2.2)\\
\gamma_t &\sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right). && (2.3)
\end{aligned}
$$

where,

$J_t^y = \xi_{t+1}^y N_{t+1}^y$,$\xi_{t+1}^y \sim N(\mu_y, \sigma_y^2)$, $Pr(N_{t+1}^y = 1) = \rho_y$, and $Pr(N_{t+1}^y = 0) = 1 - \rho_y$.

In this model y_t follows the DLM defined by $\{1, \upsilon_t, \lambda_t, \gamma_t\}$. y_t represents the log-return in percentage, defined as $y_t = 100 \times (log(P_t) - log(P_{t-1}))$, where P_t is the asset price at time t or continuous return. The jump, J_t^y, is composed by the jump indicator $N_{t+1}^y \in \{0, 1\}$ and magnitude $\xi_{t+1}^y \sim N(\mu_y, \sigma_y^2)$, following the same idea proposed by Eraker, Johannes and Polson (2003). μ_t represents the equilibrium log-return of y_t on time t. γ_t is the variance mixture component, so that using $\gamma_t \sim G(\frac{\nu}{2}, \frac{\nu}{2})$, the unconditional distribution of errors is a $t_\nu(0, 1)$ distribution, λ_t^{-1} is the volatility of returns, initialized as $\lambda_0|\mathcal{D}_0 \sim G(a_0, b_0)$, where $\mathcal{D}_0$ is the set of initial information. $\varphi = \{\beta, \nu\}$ is the vector of static parameters, where β is a discount factor and is, in general, specified, since it is a parameter of complex estimation, requiring more intensive computational methods that were out of the initial scope of Rego and Santos (2020) of keeping the estimation procedure more automatic via Gibbs Sampling. The problem of estimating of the degree of freedom parameter is addressed on this thesis, since a grid analysis was originally made to specify ν. a_{t-1} is the shape parameter of the filtering distribution of λ_t, which is described on details in Gamerman, Santos and Franco (2013) and Rego and Santos (2020).

According to Gamerman, Santos and Franco (2013), the different temporal dependence in the stochastic specifications of the γ_t^{-1} and λ_t^{-1} ensure their identification, except for an arbitrary constant c as $\gamma_t^{-1}\lambda_t^{-1} = (c\gamma_t)^{-1}(c\lambda_t)^{-1}$, causing no concern for the identification of the temporal variation of λ_t^{-1}, which is their most relevant feature.

In order to assure identifiability, we have from Stoyanov and Lin (2011) that for a mixture distribution H to be identifiable, given family $F = \{F_\theta, \theta \in \Theta\}$ of conditional distributions $f(x|\theta)$, there must be only one mixture distribution G on Θ producing H, where:

$$h(x) = \int_\Theta f(x|\theta) dG(\theta)$$

Thus, if the mixture H is identifiable, the mixture model of (X, θ) is also identifiable. According to Gamerman, Santos and Franco (2013) the Student t distribution is obtained when mixing Normal and Gamma distributions as in Equation 2.1, so that the resulting distribution is identifiable. As, in this case, H is the Student t distribution, that is identifiable, so is the mixture model.

The NGSVJ model contrasts with other stochastic volatility models available in the literature by its implementation simplicity. It has a formulation that allows the full conditional posterior distributions to be available, so that a Gibbs sampler can be used to sample from the full conditional posterior distributions, making model implementation simpler. Also, it includes a mixture on variance in order to achieve non-Gaussian distribution for innovations, is an identifiable model and has a jump structure to capture outliers and avoid their direct effect on the volatility.

2.2 Bayesian Inference

For the mean parameter of the log-returns, μ_t, a prior $N(m_0, C_0)$ is specified and the samples of its posterior distribution can be obtained through a Forward Filtering Backward Sampling (FFBS) algorithm, available in Prado and West (2010). For ease notation, let $\Phi = (\mu, \underset{\sim}{\gamma}, \underset{\sim}{\lambda}, \mu_y, \sigma_y^2, \underset{\sim}{\xi}, \underset{\sim}{\rho}, \underset{\sim}{N^y}, \nu)$, with under-tilde notation refering to vector, e.g. $\underset{\sim}{\lambda} = \lambda_1, \ldots, \lambda_n$. The brackets notation means that the parameter inside is excluded from Φ , e.g., $\Phi_{[-\lambda]} = (\mu, \underset{\sim}{\gamma}, \mu_y, \sigma_y^2, \underset{\sim}{\xi}, \underset{\sim}{\rho}, \underset{\sim}{N^y}, \nu)$, and $\mathcal{D}_t$ all available information until time t .

Independent priors were chosen so that proper posteriors are obtained, simplifying the sample procedure. When no initial information is available, non-informative priori are chosen.

With the prior distribution for λ_t, that is given by $\lambda_t|\mathcal{D}_{t-1}, \Phi_{[-\lambda]} \sim G(\beta a_{t-1}, \beta b_{t-1})$, with a_{t-1} and b_{t-1} being shape and rate hyperparameters respectively, the updating

distribution is:

$$p(\lambda_t|\mathcal{D}_t, \Phi_{[-\lambda]}) \sim G\left(\beta a_{t-1} + \frac{1}{2}, \beta b_{t-1} + \gamma_t \frac{(y_t - \mu_t - J_t^y)^2}{2}\right). \tag{2.4}$$

Sampling from $(\underset{\sim}{\lambda}|\mathcal{D}_n, \Phi_{[-\lambda]})$ follows the procedure available in Gamerman, Santos and Franco (2013), where the distribution of $(\lambda_t|\lambda_{t+1}, \Phi_{[-\lambda]}, \mathcal{D}_t)$ is given by:

$$\lambda_t - \beta\lambda_{t+1}|\lambda_{t+1}, \Phi_{[-\lambda]}, \mathcal{D}_t \sim G\left((1-\beta)a_t, b_t\right), \forall t \geq 0, \tag{2.5}$$

Thus, for a fixed discount factor β, an exact sample of the joint distribution $(\underset{\sim}{\lambda}|\mathcal{D}_n, \Phi_{[-\lambda]})$ can be obtained following the algorithm:

1. set $t = n$ and sample $p(\lambda_n|\Phi_{[-\lambda]}, \mathcal{D}_n)$;
2. set $t = t - 1$ and sample $p(\lambda_t|\lambda_{t+1}, \Phi_{[-\lambda]}, \mathcal{D}_t)$;
3. if $t > 1$, go back to step 2; otherwise, the sample of $(\lambda_1, ..., \lambda_n|\Phi_{[-\lambda]}, \mathcal{D}_n)$ is complete.

For the mixture component γ_t, a prior $G(\frac{\nu}{2}, \frac{\nu}{2})$ is defined, which, when mixed as γ_t^{-1}, resulting in Inverse-Gamma, leads to a Student-t with ν degrees of freedom to the innovations, where ν is specified in order to avoid Metropolis step, keeping an automatic and simple procedure for the model estimation. The full conditional posterior distribution is:

$$p(\gamma_t|\mathcal{D}_n, \Phi_{[-\gamma]}) \sim G\left(\frac{\nu}{2} + \frac{1}{2}, \frac{\nu}{2} + \lambda_t \frac{(y_t - \mu_t - J_t)^2}{2}\right). \tag{2.6}$$

The jump sizes ξ_{t+1}^y follow a $N(\mu_y, \sigma_y^2)$. For the mean μ_y a prior $N(m, v)$ is set, resulting in a full conditional posterior:

$$p(\mu_y|\mathcal{D}_n, \Phi_{[-\mu_y]}) \sim \mathrm{N}\left(\frac{m\sigma_y^2 + vn_j\bar{\xi^y}}{\sigma_y^2 + n_j v}, \frac{v\sigma_y^2}{\sigma_y^2 + n_j v}\right). \tag{2.7}$$

For the variance σ_y^2 a prior $IG(\alpha, \beta)$ is assumed, resulting in the full conditional posterior:

$$p(\sigma_y^2|\mathcal{D}_n, \Phi_{[-\sigma_y^2]}) \sim IG\left(\alpha + \frac{n_j}{2}, \beta + \frac{\sum_{\substack{i=1 \\ J_i \neq 0}}^{n} (\xi_{i+1}^y - \mu_y)^2}{2}\right). \tag{2.8}$$

In both cases, n_j is the number of times that the jump is observed, and $\bar{\xi^y}$ the mean of jump sizes ξ_{t+1}^y. As the prior of jump sizes is assumed to follow a Normal distribution, the full conditional posterior is also Normal, given by:

$$p(\xi_{t+1}^y|\mathcal{D}_n, \Phi_{[-\xi]}) \sim N\left(m_\xi^*, v_\xi^*\right). \tag{2.9}$$

where:

$$m_\xi^* = \frac{\mu_y \gamma_t^{-1}\lambda_t^{-1} + y_t\sigma_y^2 - \mu\sigma_y^2}{\sigma_y^2 + \gamma_t^{-1}\lambda_t^{-1}}, \tag{2.10}$$

$$v_\xi^* = \frac{\sigma_y^2\gamma_t^{-1}\lambda_t^{-1}}{\sigma_y^2 + \gamma_t^{-1}\lambda_t^{-1}}. \tag{2.11}$$

For jump probability ρ_y, a prior $Beta(\alpha, \beta)$ is set. The full conditional posterior is given by:

$$p(\rho_y|\mathcal{D}_n, \Phi_{[-\rho_y]}) \sim Beta\left(\alpha + \sum_{i=0}^{n} N_i^y, \beta + n - \sum_{i=0}^{n} N_i^y\right). \tag{2.12}$$

Since the jump indicator N_{t+1}^y can assume only two values, 0 or 1. The posterior probability of observation at $t+1$ be a jump is given by:

$$P(N_{t+1}^y = 1|\mathcal{D}_{t+1}, \Phi_{[-N]}) \propto \rho_y P(y_{t+1}|N_{t+1}^y = 1, \Phi_{[-N]}). \tag{2.13}$$

which is easy to calculate, since $P(y_{t+1}|N_{t+1}^y = 1, \Phi_{[-N]})$ is a Normal distribution. Using the concept proposed by Brooks an Prokopczuk (2011), if $P(N_{t+1}^y = 1|\mathcal{D}_{t+1}, \Phi_{[-N]})$ is greater then a threshold α, then $N_{t+1}^y = 1$. The threshold α is chosen such that the number of jumps identified corresponds to the estimate of the jump intensity ρ_y.

2.3 General Procedure

Let $Y_n = \{y_t\}_{t=1}^n$, $\underset{\sim}{\mu} = \{\mu_t\}_{t=1}^n$, $\underset{\sim}{J} = \{J_t^y\}_{t=1}^n = \{\xi_{t+1}^y N_{t+1}^y\}_{t=1}^n$, $\underset{\sim}{\gamma} = \{\gamma_t\}_{t=1}^n$, $\underset{\sim}{\lambda} = \{\lambda_t\}_{t=1}^n$, $\underset{\sim}{\xi} = \{\xi_{t+1}^y\}_{t=1}^n$, $\underset{\sim}{N} = \{N_{t+1}^y\}_{t=1}^n$ and independent prior probability density $p(\underset{\sim}{\gamma}), p(\mu_y), p(\sigma_y^2), p(\xi), p(\rho_y)$ are set for $\underset{\sim}{\gamma}, \mu_y, \sigma_y^2, \underset{\sim}{\xi}, \rho_y$. Then, a sample of size M from the joint posterior distribution $p(\mu, \underset{\sim}{\lambda}, \underset{\sim}{\gamma}, \mu_y, \sigma_y^2, \underset{\sim}{J}, \rho_y|\mathcal{D}_n)$ is drawn via Gibbs Sampler, which is given by:

1. Initialize $\underset{\sim}{\mu}^{(0)}, \underset{\sim}{\lambda}^{(0)}, \underset{\sim}{\gamma}^{(0)}, \mu_y^{(0)}, (\sigma_y^2)^{(0)}, \underset{\sim}{\xi}^{(0)}, \underset{\sim}{N}^{(0)}$ and $\rho_y^{(0)}$.
2. Set $j = 1$.
3. Sample $\underset{\sim}{\mu}^{(j)}|\mathcal{D}_n, \underset{\sim}{J}^{(j-1)}, \underset{\sim}{\lambda}^{(j-1)}, \underset{\sim}{\gamma}^{(j-1)}$ using FFBS algorithm.
4. Block sample $\underset{\sim}{\lambda}^{(j)}|\mathcal{D}_n, \underset{\sim}{\mu}^{(j)}, \underset{\sim}{J}^{(j-1)}, \underset{\sim}{\gamma}^{(j-1)}$ using algorithm proposed by Gamerman, Santos and Franco (2013).
5. Block sample $\underset{\sim}{\gamma}^{(j)}|\mathcal{D}_n, \underset{\sim}{\mu}^{(j)}, \underset{\sim}{J}^{(j-1)}, \underset{\sim}{\lambda}^{(j)}$, as in Eq. (2.6).
6. Sample $\mu_y^{(j)}|\underset{\sim}{\xi}^{(j-1)}, (\sigma_y^2)^{(j-1)}$ as in Eq. (2.7).
7. Sample $(\sigma_y^2)^{(j)}|\underset{\sim}{\xi}^{(j-1)}, \mu_y^{(j)}$ as in Eq. (2.8).

8. Block sample $\underset{\sim}{J}^{(j)}|\mathcal{D}_n, \underset{\sim}{\mu}^{(j)}, \underset{\sim}{\lambda}^{(j)}, \underset{\sim}{\gamma}^{(j)}, \mu_y^{(j)}, (\sigma_y^2)^{(j)}$ by

   a) Block sample $\underset{\sim}{\xi}^{(j)}|\mathcal{D}_n, \underset{\sim}{\mu}^{(j)}, \underset{\sim}{\lambda}^{(j)}, \underset{\sim}{\gamma}^{(j)}, \mu_y^{(j)}, (\sigma_y^2)^{(j)}$ as in Eq. (2.9).

   b) Block sample $\underset{\sim}{N}^{(j)}|\mathcal{D}_n, \underset{\sim}{\mu}^{(j)}, \underset{\sim}{\lambda}^{(j)}, \underset{\sim}{\gamma}^{(j)}, \underset{\sim}{\xi}^{(j)}$ as in Eq. (2.13).

9. Sample $\rho_y^{(j)}|\underset{\sim}{J}^{(j)}$ as in Eq. (2.12).

10. Set $j = j + 1$.

11. If $j \leq M$, go to 3, otherwise stop.

Since all full conditional posterior distribution has closed-form, only Gibbs sampler steps are used in this procedure.

2.4 Additional Comments

The NGSVJ is innovative on presenting the DM class as an alternative to SV and GARCH classes for volatility modeling with the inclusion of relevant features for finance applications, jumps and a mixture for modeling heavy-tailed distributed observations. Despite the simpler implementation structure, the NGSVJ could benefit of a procedure for estimating the degree of freedom of the mixture component. The inclusion of such procedure would eliminate the grid analysis that is undesirable for practical applications, since it is not an efficient method for setting the parameter ν. Chapter 3 will address this modification by proposing a Jeffrey's prior together with a Metropolis step to estimate ν directly on the MCMC algorithm.

Another feature that can potentially increase the usage of the NGSVJ model on practical applications is the inclusion of a structure to help users to interpret model results in terms of market states. Non professional users, specially those not used to statistical theory, may have difficult to interpret volatility values in numerical terms. As the philosophy behind the NGSVJ is being a simple implementation and universal usage model, the inclusion of such structure is justified. The inclusion of a HMM structure for classifying model states will be addressed on Chapter 4.

3 Estimating degrees of freedom

The problem of estimating the degree of freedom is particularly important when dealing with financial time series and risk management, where heavy-tailed distributions better describe the log-returns data. Lin and Shen (2006) reported substantial improvements of using Student-t distribution over Gaussian distribution, especially when using a credibility interval over 98.5%, when studying the empirical performance of value-at-risk calculations.

One natural extension of the NGSVJ model is to include a Metropolis step to estimate the degrees of freedom, ν, and avoid a grid analysis in order to set this parameter. By estimating the degree of freedom, the automaticity of the model is enhanced, since grid analysis is not an efficient method for setting a parameter, requiring overhead effort each time the model is fit for a new time series. To keep the original essence of the NGSVJ model, the chosen method for estimating ν must be such that requires minimum effort from the user to tune parameters for estimation.

According to Villa and Walker (2014), the estimation of degrees of freedom parameter of a Student-t distribution is typically problematic in Bayesian inference since usual proper priors such as Gamma or Exponential distribution, as proposed by Juárez and Steel (2010) and Ding (2014), do not result in a precise estimation and improper priors lead to improper posteriors, whilst proper priors may dominate the data likelihood. The reason is, as stated by Anscombe (1967) and Fonseca, Ferreira, and Migon (2008) the estimation of degree of freedom parameter, ν, of the Student-t distribution is not straightforward, as the likelihood function tends to infinity as $\nu \to 0$, so that proper priors such as Exponential and Gamma distribution, and improper prior distributions, as proposed by Jacquier, Polson and Rossi (2004) and Juárez and Steel (2006), might lead to improper posterior distributions and, even when the parameter space is restricted to a desired region in which likelihood is bounded, the maximum likelihood estimator may not exist with a positive probability.

Another attempt to address the question of estimating the degree of freedom using a parametric prior is proposed Gelman and Hill (2006), that restricts the degree-of freedom parameter to be at least 2, supporting that it is possible and convenient to assign a uniform distribution to its inverse in this case. Therefore, Simpson et al (2017) refutes the usage of uniform priors on a fixed interval for the degrees of freedom in a Student-t distribution, proposing a penalized complexity prior, and supporting it to have a good behavior for the degrees of freedom estimation, yet it requires numerical computation of the prior, making it a computationally intensive method.

Fonseca, Ferreira and Migon (2008) and Villa and Walker (2014), on the other hand, defend the usage of an objective prior. Jeffrey priors have the disadvantage of the need to be derived for the specific model at hand and the derivation may be difficult, but it allows a non-subjective statistical analysis with adaptive robustness to outliers and may be a good option provided that the posterior distribution is proper. The biggest advantage of adopting an objective prior to estimate the degrees of freedom parameter for the NGSVJ model is keeping its automatic structure, by the addition of a Metropolis step without requiring further tunning or any other computationally extensive method.

Kwok-Wah (2012) compares the usage of two Jeffreys priors, Jeffreys-rule prior and the marginal independence Jeffreys prior, motivated by a practical financial risk management application, with other priors proposed on the literature to estimate quantiles for the Student-t model with unknown degrees of freedom, and concludes that both priors perform better than common priors suggested in the literature. By studying a similar application, on financial data, and obtaining favorable results on the usage of objective priors Kwok-Wah's work provides a strong insight on how to enhance the NGSVJ model on including a method for estimating ν while keeping its original essence of computational simplicity.

3.1 Proposed procedure

In order to estimate the degrees of freedom, ν, in the NGSVJ model, as there is not previous information about ν available, the Jeffreys reference prior proposed by Kwok-Wah (2012) is used:

$$p(\nu \mid \mathcal{D}_0) \propto \frac{\nu^{1/2}(\nu+1)^{1/2}}{\nu+3}\left\{\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right\}^{1/2} \tag{3.1}$$

where $\psi(a) = \Gamma'(a)$ and $\psi'(a) = \Gamma''(a)$ are the digamma and trigamma functions, respectively.

The full conditional posterior distribution for ν is then given by:

$$p(\nu|\mathcal{D}_n, \Phi_{[-\nu]}) \propto p(\nu) \times \frac{(\frac{\nu}{2})^{n\nu/2}}{\prod_{t=1}^{n}\Gamma(\frac{\nu}{2})}\prod_{t=1}^{n}\gamma_t^{-1-\nu/2} exp\left\{-\sum_{t=1}^{n}\frac{\nu}{2\gamma_t}\right\} \tag{3.2}$$

which has no closed-form. A Metropolis step is included in the general procedure exposed in Section 2.3 in order to estimate the degree of freedom parameter.

3.2 Simulation Study

A Monte-Carlo study with 1,000 replications is made to test the efficiency of the proposed procedure to estimate the degrees of freedom, ν, of the gamma mixture parameter. Replications are built using the same procedure and parameter scenario proposed in Rego and Santos (2020).

To generate the volatility, we use:

$$v_t = v_{t-1} + \kappa(\theta - v_{t-1})\Delta + \rho\sigma_v\sqrt{v_{t-1}\Delta}\epsilon_{1,t} + \sigma_v\sqrt{(1-\rho^2)v_{t-1}\Delta}\epsilon_{2,t} \tag{3.3}$$

where $\epsilon_{1,t}$ and $\epsilon_{2,t} \sim N(0,1)$. Synthetic data for returns is then generated from:

$$r_t = N(\mu + \mathcal{J}_t, \gamma_t^{-1}v_t), \qquad \mathcal{J}_t = N_t\xi_t, \tag{3.4}$$

where the jump times, N_t are generated from a Bernoulli(ρ_y), jump sizes ξ_t from $N(\mu_y, \sigma_y^2)$, and γ_t from $G(\frac{\nu}{2}, \frac{\nu}{2})$. Setup of parameters was: log-returns mean $\mu = 0.05$; jump probability $\rho_y = 0.015$; jump magnitude mean $\mu_y = -2.5$ and standard deviation $\sigma_y = 4$; volatility components $\Delta = 1$, $\theta = 0.8$, $\kappa = 0.015$, $\sigma_v = 0.1$, $\rho = 0.4$, $n = 1000$ observations, same used by Rego and Santos (2020).

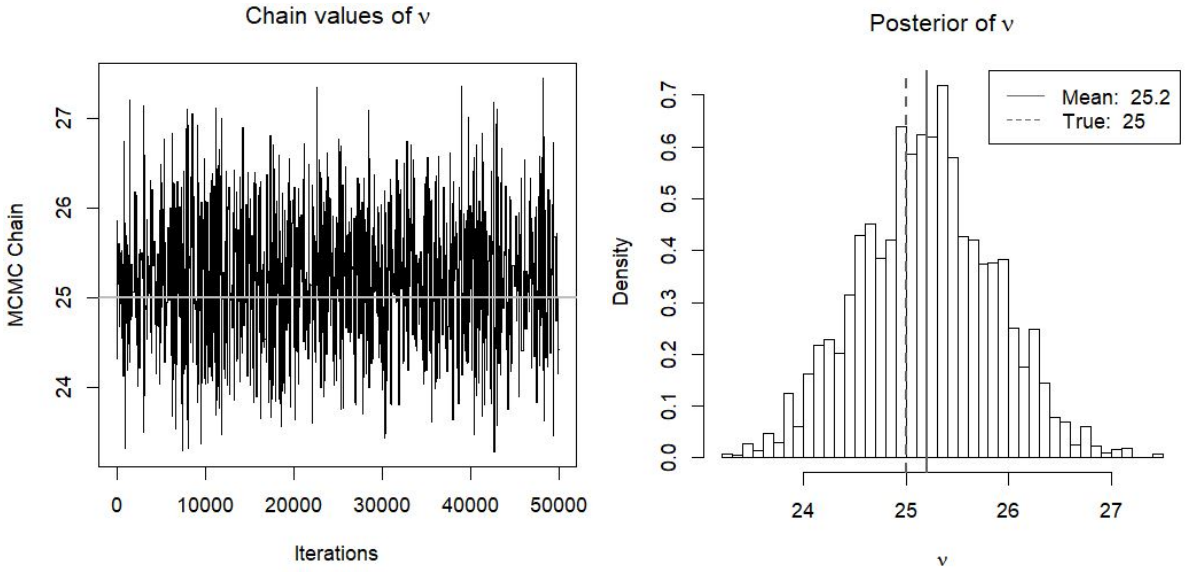
Table 4 shows true values and the summary estimates of the posterior distribution for the degrees of freedom of the NGSVJ, over the 1,000 simulation scenarios. The model was able to closely estimate all the values ranging from $\nu = 1$ to $\nu = 50$.

True Value	**MCMC Mean**	**RMSE**	$\mathbf{Q}_{2.5}$	$\mathbf{Q}_{50}$	$\mathbf{Q}_{97.5}$	**Min**	**Max**
1	1.0001	0.0212	0.9601	0.9999	1.042	0.9179	1.0603
2	2.0272	0.0275	2.0198	2.0272	2.0347	2.0139	2.0386
3	2.8999	0.1002	2.8893	2.9000	2.9106	2.8798	2.9167
4	3.8445	0.1556	3.8293	3.8444	3.8590	3.8224	3.8692
5	5.002	0.1263	4.7623	5.0018	5.2458	4.5341	5.4033
8	7.7958	0.2046	7.7678	7.7956	7.8245	7.7514	7.8490
10	9.9980	0.2593	9.4966	9.9828	10.5221	9.094	10.9532
15	14.9732	0.3763	14.2059	14.975	15.6998	13.9075	16.0706
20	19.9782	0.4914	19.0521	19.9745	20.9383	18.5055	22.0725
25	25.0133	0.6485	23.7900	25.0005	26.2920	23.2166	27.2702
30	29.9944	0.7798	28.5743	29.9658	31.4994	27.6653	33.0256
50	49.9648	1.3626	47.3766	49.9561	52.8543	45.9504	54.3098

Table 4 – Summary estimates of the posterior distribution for the degrees of freedom of the NGSVJ model, ν, over 1,000 simulation scenarios.

Figure 12 shows one realization of a simulation scenario. The proposed procedure was able to estimate the degree of freedom,with $\nu = 25$, of the mixture parameter. MCMC chain convergence was verified trough graphical methods, using different initial values.

Figure 13 shows MCMC chains for different initial values: $\nu_0 = \{5, 20, 50, 100\}$, and convergence is achieved regardless of the initial value used, with $\nu = 25$. Figure 14 shows histograms of posterior distributions for ν for each MCMC chain. Posterior mean estimate is the same up to one decimal place.



(a) MCMC chain for df ν.

(b) Histogram of df ν.

Figure 12 – MCMC results for degree of freedom parameter ν over a simulation scenario. Right graph shows MCMC effective chain with gray line being the true value and left graph shows histogram of MCMC sample with gray dashed line being true value and gray solid line the posterior mean.

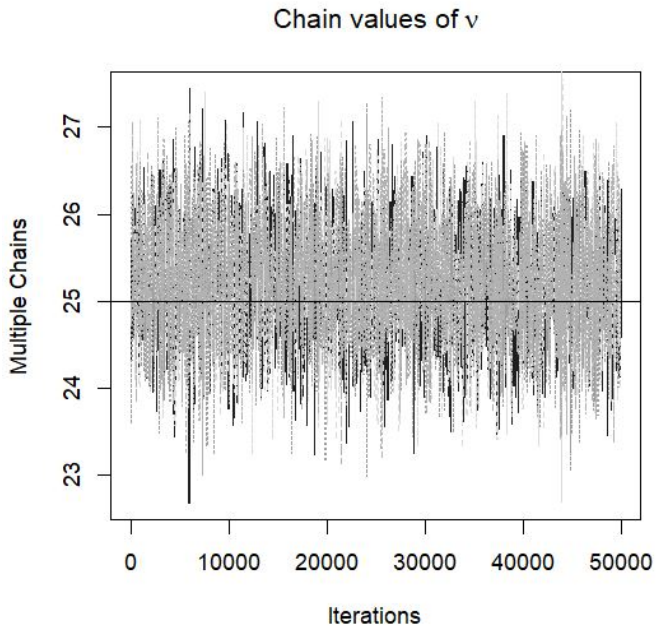


Figure 13 – Multiple MCMC chain using different start values for one realization of simulation study, with black solid line being the true value ($\nu = 25$).

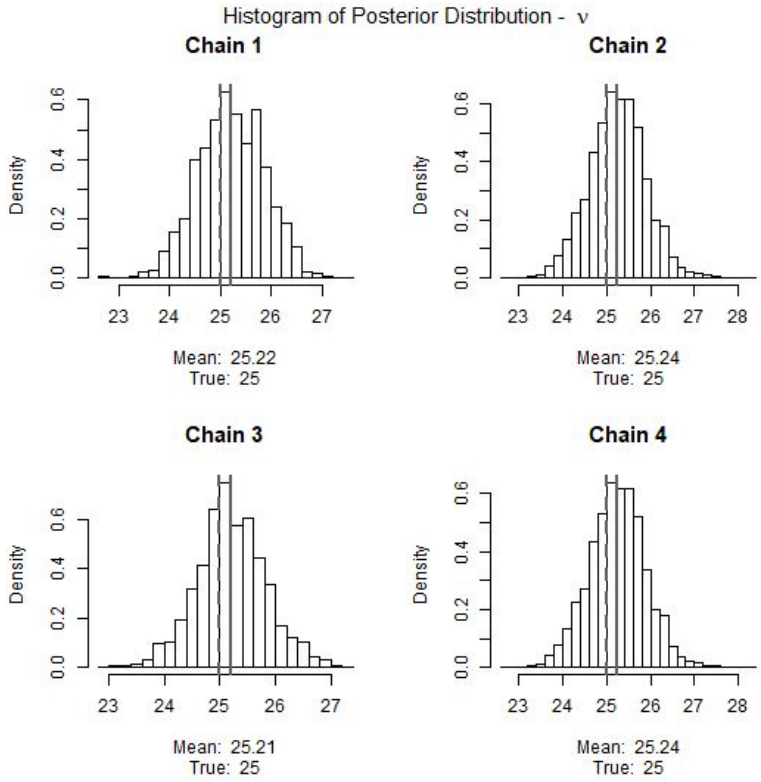


Figure 14 – Multiple histograms for posterior distributions for ν parameter using different start values for one realization of a simulation study with $\nu = 25$. Gray dashed line is the true value and gray solid line is the posterior mean.

Figure 15 and Table 5 show posterior estimates for one realization of a simulation scenario. Posterior estimates for volatility can capture the true values shown in dashed line most of the time. For static parameters the model has difficult in estimate jump magnitude parameters with good precision, since the effective sample for those parameters is limited, *e.g.*, for a simulated scenario with 1,000 observations and jump probability of 5%, the expected number of jump samples is 50 observations. This issue tends to be smaller for higher observation number and higher jump probability parameter, so that the number of jump samples is enough to estimate these parameters with better precision. BIC and AICc are shown on Table 5 so that the model results can be easily compared to other models available on literature, since those criteria are more commonly available on statistical models available on statistical softwares.

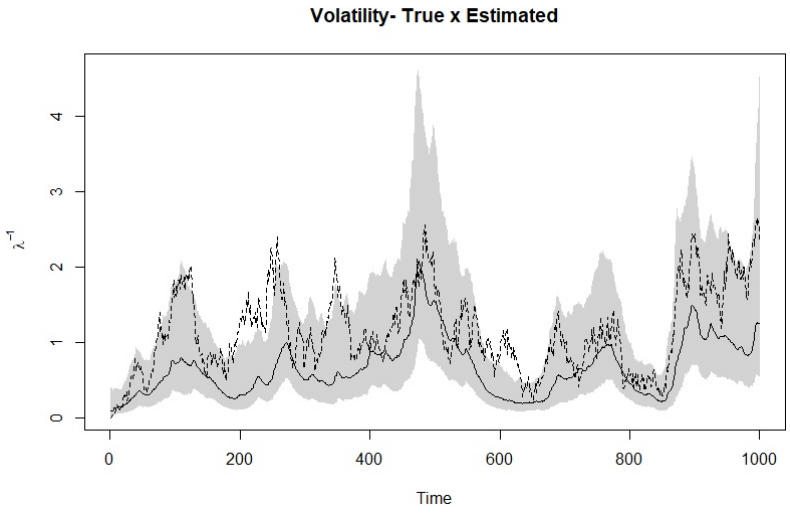


Figure 15 – Simulation study: Posterior mean estimates for instantaneous volatility v_t for simulated data. True volatility series shown in dashed line; posterior mean estimates λ_t^{-1} in solid line; and 95% credibility interval is the light gray area.

	True	Mean	SD
μ	0.05	0.020	0.011
ν	10	9.915	1.439
ρ_y	0.05	0.058	0.013
μ_y	-2.5	-1.086	0.669
σ_y	4.0	4.385	0.542
log L	-	-1308	
BIC	-	2665	
AICc	-	2632	

Table 5 – Posterior estimate of MSVJM static parameters for simulated daily returns (n = 1,000).

3.3 Application to S&P log-returns data set

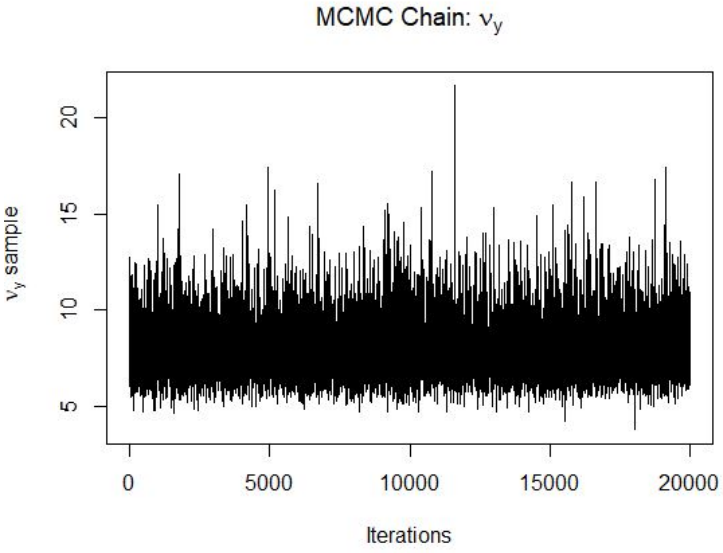
The model is applied on S&P500 daily log-returns data set from January 2, 1980, to December 31, 1999, consisting of 5,054 daily observations, the same data set used in Rego and Santos (2020). The objective here is to see if other parameters estimates will significative change whereas setting the degree of freedom parameter, ν, or estimating it using proposed procedure. The log returns equilibrium parameter μ_t was set static as $\mu_t = \mu$ to keep results comparable to NGSVJ and SV results in Rego and Santos (2020).

Table 6 present model estimates for static parameters in NGSVJ by both fixing ν parameter (NGSVJ) and estimating it using Metropolis steps (NGSVJ - MS) and SV model results shown in Rego and Santos (2020). There was significative change on jump

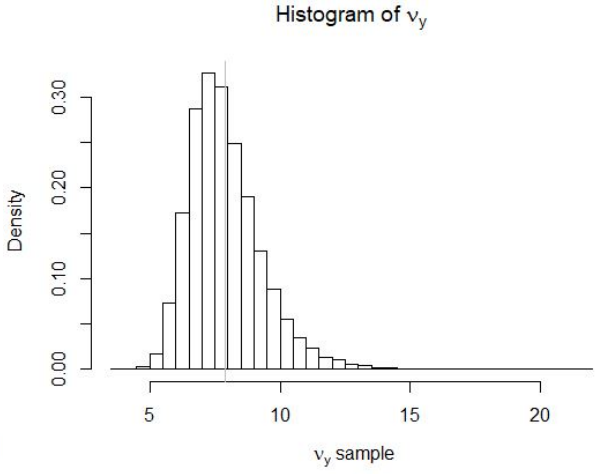
related parameters since a heavier tail caused by a lower value of the degree of freedom ν tends to reduce the effect of jumps on capturing extreme values, since they are already captured by the heavy tail. Figure 16 shows MCMC results for degree of freedom parameter ν, convergence was verified trough graphical methods.

	NGSVJ		NGSVJ - MS		SV	
	Mean	SD	Mean	SD	Mean	SD
μ	0.0616	0.0020	0.0591	0.0038	0	-
ν	30	-	7.8983	1.4080	-	-
ρ_y	0.0042	0.0012	0.0238	0.0061	-	-
μ_y	-2.4598	1.4302	-0.3139	0.3142	-	-
σ_y	5.2793	1.2598	2.5447	0.3136	-	-
log L	-5,972		-5,633		-8,468	
BIC	12,012		11,343		12,206	
AICc	11,960		11,284		-	

Table 6 – Posterior estimates of static parameters for NGSVJ, and NGSVJ with Metropolis step for estimating degrees of freedom, ν, (NGSVJ - MS) for S&P500 daily returns.



(a) MCMC chain for the parameter ν.



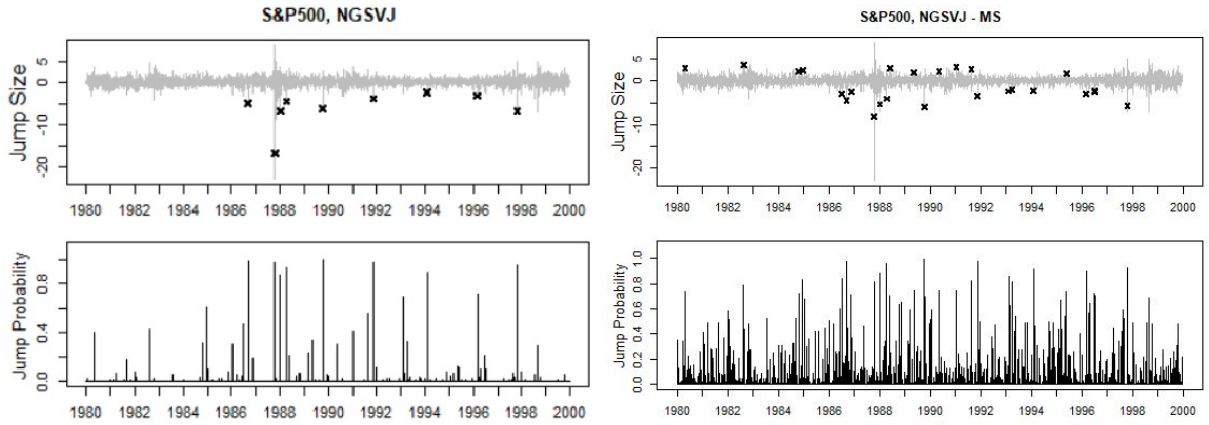
(b) Histogram of the parameter ν.

Figure 16 – MCMC results for degree of freedom parameter ν. Right graph shows MCMC effective chain and left graph shows histogram of MCMC sample. Grey line is the posterior mean.

Figure 17 shows the comparison between estimated jumps on NGSVJ and NGSVJ-MS. By estimating a smaller value of ν, NGSVJ-MS tends to have more frequent jumps, but with smaller magnitude, since part of the excess of returns is captured by the heavy tail. Points identified at jumps on NGSVJ are still present on NGSVJ-MS.

Figure 18 shows the comparison between estimated spot volatility, $\lambda_t^{-1/2}$, on NGSVJ and NGSVJ-MS. A smaller degree of freedom, ν, leads to a smaller volatility estimative since more points are now captured by the heavier tail.

Figure 19 shows the posterior estimates for instantaneous volatility, $\lambda_t^{-1/2}$, for the S&P500 index, using NGSVJ-MS model over the entire period of analysis and two specific periods of time.



(a) NGSVJ jumps. (b) NGSVJ-MS jumps.

Figure 17 – Posterior estimates for instantaneous jumps and jump probabilities for NGSVJ (a) and NGSVJ-MS (b). Grey line on top graphs is the S&P 500 log-returns series, jumps are the black crosses.

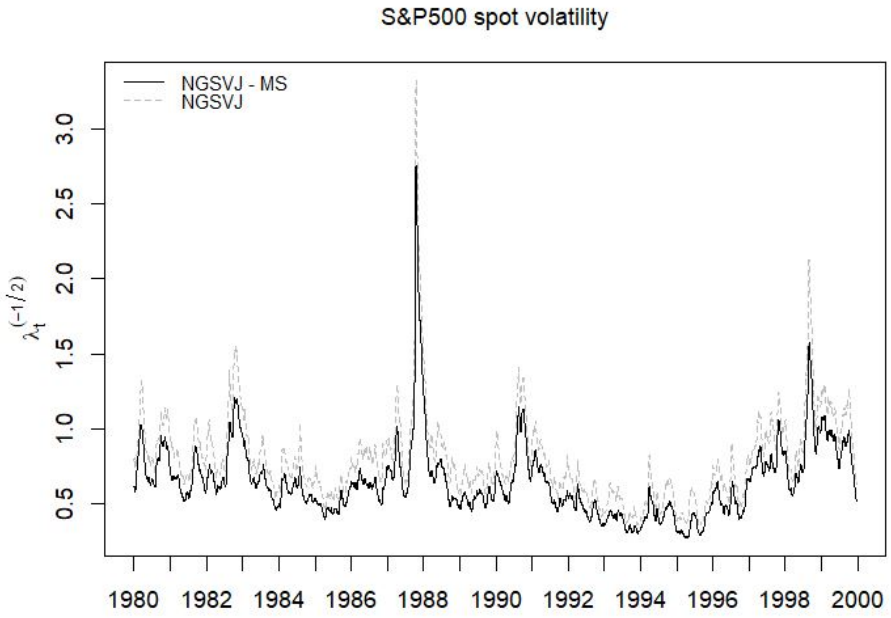


Figure 18 – Posterior estimates for instantaneous volatility $\lambda_t^{-1/2}$. NGSVJ-MS mean estimates in solid line and NGSVJ mean estimates in dashed gray line.

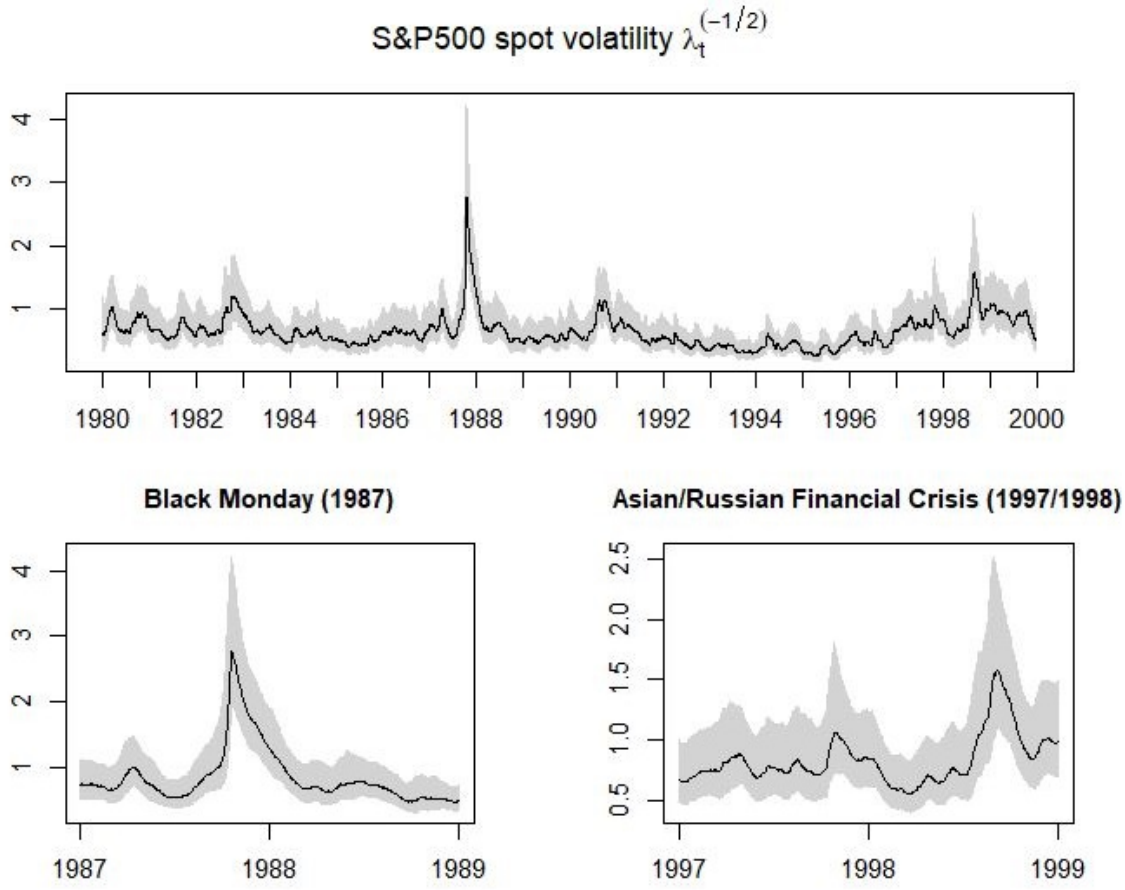


Figure 19 – Posterior estimates for instantaneous volatility, $\lambda_t^{-1/2}$, for the S&P500 index. Solid line is the posterior mean. The gray area indicates the 95% credibility intervals.

Table 7 shows descriptive statistics for the residuals for NGSVJ-MS model. The skewness that exists on S&P500 log-returns is captured by the model trough the jump component so that even though the model does not have a parameter for modeling skewness directly, it can deal with existing asymmetry on data trough the jump component.

	Mean	Standard Dev.	Min	$Q_{25\%}$	$Q_{50\%}$	$Q_{75\%}$	Max
Residuals	0.0044	1.098	-4.5668	-0.7805	-0.0076	0.7961	4.2103

Table 7 – Posterior mean residuals: Summary statistics

Table 8 show results for Jarque-Bera test (Jarque and Bera, 1980), for testing normality of residuals. As p-value is higher than the significance level of 5% for every statistic tested, the normality of residuals can be assumed, as shown on Figure 20 where the histogram of residuals is plotted with the density of a standard Gaussian distribution in solid gray line.

	Statistic	df	p-value	Kurtosis [IC 95%]	Skewness [IC 95%]
JB Test	0.57105	2	0.7516	0.0181 [-0.1168 ; 0.1530]	-0.0385 [-0.1060 ; 0.0290]

Table 8 – Jarque-Bera test results

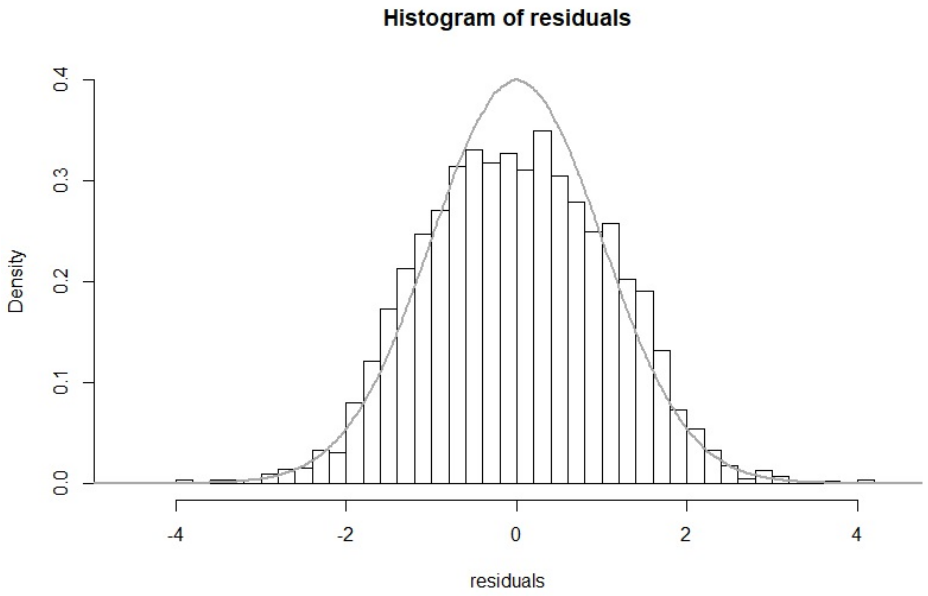


Figure 20 – Histrogram of posterior mean residuals is shown in bars and standard normal curve in gray solid line.

Figure 21 shows ACF and PACF for residuals and squared residuals. There is significant correlation between lags for both cases. This is expected since the model treats the autocorrelation trough the beta evolution for volatility. Table 9 show results for Box-Pierce and Box-Ljung tests for autocorrelation for lags 1, 6, 12, 24 and 36, excluding residuals for first 100 observations. There is no evidence of violation in model hypothesis.

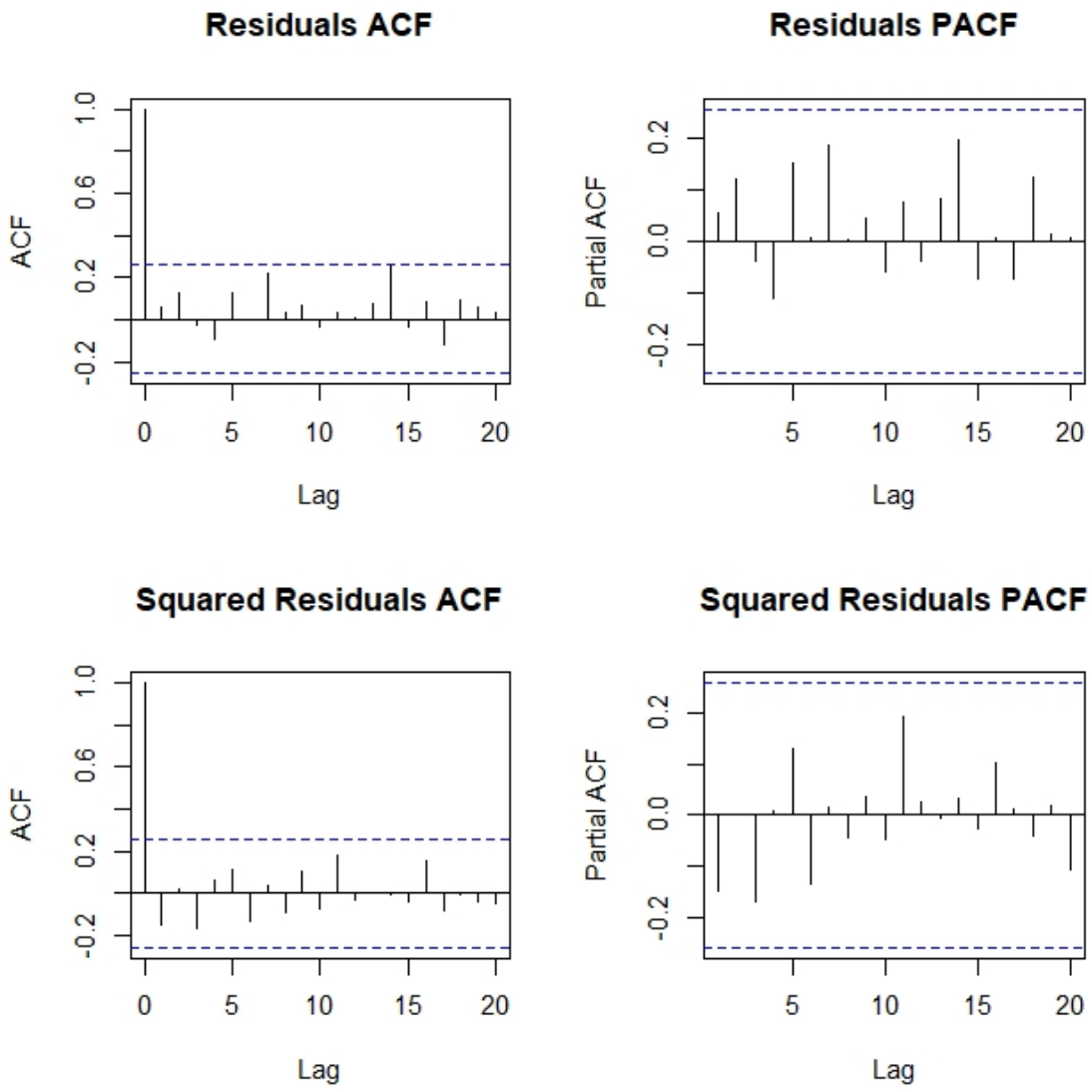


Figure 21 – Top left graph shows ACF for posterior mean residuals; Top right graph shows PACF for posterior mean residuals; Bottom left graph shows ACF for posterior mean squared residuals; Bottom right graph shows ACF for posterior mean squared residuals. Dashed line is the 5% significance interval around zero.

	Residuals			Squared Residuals		
	Statistic	df	p-value	Statistic	df	p-value
Box-Pierce	0.30884	1	0.5784	2.0768	1	0.1496
Box-Ljung	0.3182	1	0.5727	2.1397	1	0.1435
	4.8483	6	0.5634	8.8226	6	0.1838
	11.15	12	0.5161	15.646	12	0.208
	28.834	24	0.2264	21.011	24	0.6381
	35.762	36	0.4798	29.161	36	0.7834

Table 9 – Box-Pierce and Box-Ljung tests results

3.4 Additional Comments

The method proposed was effective in estimating ν by using a Jeffreys prior proposed by Kwow-Wah (2012). Simulation studies show that the proposed prior and the usage of a Metropolis algorithm, since full conditional posterior distribution has no closed-form, was able to estimate the degrees of freedom with good precision, over different simulation scenarios.

The inclusion of a Metropolis step to estimate the gamma mixture degree of freedom ν adds more flexibility to the model and excludes the need for a grid analysis in order to set ν. This extra flexibility comes with a cost since the Metropolis steps require a larger number of iterations to achieve convergence when compared to the Gibbs Sampler. This is a drawback when dealing with situations where a fast response is required from the model, i.e. intra-day operations.

In terms of model applications on financial risk measurement, the NGSVJ-MS is able to better capture the heavy tail effect from data, reducing jumps magnitudes and volatility, while increasing jumps frequency. This effect was already expected, since there is an overlap between the jumps and heavy tails structures, since both are designed to capture tail events.

4 The NGSVJ-HMM Model

The non-Gaussian Stochastic Volatility with Jumps and Hidden Markov Model (NGSVJ-HMM) model is an extension of the NGSVJ model that includes a Hidden Markov Model structure to classify different levels of volatility. This extension was originally developed to be applied to financial markets, supporting strategic decisions. Nevertheless, it can be used in any other applications where classifying different states of volatility is crucial for decision making.

The Hidden Markov Models (HMM) are widely used for classification, also referred to as decoding, in many areas. Rabiner (1989) gives an overview of HMM functioning, as well as a detailed tutorial on its implementation, and applies it to a speech recognition problem, where the sound waves that come from a speech must be classified as letters to form a text. Diez, Burget and Matejka (2018) uses a Bayesian HMM to make a classification in a speaker diarization context, on identifying who spoken and when in an audio recording of a conversation. Ji (2018), Sebastian *et. al* (2018) and Pohle *et. al* (2017) also use this approach, in a biological application context.

Further developments on the HMM were done by Benyacoub *et. al.* (2014) on expanding the usage of HMM to problems where supervised learning is not available directly. Their approach consists of evaluating the probability to belonging in one group, given the observations by a linear classifier, to estimate HMM parameters. Liu (2019) develops a computational method for a CUDA implementation of HMM training and classification, for increasing its computational efficiency.

The traditional approach to Hidden Markov Models relies on Expectation Optimization algorithms to estimate HMM parameters. Xu and Jordan (1996) defend the usage of EM algorithms for HMM problems, despite the reputation of being a slow algorithm, specially when mixture components are not well separated. On the Bayesian perspective, the work of Meeden and Vardeman (2000) propose a method for a Bayesian approach to the HMM modeling with time dependent data, allowing the incorporation of different types of prior information and discussing the implementation of a Bayesian HMM.

Nkemnole and Wulu (2017) stated that HMMs have been applied for at least three decades in signal-processing applications, especially in automatic speech recognition, and now this theory and application has expanded to other fields, such as finance. According to Smith (2002), Markov-switching models have proven to be very useful in modeling a range of economic time series including business cycles, stock market, exchange rates and short-term interest rates. Smith (2002) develops the Markov-switching stochastic volatility model, a generalization of the SV model to estimate the elasticity of variance and evaluate

whether the interest rate process in explosive.

Hassan and Nath (2005) propose the usage of a HMM to forecast airline stock prices, by modeling behavioral data patterns from the past dataset of prices data, comparing its results to those obtained by applying a Artificial Neural Network for prices forecasting, concluding that the HMM has the advantage of being an explainable model with solid statistical foundation. In a later work, Hassan, Nath and Kirley (2007) propose a model for stock market forecasting by combining HMM, Artificial Neural Networks and Genetic Algorithms, for using on deep analysis of the stock market, so that the advantages of HMM can be combined to the Artificial Neural Network model, to produce a model with good results for forecasting and interpretability. Their work gives insights on the versatility of Hidden Markov Models on being able to be included inside other models, giving interpretability to results.

Lopes and Carvalho (2007) develop a Factor Stochastic Volatility model using Markov Switching regimes to allow Markovian jumps in the levels of log-volatilities. These jumps allow for fast track of possible abrupt changes in the variance and covariance structures while addressing the issue of high persistency commonly present in financial data. Sun (2018) uses a HMM approach to implement a trading strategy, by determining the current market price level for stock and suggesting to sell when stock is in high trend level and buy when it is in low trend. Nkemnole and Wulu (2017) use an HMM-SV approach, in two stages, using a HMM to divide the entire time series into regimes with different volatility levels, so that the return of the time series is assumed to be modeled by a mixture of probability densities and each density function corresponds to a hidden state with its mean and variance. Subsequently the subsets of original time series corresponding to different states (volatility levels) are modeled using a SV model with different parameter sets to model the conditional variance.

The aforementioned works give insights on how to include a Markovian structure inside the NGSVJ to classify volatility states for enhancing model interpretability and usage on real applications on financial markets. HMM methods have strong appeal on finance applications specially on giving interpretability for statistical models. The inclusion of a HMM structure inside the NGSVJ is not straightforward and a well documented method for addressing this question, into a Bayesian perspective and computationally effective way, strengthens the DM class of models and makes the NGSVJ more competitive in relation to SV models, since it shows that the model is versatile enough to include features with great appeal on SV class of models, that were not yet developed on the NGSVJ perspective.

In this chapter, we present the Non-Gaussian Stochastic Volatility with Jumps and Hidden Markov Model to estimate market volatility and provide an interpretable result to the final user by classifying the volatility in a high or low state, developing the

required theoretical background needed for the inclusion of the structure on the NGSVJ model. This information can be used in a market operation context, to take investment or arbitrage strategic decisions, in an intraday or daily time frame. In this scenario, the return data are high dimensional. When classified as high volatility state, it means that the change in log-returns is higher for the asset's standards, which means there is an arbitrage window for speculative traders. Note that the model is not able to determine if the price will rise or fall, only that the price is changing faster than usual, in either direction. We apply the model on intraday log returns from Brent Crude Oil future contracts, to show its application.

The main objective of this chapter is to introduce the NGSVJ-HMM which present to the final user an interpretable and intuitive result for market volatility to use this information on financial decisions. This model includes all characteristics from NGSVJ: computational simplicity, flexibility, the inclusion of a heavy-tailed distribution and jumps for accommodating financial returns data, together with the HMM structure to classify the volatility state based on both log-returns and estimated volatility.

The proposal is to combine these two models, finding an optimal structure for the HMM classifier to be able to effectively classify the observations that belong to each state inside the NGSVJ structure, given that the estimated volatility is subject to variations that can affect on classification.

Our first attempt of implementing NGSVJ-HMM used only estimated log-volatility as information for the classifier to separate observations on two groups: high and low volatility state. Therefore, given the shape of density for the log-volatility, the model was only able to effective classify states when using the EM algorithm proposed by Xu and Jordan (1996), using two stages: first to estimate volatility, second to classify states. Although, the two stage strategy is not desirable.

By including log-returns as a piece of additional data for the classifier, and the usage of a bivariate density for volatility and log-returns into a Bayesian mixture approach, we were able to effectively classify states in a one-stage model, so that the HMM structure is included inside the MCMC algorithm for the NGSVJ. This approach leads to a more elegant mathematical solution for the classification structure and a more computational efficient approach, which we will discuss in the next pages with more details.

4.1 Model Structure

The NGSVJ-HMM model is an enhancement of the NGSVJ model to include a Markov classifier for market volatility states. The advantage of this structure is keeping an automatic sampling procedure, mostly using Gibbs steps, since most conditional posterior distributions have closed-form and are known.

The NGSVJ-HMM for log-return time series $\{y_t\}_{t=1}^n$ is given by:

$$
\begin{aligned}
y_t &= \mu_t + J_t^y + \upsilon_t,\ \upsilon_t|\gamma_t \sim N(0, \gamma_t^{-1}\lambda_t^{-1}), && (4.1)\\
\lambda_t &= \beta^{-1}\lambda_{t-1}\zeta_t,\ \zeta_t|\mathcal{D}_{t-1}, \varphi \sim Beta(\beta a_{t-1}, (1-\beta)a_{t-1}), && (4.2)\\
\gamma_t &= G(\frac{\nu}{2}, \frac{\nu}{2}), && (4.3)\\
\kappa &= (A, B, \underset{\sim}{\pi}). && (4.4)
\end{aligned}
$$

where

$$
\begin{aligned}
&J_t^y = \xi_{t+1}^y N_{t+1}^y, \quad \xi^y \sim N(\mu_y, \sigma_y^2), \quad and \quad Pr(N_{t+1}^y = 1) = \rho_y.\\
&A = \{a_{i,j}^* \sim Dir(\underset{\sim}{\delta_j})\}, \quad B = \{b_j(y_t, log\lambda_t|\gamma_t, s_t = j)) \approx N_2(\mu_{D_j}, \Sigma_{D_j})\},\\
&and \quad \underset{\sim}{\pi} \sim Dir(\underset{\sim}{\delta_j^*}).
\end{aligned}
$$

In this model y_t follows the DLM defined by $\{1, v_t, \lambda_t, \gamma_t\}$. y_t represents the log-return in percentage, defined as $y_t = 100 \times (log(P_t) - log(P_{t-1}))$, where P_t is the asset price on time t. J_t^y is the jump, composed by the jump indicator $N_{t+1}^y \in \{0, 1\}$ and magnitude $\xi^y \sim N(\mu_y, \sigma_y^2)$. μ_t represents the equilibrium log-return of y_t on time t. γ_t is the variance mixture component and λ_t^{-1} is the volatility of returns and it's has a $G(\frac{\nu}{2}, \frac{\nu}{2})$ distribution in order to achieve a Student-t distribution with ν degrees of freedom for innovations, initialized as $\lambda_0|\mathcal{D}_0 \sim G(a_0, b_0)$, where $\mathcal{D}_0$ is the set of initial information. $\varphi = \{\beta, \nu, \mu, \kappa\}$ is the vector of static parameters, where β controls information loss over time and is specified, and a_{t-1} is the shape parameter of the filtering distribution of λ_t. Both y_t and λ_t are subject to a state classification s_t.

The HMM is described by the static parameters $\kappa = (A, B, \pi)$. We define $N = 2$ the number of states so that $s_t \in \{1, 2\}$, namely $s_t \in \{low, high\}$, and our observation vector is the pair log-returns, y_t, and estimated volatility, the log posterior volatility measure, $log\lambda_t$. $A = \{a_{i,j}^*\}$ is the state transition matrix, so that for each row i, $a_{i,j}$ indicates the probability that the next state becomes the state i, for $j = 1, 2$, $B = \{b_j(y_t, log\lambda_t|\gamma_t, s_t = j))\}$ is the observation probability density distribution in state j for the position of centroid, that is given by a multiplication of a conditional Gaussian and Log-Gamma distribution, that will be approximated by a $p_t = N_2(\mu_{D_j}, \Sigma_{D_j})$, and $\underset{\sim}{\pi} = \{\pi_i\}$, with a $Dir(\underset{\sim}{\delta_j^*})$ prior, the initial state probability. Dirichlet prior was chosen so that the model can be expanded for any desired number of states that may be needed for user application,

e.g. $s_t \in \{low, medium, high\}$. Our goal is to sample from the posterior distribution of $p(\kappa|\mathcal{D}_n)$.

The probability distribution $b_j(y_t, log\lambda_t|\gamma_t, s_t = j))$ is approximated by a Bivariate Gaussian to estimate the position and dispersion of the centroids in that distribution, since it is more convenient to the classification approach under the HMM structure, as in Rabiner (1989) and Marin *et. al.* (2005). Also, as $log\lambda_t$ is distributed as a Log-Gamma, it can be approximated by a Gaussian distribution for usual values of shape and scale parameter, so that there is no loss in generality by the model by approximating b_j as a bivariate Gaussian for estimating centroid position and range of influence for the HMM classifying structure. From now on $\gamma_t, s_t = j$ is omitted from b_j to simplify notation.

4.2 Bayesian Inference

For ease notation, let $\Phi = (\underset{\sim}{\mu}, \underset{\sim}{\gamma}, \underset{\sim}{\lambda}, \underset{\sim}{\mu_y}, \underset{\sim}{\sigma_y^2}, \underset{\sim}{\xi}, \underset{\sim}{\rho}, \underset{\sim}{N^y}, \kappa, \underset{\sim}{s}, \underset{\sim}{\mu_{D_j}}, \underset{\sim}{\Sigma_{D_j}}, \nu)$, excluding the parameter being evaluated, i.e. $\Phi_{[-\lambda]} = (\underset{\sim}{\mu}, \underset{\sim}{\gamma}, \underset{\sim}{\mu_y}, \underset{\sim}{\sigma_y^2}, \underset{\sim}{\xi}, \underset{\sim}{\rho}, \underset{\sim}{N^y}, \kappa, \underset{\sim}{s}, \underset{\sim}{\Sigma_{D_j}}, \nu)$. Given proper priors for the parameters, a sample from the posterior of Φ after observing the series up to time n is drawn. Independent priors were chosen so that proper posteriors are obtained, simplifying the sample procedure. When no initial information is available, non-informative priori are chosen. The joint posterior is:

$$
\begin{aligned}
p(\Phi|\mathcal{D}_n) &= p(\Phi_{[-(\kappa,\underset{\sim}{s})]}|\kappa, \underset{\sim}{s}, \mathcal{D}_n) && (4.5)\\
&= p(\Phi_{[-(\kappa,\underset{\sim}{s})]}|\mathcal{D}_n) \times p(\kappa, \underset{\sim}{s}|\Phi_{[-(\kappa,\underset{\sim}{s})]}, \mathcal{D}_n) && (4.6)\\
&= p(\Phi_{[-(\kappa,\underset{\sim}{s})]}|\mathcal{D}_n) \times p(\underset{\sim}{s}|\kappa, \Phi_{[-\underset{\sim}{s}]}, \mathcal{D}_n) \times p(\kappa|\Phi_{[-\underset{\sim}{s}]}, \mathcal{D}_n) && (4.7)
\end{aligned}
$$

For NGSVJ related parameters,(μ_t, λ_t, γ_t, μ_y, σ_y^2, ξ_{t+1}^y, N_{t+1}^y, ρ, ν), prior and posterior distributions are exactly the same of that in Section 2.2, the reason why they are here omitted. Note that, as can be seen on Eq. (4.5), none of the NGSVJ related parameters are influenced by the HMM related parameters κ and $\underset{\sim}{s}$, so that the full conditional posterior distribution stays unchanged, but HMM parameters are affected by the parameters of NGSVJ.

For state transition probabilities, $A = \{a_{i,j}^*\}$, a Dirichlet$(\delta_1, \ldots, \delta_j)$ prior is specified for each row a_i, for $i = 1, 2$ and $j = 1, 2$, resulting in a full conditional posterior distribution:

$$
\begin{aligned}
a_i^*|\underset{\sim}{y}, \underset{\sim}{\lambda}, s_t, \Phi_{[-a_i]} \sim \quad Dir \Bigg(& \delta_1 + \sum_{t=1}^{n} \mathbb{1}[s_t = 1, s_{t-1} = i], \\
& \ldots, \delta_j + \sum_{t=1}^{n} \mathbb{1}[s_t = j, s_{t-1} = i] \Bigg).
\end{aligned} \tag{4.8}
$$

where $\mathbb{1}[\cdot]$ stands for the indicator function.

For the initial state probabilities, $\pi = \{\pi_1, \pi_2\}$, a Dirichlet(δ_1^*, δ_2^*) prior is specified, resulting in a full conditional posterior distribution:

$$
\pi|\underset{\sim}{y}, \underset{\sim}{\lambda}, s_t, \Phi_{[-\pi]} \sim \quad Dir \left(\delta_1^* + \sum_{t=1}^{n} \mathbb{1}[s_t = 1], \delta_2^* + \sum_{t=1}^{n} \mathbb{1}[s_t = 2] \right). \tag{4.9}
$$

For estimating the probability distribution $b_j(\underset{\sim}{y}, log\underset{\sim}{\lambda})$ parameters mean μ_{D_j} and covariance matrix Σ_{D_j}, a prior $N_2(m_0, \Sigma_0)$ and $IW(\eta_0, \Psi_0)$ were specified respectively, resulting in a full conditional posterior distribution:

$$
\mu_{D_j}|\underset{\sim}{y}, \underset{\sim}{\lambda}, s_t, \Sigma_{D_j} = N_2 \left(V_j \left(\Sigma_0^{-1} m_0 + \Sigma_{D_j}^{-1} S_{x_j} \right), V_j \right) \tag{4.10}
$$

$$\Sigma_{D_j}|\underset{\sim}{y}, \underset{\sim}{\lambda}, s_t, \mu_{D_j} \;=\; IW\left(n_j + \eta_0, \Psi_0 + S_{x_j^2}\right) \tag{4.11}$$

where $V_j = \left(\Sigma_0^{-1} + n_j\Sigma_{D_j}^{-1}\right)$, $j = 1, 2$ refers to the states (low, high), x stands for the pair $(y, log\lambda)$, $n_j = \sum_{t=1}^{n} \mathbb{1}[s_t = j]$ is the number of times a state was classified as 1 (low) or 2 (high), $S_{x_j} = \sum_{t=1}^{N} x\mathbb{1}[s_t = j]$ is the sum of x for state j, and $S_{x_j^2} = \sum_{t=1}^{n}(x - \mu_{D_j})(x - \mu_{D_j})^T\mathbb{1}[s_t = j]$.

A sample for the states s_t given all information available and κ, can be obtained by the Viterbi algorithm, described in Rabiner (1989). According to Rabiner (1989), the Viterbi procedure has the advantage of avoiding transitions where state transition probability is zero so that it guarantees a valid state sequence. The procedure consists in finding the single best state sequence, $Q = (q_1, q_2, \ldots, q_n)$, for the given observation sequence $(y, log\lambda) = ((y_1, log\lambda_1), (y_2, log\lambda_2), \ldots, (y_n, log\lambda_n))$, where $\delta_t(i)$ is the best score (higher probability) along a single path, at time t, which accounts for the first t observations and ends in state s_i. To retrieve the state sequence, we keep track of which argument maximized $\delta_{t+1}(j)$ for each time t and state j, via the array $\Psi_t(j)$, and do a path backtracking to determine the best path. The Viterbi algorithm steps are:

1. Initialization:

$$\delta_1(i) \;=\; \pi_i b_i((y, log\lambda)_1), \quad 1 \leq i \leq N \tag{4.12}$$

$$\Psi_1(i) \;=\; 0. \tag{4.13}$$

2. Recursion:

$$\delta_t(j) \;=\; \max_{1\leq i\leq N}\left[\delta_{t-1}(i)a_{ij}^*\right] b_i((y, log\lambda)_t),\ 2 \leq t \leq n,\ 1 \leq j \leq N \tag{4.14}$$

$$\Psi_t(j) \;=\; \max_{1\leq i\leq N}\left[\delta_{t-1}(i)a_{ij}^*\right], \quad 2 \leq t \leq n, \quad 1 \leq j \leq N. \tag{4.15}$$

3. Termination:

$$P^* \;=\; \max_{1\leq i\leq N}\left[\delta_T(i)\right] \tag{4.16}$$

$$q_T^* \;=\; \max_{1\leq i\leq N}\left[\delta_T(i)\right]. \tag{4.17}$$

4. Path (state sequence) backtracking:

$$q_t^* \;=\; \Psi_{t+1}(q_{t+1}^*), \quad t = n-1, n-2, \ldots, 1. \tag{4.18}$$

4.3 General Procedure

Let be $Y_n = \{y_t\}_{t=1}^n$, the NGSVJ parameters $\underset{\sim}{\mu}$ $\underset{\sim}{J} = \{J_t^y\}_{t=1}^n = \{\xi_{t+1}^y N_{t+1}^y\}_{t=1}^n$, $\underset{\sim}{\gamma} = \{\gamma_t\}_{t=1}^n$, $\underset{\sim}{\lambda} = \{\lambda_t\}_{t=1}^n$, $\underset{\sim}{\xi} = \{\xi_{t+1}^y\}_{t=1}^n$, $\underset{\sim}{N} = \{N_{t+1}^y\}_{t=1}^n$, and the HMM parameters $\underset{\sim}{s} = \{s_t\}_{t=1}^n$, $A = \{a_{i,j}\}$, $\underset{\sim}{\pi} = \{\pi_i\}$, $\underset{\sim}{\mu}_{D_i}$, Σ_{D_i}, for $i, j = 1, 2$ and the prior probability density $p(\underset{\sim}{\gamma})$, $p(\nu)$, $p(\mu_y)$, $p(\sigma_y^2)$, $p(\underset{\sim}{\xi})$, $p(\rho_y)$, $p(a_{i,j})$, $p(\pi_i)$, $p(\underset{\sim}{\mu}_{D_i})$, $p(\Sigma_{D_i})$ are set for $\underset{\sim}{\gamma}, \nu, \mu_y, \sigma_y^2, \underset{\sim}{\xi}, \rho_y, a_{ij}^*, \pi_i$, $\underset{\sim}{\mu}_{D_i}$, Σ_{D_i}, for $i, j = 1, 2$, respectively. Then, a sample of size M from the posterior distribution of the parameters is drawn from algorithm is described as follows:

1. Initialize $\underset{\sim}{\mu}^{(0)}, \underset{\sim}{\lambda}^{(0)}, \underset{\sim}{\gamma}^{(0)}, \mu_y^{(0)}, (\sigma_y^2)^{(0)}, \underset{\sim}{\xi}^{(0)}, \underset{\sim}{N}^{(0)}, \rho_y^{(0)}$ and $\nu^{(0)}$.
2. Initialize $\underset{\sim}{s}^{(0)}, A^{(0)}, \underset{\sim}{\pi}^{(0)}, \mu_{D_i}^{(0)}, \Sigma_{D_i}^{(0)}$, for $i = 1, 2$.
3. Set $j = 1$.
4. Sample from $\underset{\sim}{\mu}^{(j)}, \underset{\sim}{\lambda}^{(j)}, \underset{\sim}{\gamma}^{(j)}, \mu_y^{(j)}, (\sigma_y^2)^{(j)}, \underset{\sim}{\xi}^{(j)}, \underset{\sim}{N}^{(j)}$ and $\rho_y^{(j)}$, as in Section 2.3.
5. Sample $\nu^{(j)}$ from Eq. (3.2), using a Metropolis step.
6. Sample $A^{(j)}$ as in Eq. (4.8).
7. Sample $\underset{\sim}{\pi}^{(j)}$ as in Eq. (4.9).
8. Sample $\mu_{D_i}^{(j)}$, for $i = 1, 2$ as in Eq. (4.10).
9. Sample $\Sigma_{D_i}^{(j)}$, for $i = 1, 2$ as in Eq. (4.11).
10. Block sample $\underset{\sim}{s}^{(j)}$ using the Viterbi algorithm in Rabiner (1989).
11. Set $j = j + 1$.
12. If $j \leq M$, go to step 4, otherwise stop.
13. Reorder states so that $s = 1$ corresponds to lowest and $s = 2$ corresponds to highest values of volatility.

Since most full conditional posterior distribution has closed-form, there is no need to appeal to Metropolis algorithms, except for the degree of freedom parameter ν. The increase of dimensionality by using the pair $(y, log\lambda)$ for classification instead of using only the volatility as observation for the HMM classifier assures that no identifiability issues will occur when estimating parameters for B. As in NGSVJ, static such as β are set.

4.4 Simulation Study

The procedure in Rego and Santos (2020) was adapted to build simulation scenarios for a market daytrade operation with two volatility states. A Monte-Carlo study with 1,000 replications is made to test the efficiency of the proposed procedure to estimate the market states using the HMM structure.

To generate the volatility, we use:

$$v_t = v_{t-1} + \kappa(\theta - v_{t-1})\Delta + \rho\sigma_v\sqrt{v_{t-1}\Delta}\epsilon_{1,t} + \sigma_v\sqrt{(1-\rho^2)v_{t-1}\Delta}\epsilon_{2,t} \tag{4.19}$$

where $\epsilon_{1,t}$ and $\epsilon_{2,t} \sim N(0,1)$. A transition matrix A was built so that for each state $s_t \in \{1,2\}$ there is a probability of moving to the next state or remaining in the current state. If for an instant t votatility state is High, $s_t = 2$, then a fixed multiplier, χ, is applied to v_t to increase volatility in this specific instant, highlighting the increased volatility state.

Synthetic data for returns is then generated from:

$$r_t = N(\mu_t + \mathcal{J}_t, \gamma_t^{-1}(\mathbb{1}[s_t = 1]v_t) + \mathbb{1}[s_t = 2]\chi v_t), \qquad \mathcal{J}_t = N_t\xi_t, \tag{4.20}$$

where the jump times, N_t are generated from a Bernoulli(ρ_y), jump sizes ξ_t from $N(\mu_y, \sigma_y^2)$, and γ_t from $G(\frac{\nu}{2}, \frac{\nu}{2})$, with $\nu = 15$. Setup of parameters was: log-returns mean $\mu_t = \mu = 0.05$; jump probability $\rho_y = 0.015$; jump magnitude mean $\mu_y = -2.5$ and standard deviation $\sigma_y = 4$; volatility components $\Delta = 1$, $\theta = 0.8$, $\kappa = 0.015$, $\sigma_v = 0.1$, $\rho = 0.4$, $n = 1000$. For transition matrix $a_{1,1} = 0.99$, $a_{1,2} = 0.01$, $a_{2,1} = 0.02$ and $a_{2,2} = 0.98$. The multiplier was fixed in $\chi = 2$.

Tables 10 to 22 show model estimatives for simulated parameters. Estimative for jump magnitude mean, μ_y, is influenced by the increase in volatility caused by the market state change in the simulation scenarios, since with a higher estimative for instantaneous volatility the model capability to recognize jumps is reduced. Other static parameters are well estimated by the model. Another point of attention is on estimative for instantaneous state probability π, since this value is not set in the simulation procedure, and true value is obtained by the reason of the number of observations in each state by the total number of observations. As some points can still remain in a lower volatility region even when multiplied by the constant χ, it is expected that estimates for π are different from true values, which does not influence the final analysis.

	True	Mean	SD		True	Mean	SD
μ	0.05	0.0516	0.0077	μ_{D_1}	-	(0.0786, -0.1029)	(0.0715, 0.1701)
ρ_y	0.015	0.0154	0.0075	μ_{D_2}	-	(0.0215, 1.0388)	(0.0345, 0.1505)
μ_y	-2.5	-0.2858	1.6032	π_1	0.777	0.4491	0.0582
σ_y	-4	4.1659	1.5123	π_2	0.223	0.5509	0.0582
ν	15	16.1564	1.277	BIC		2553	
log L		-1245		AICc		2509	

Table 10 – Posterior inference of static parameters for simulated data.

	Σ_{D_1}			Σ_{D_2}	
	1	2		1	2
1	2.0691 (0.2077)	-0.0447 (0.0905)	1	0.4876 (0.0529)	0.0034 (0.0253)
2	-0.0447 (0.0905)	0.3571 (0.0725)	2	0.0034(0.0253)	0.4681 (0.0877)

Table 11 – Posterior means for HMM classifier covariance matrix Σ_{D_i}. Standard deviations are in parenthesis.

	Low		High	
	True	Estimated	True	Estimated
Low	0.99	0.9837 (0.0067)	0.01	0.0163 (0.0067)
High	0.02	0.0149 (0.006)	0.98	0.9851 (0.006)

Table 12 – Posterior transition probability, $a_{i,j}$, means for simulated data, log returns in an intuitive approach. Standard deviations are in parenthesis.

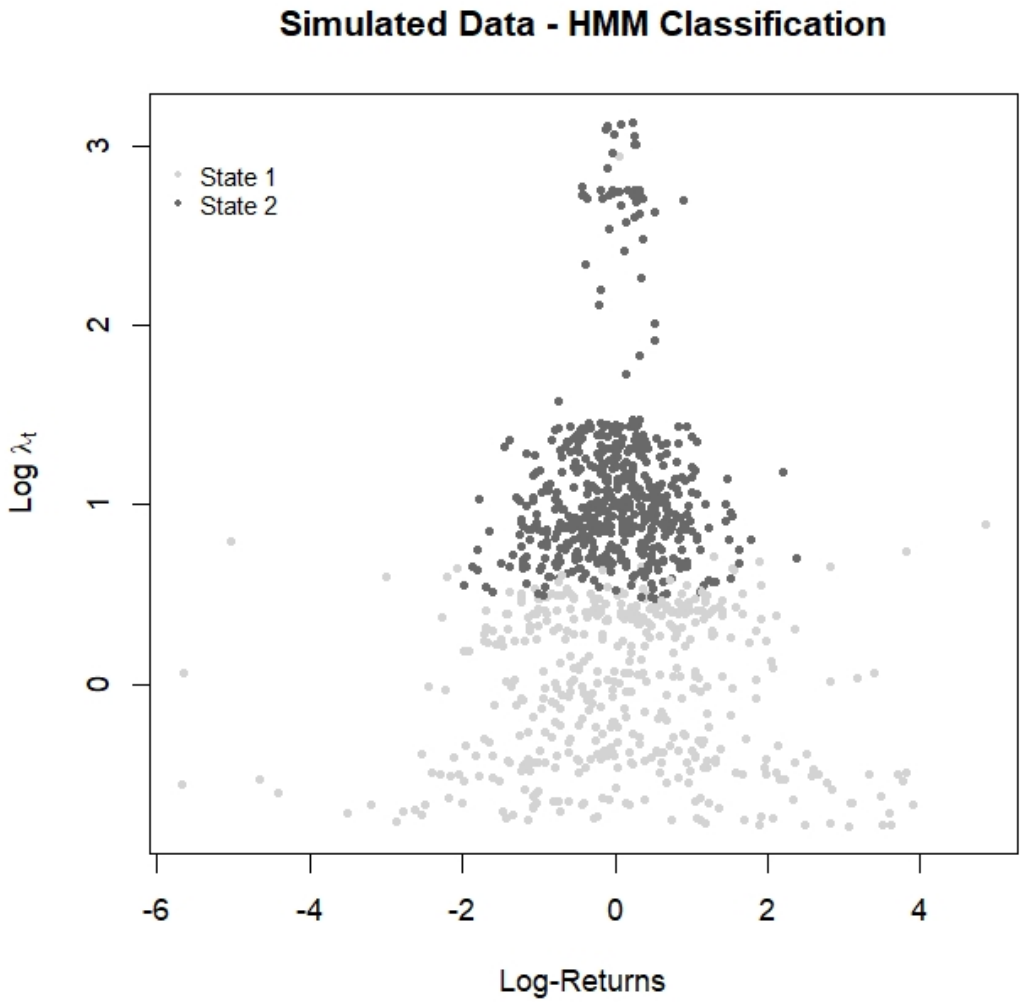


Figure 22 – HMM classification for market states according to values of log-return, y_t, and mean posterior log volatility, $log\lambda$.

Figure 22 show model classification for one simulated scenario. There is a well defined cut-point between states in the region around $log\lambda_t = 0.6$, which corresponds to a spot volatility of approximately $\lambda_t^{-1/2} \approx 0.74$, as can be seen on Figure 23, where volatility posterior estimative (solid line) is plotted together with true value (dashed line) and market states are shown on bottom graph. In moments with increased volatility, as around observation number 800, the model capability of detecting of jumps is reduced due to the higher volatility behavior, which caused the estimative for μ_y to be underestimated.

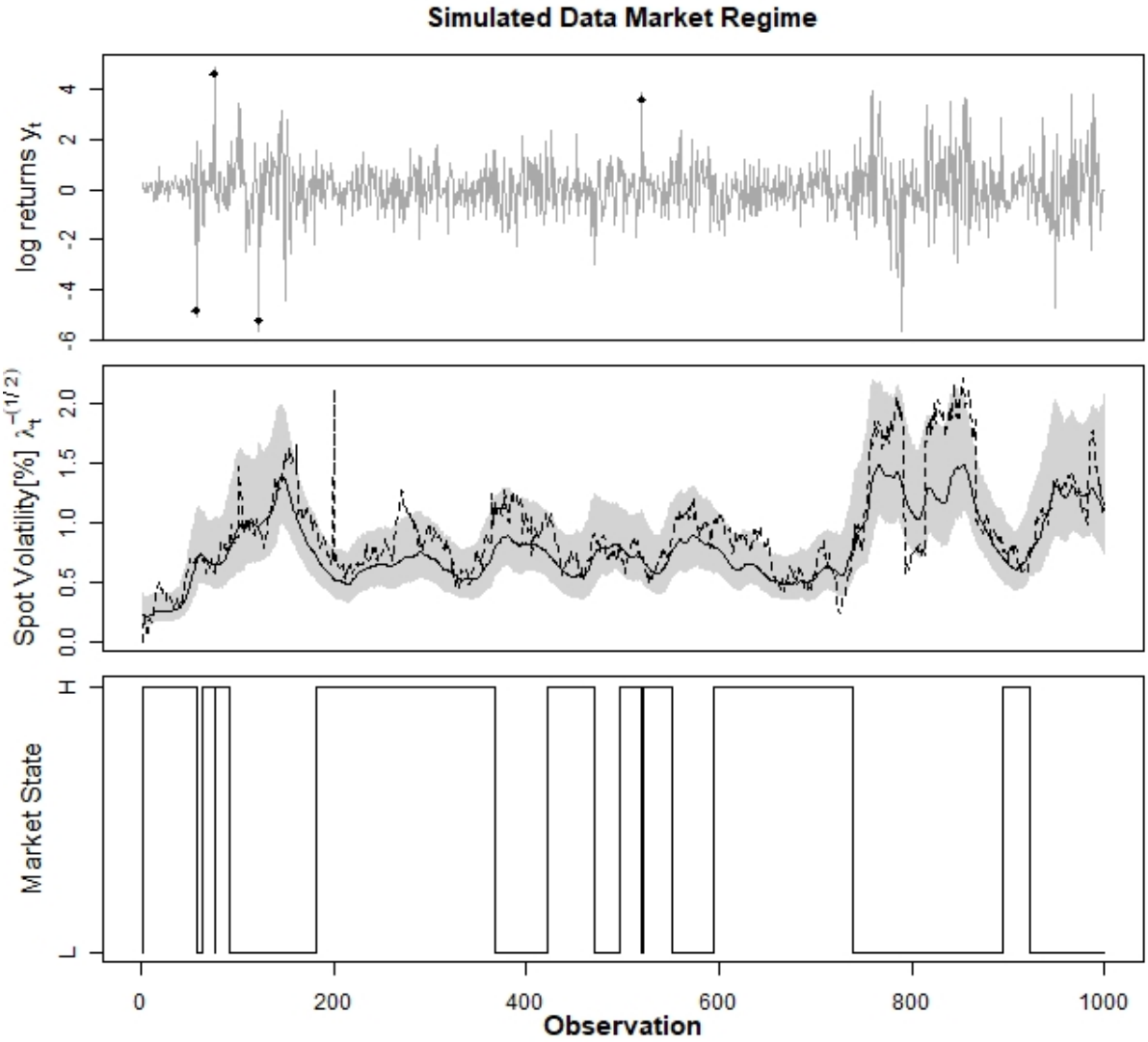


Figure 23 – Simulated log returns data are shown at the top graph, together with jumps as dots; Mean posterior estimates of spot volatility in percentage $\lambda_t^{-1/2} \times 100\%$ at the middle graph in solid line, the gray area is the 95% credibility interval, the true value is the dashed line; and estimates of instantaneous Market Regime at the bottom graph.

Figure 24 show posterior estimate for jump probabilities in solid line together with true values in dashed line. The model is able to capture most regions of high and low volatility correctly, with a small delay, which is acceptable for the purposes here presented.

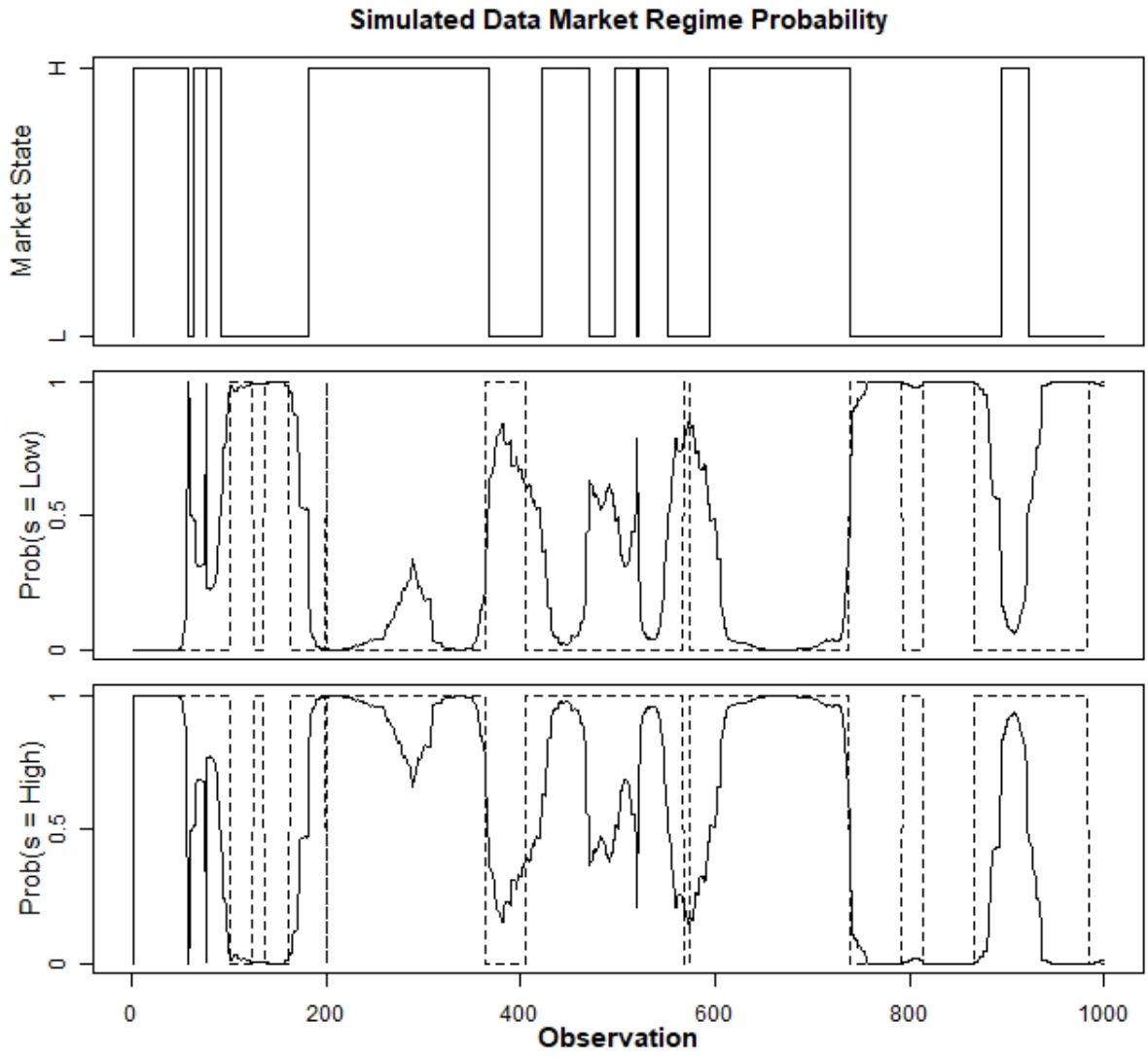


Figure 24 – Market Regime probability for each observation of simulated log returns data. Dashed line is the true value of regime.

4.5 Intraday Log-Returns of Brent Crude Oil Futures

In this section we show an application of the model to Brent Crude futures, ICE:BRN, intraday log-returns time series. The particular interest in this application is to identify moments of increased volatility so that a speculator can use trading strategies to obtain profit from arbitrage. The data set consists of Brent Crude futures, ICE:BRN, log

returns from August 15, 03:00, 2018 to August 16, 23:59, 2018, in a total of 2,472-minute observations.

Parameter specification for the NGSVJ-HMM model follows specifications used and presented in previous sections. For state transition probabilities, $a_{i,j}$, and for initial state probabilities π_i, $i = 1, 2$, a Dirichlet(0.1,0.1) is defined, so that there are only two states, namely low and high volatility states. For this application, the log-returns mean, μ_t will be assumed as a static parameter, so that $\mu_t = \mu$. MCMC specification is 50,000 iteration chain, a burn-in of 20,000 observations, with a lag of 10 observations, resulting in 2,728 samples.

Convergence of MCMC chains was verified trough graphical methods. Residual analysis was made in the same way as in Section 3.3, with no evidence of strong violations of the model hypotheses.

Table 13 and Table 14 shows posterior inference of static parameters. Convergence was achieved after approximately 10,000 iterations. As expected based on previous applications, the jump probability is small, $\rho_y = 3.19\%$, and the presence of heavy-tail is observed by the estimated degree of freedom, $\nu = 8.6094$. Physically, the estimates of μ_{D_1} and μ_{D_2} represent the positions of the centroids used by the HMM classifier. As can be seen in Table 15, there is a high probability that a low(high) state remains the same on the next observation, which brings the desirable amount of predictability for the user to take arbitrage decisions based on model results.

	Mean	SD		Mean	SD
μ	0.0000	0.0002	μ_{D_1}	(0.0000, 8.3042)	(0.0037, 0.1527)
ρ_y	0.0319	0.0088	μ_{D_2}	(0.0000, 6.5737)	(0.0026, 0.1135)
μ_y	0.0002	0.0209	π_1	0.3607	0.0297
σ_y	0.1545	0.0253	π_2	0.6393	0.0297
ν	8.6094	2.6278	BIC	-9,981	
log L	5,045		AICc	-10,062	

Table 13 – Posterior inference of static parameters for ICE:BRN.

	Σ_{D_1}			Σ_{D_2}	
	1	2		1	2
1	0.0118 (0.0010)	0.0001 (0.0025)	1	0.0097 (0.0005)	0.0001 (0.0017)
2	0.0001 (0.0025)	0.5020(0.0970)	2	0.0001(0.0017)	0.4622 (0.0580)

Table 14 – Posterior means for HMM classifier covariance matrix Σ_{D_i}. Standard deviations are in parenthesis.

	Low	High
Low	0.9936 (0.0030)	0.0064 (0.0030)
High	0.0043 (0.0019)	0.9957 (0.0019)

Table 15 – Posterior transition probability, $a_{i,j}$, means for ICE:BRN, log returns in an intuitive approach. Standard deviations are in parenthesis.

Figure 25 shows the pair log-returns y_t and posterior mean log volatility, $log\lambda$. Light gray dots are the observations classified as Low volatility state and dark gray dots are the observations classified as High volatility state. As can be seen in Table 14, values that belong to State 1 are more uniform, whereas the ones that belong to State 2 are more sparse and have higher values for the covariance matrix.

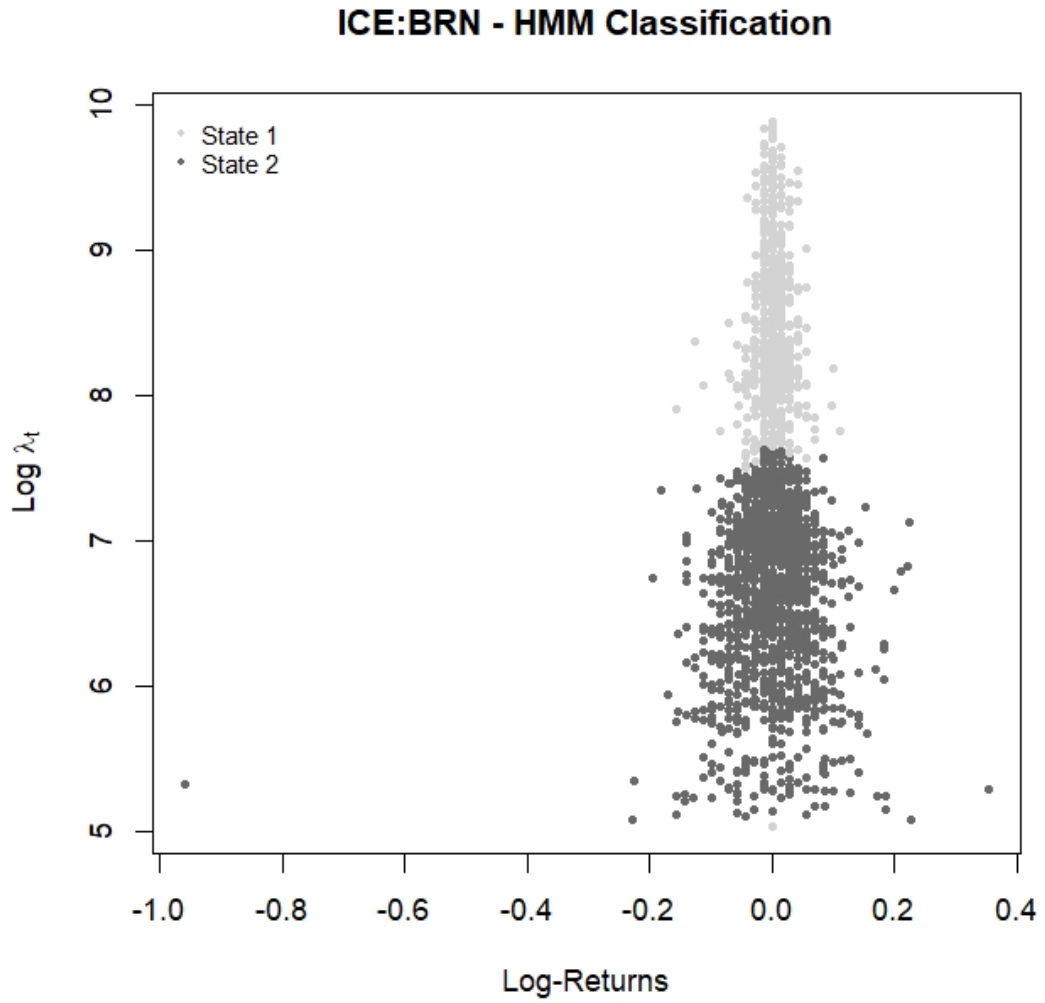


Figure 25 – HMM classification for market states according to values of log-return, y_t, and mean posterior log volatility, $log\lambda$.

Figure 26 shows intraday observations of ICE:BRN log-returns together with the

jump component, as black dots. The second graph shows mean posterior estimates of instantaneous volatility and the 95% credibility interval, obtained by the NGSVJ-HMM model. The third graph shows the instantaneous market regime for each observation, in an intuitive approach for the final user. The model was able to classify volatility increases on both days, which are propitious moments for arbitrage strategies.

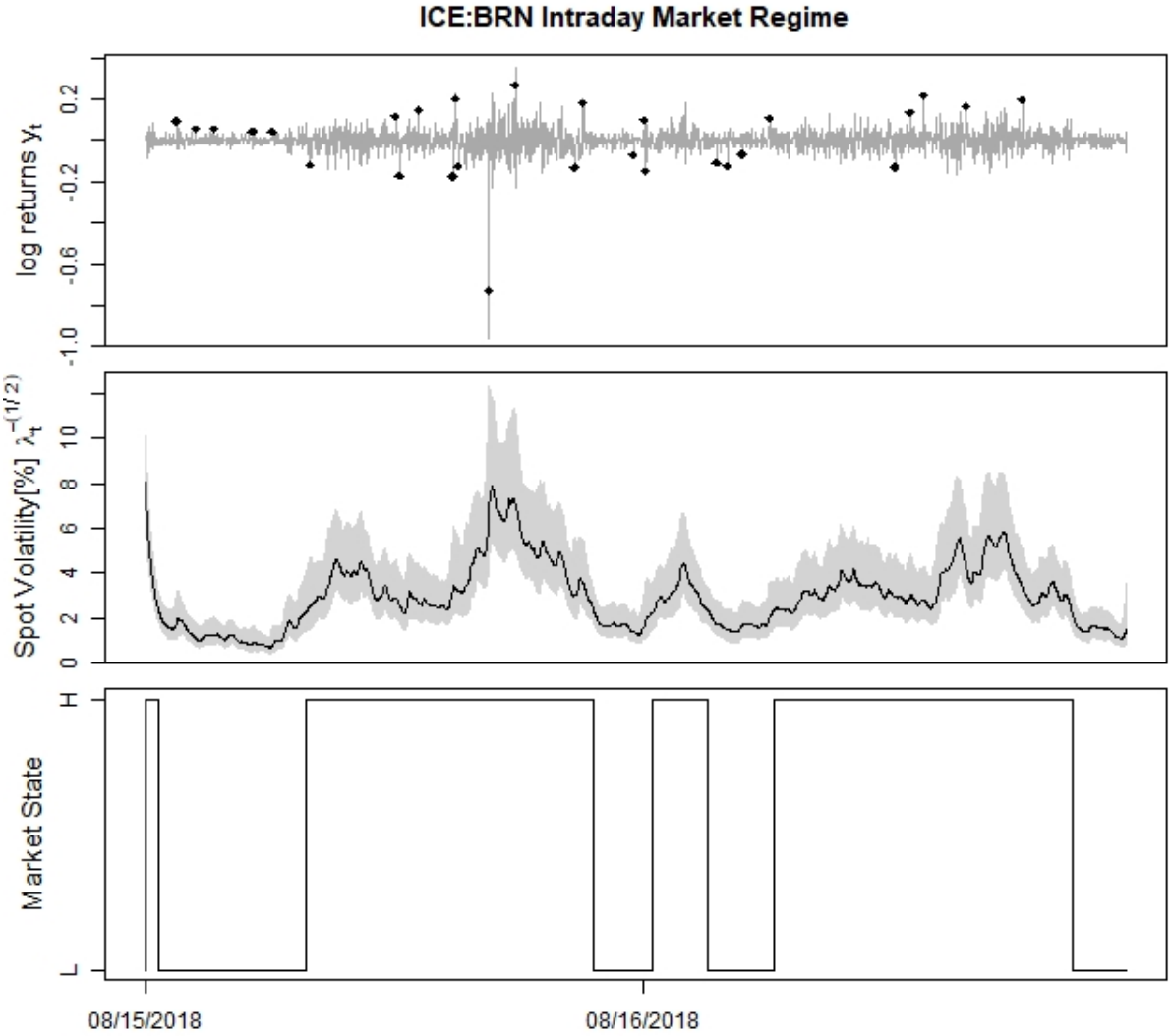


Figure 26 – Log returns for intraday ICE:BRN are shown at the top graph, together with jumps as dots; Mean posterior estimates of spot volatility in percentage $\lambda_t^{-1/2} \times 100\%$ at the middle graph in solid line and the gray area is the 95% credibility interval; and estimates of instantaneous Market Regime at the bottom graph.

Figure 27 shows market regime probabilities for each observation. The market state on a specific time t is that of higher probability. Also, as estimated state probability goes

closer to 1, the more certain the model is about its classification.

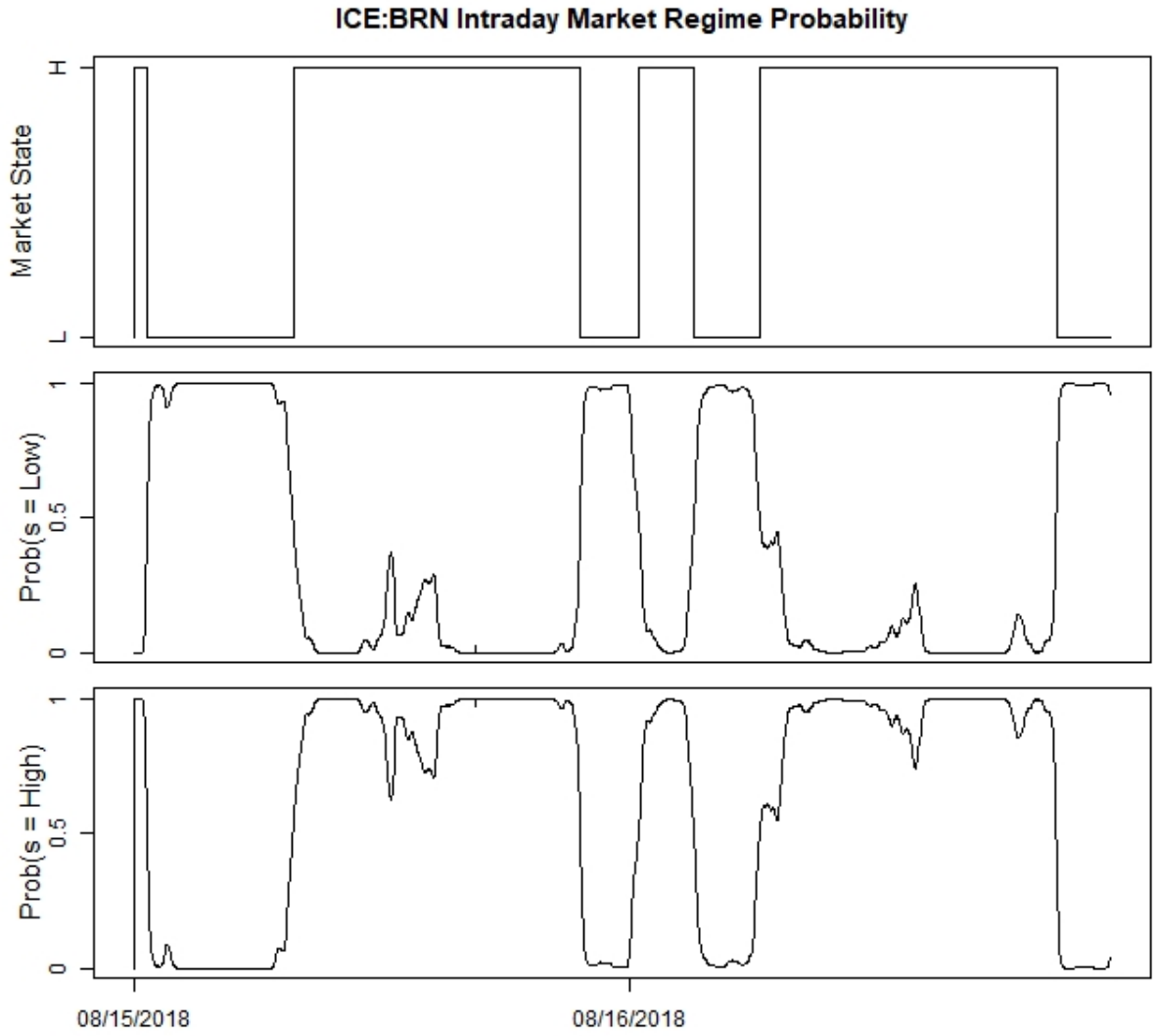


Figure 27 – Market Regime probability for each intraday observation of BRN:ICE log returns.

The model was able to effectively classify market states and the classification matches moments with higher estimated spot volatility. The model user can rely on market states classified by the model to take strategic arbitrage decisions in order to pursue short terms gain on financial markets, by speculating.

4.6 Additional Comments

The NGSVJ-HMM model was able to capture speculative movements in the market through the jump components and detect periods with increased market risk through the volatility component. Also, the decoding structure allows the model to deliver an intuitive approach to the final user, to assist arbitrage via HMM and investment decisions.

The most notable advantage of using NGSVJ-HMM is its computational simplicity derived from the structure of the NGSVJ model presented by Rego and Santos (2020), which grants an automatic sampling process for parameters, that allows to sample the volatility in block mostly using Gibbs sampler, into one single stage. Furthermore, the HMM structure can be included inside the NGSVJ structure allowing it to achieve fast convergence and implementation simplicity, which allows it to get reliable estimates with less MCMC iterations, since the model is mostly based on Gibbs Sampling steps, achieving convergence after approximately 10,000 iterations, so that it can give results fast enough to be used on practical situations. Another advantage is the model flexibility, since it include jumps and heavy-tailed distribution for innovations, allowing the model to give a more precise estimative of volatility, by not letting tail events interfere on the estimative.

The development of NGSVJ-HMM fills a gap of NGSVJ when compared to other models existing on literature, as SV with inclusion of Markov Switching structure, presenting detailed theory behind the inclusion of such structure on NGSVJ and presenting the NGSVJ-HMM model as an alternative model with a simpler computational structure, sice it is mostly based on Gibbs Sampling steps.

As a disadvantage when compared to models for similar application on literature is that the NGSVJ-HMM was conceived only for classifying the volatility, without allowing the feedback of this information into the model to achieve the jump effect on volatility. This approach requires a more complex structure and will be addressed to future works.

5 Multivariate Stochastic Volatility Model with Jumps

In finance, diversification is a technique of allocating the available capital in different investments to reduce the investment risk without sacrificing return. According to Drake and Fabozzi (2010), to build an efficient portfolio (that is, a group of available assets that jointly have the lowest risk, or volatility, for a given value of expected return) the investor or portfolio manager must be able to estimate both expected return and the covariance matrix so that the combination of assets and weights for each one of them can be chosen according to the investors strategy. The covariance matrix, in this case, not only provides information about the variance of each asset, which is a measure of the risk of this particular asset. It also provides the covariance between available investment options, which is essential for choosing the composition of the portfolio in order to reduce its risk.

As the ARCH-GARCH and stochastic volatility (SV) models are the most dominant approaches to modeling volatility in literature, extending them to the multivariate case was a natural way to deal with the portfolio risk management challenge. Early works as Harvey, Ruiz, and Shephard (1994) develop a generalization of the SV model but highlight that such generalizations to multivariate series can be difficult to estimate and interpret, Engle and Kroner (1995) present a multivariate generalized ARCH model within simultaneous equation systems, Kroner and Ng (1998) develop a general model based on GARCH to estimate time-varying covariances between asset returns. With the consolidation of such models, recent works improvements focus on including heavy-tailed distributions and a jump structure to obtain a more accurate estimation of risk. Chib, Nadari, and Shephard (2006) present a multivariate SV model that includes both heavy-tailed student-t distribution and series-specific jump components. Chib, Omori, and Asai (2009) present a collection of works in multivariate stochastic volatility (MSV) models available in the literature that includes leverage effects, a mean factor MSV model and the inclusion of dynamic correlations trough a Wishart Process on the MSV model. Luciano and Schoutens (2006) develop a Lévy multivariate model for financial assets which incorporates jumps, skewness, kurtosis and stochastic volatility. More recent works on this subject include new features to the existing class of SV multivariate models, as Izzeldin, Tsionas, and Michaelides (2019) that propose a MSV model that generalizes the approach of Dendramis *et al.* (2015) by including impulse response functions to capture e impact of large shocks on returns or structural breaks in volatility, as contrast to the jump structure used on literature. Clark and Marcellino (2019) investigate the usage of Normal-Wishart and non-conjugate priors on a vector autoregressive SV model when dealing with high dimensional

data, defending that the usage of non-conjugate priors give extra flexibility on their model. Fengler, Herwartz, and Raters (2017) explore the BEKK algorithm, named after its creator names initials, to estimate volatility for foreign exchange rates using a multivariate GARCH model. However, the model is only able to estimate volatility for pairs of assets, instead of the entire portfolio simultaneously, which is a drawback when dealing with real applications on financial markets, that require to estimate volatility and covariance matrix for a large number of assets inside the same portfolio.

A strong disadvantage of ARCH-GARCH and SV classes for multivariate applications is that, as stated by Yu and Meyer (2006), apart from the inherent problems of multivariate models, such as high dimensionality of the parameter space and the required positive semidefiniteness of covariance matrices, the likelihood function has no closed form for the multivariate ARCH-GARCH and SV models, requiring more complex computational methods to sample from the parameters posterior distributions or even approximations for those distributions. This turns our attention to the dynamic models (DM) class for modeling volatility, where resides the NGSVJ developed by Rego and Santos (2020).

The DM approach has the advantage of a simpler implementation and a flexible model structure since the model is written in the space state form. Quintana, Lourdes, Aguilar and Liu (2003) present the General Multivariate Dynamic Linear Model (GMDLM) and the Bayesian Dynamic Factor Model (BDFM) as models used for investment management, specifically for building mean-variance efficient portfolios.

Approaches using DM to multivariate financial time series on literature lacks relevant features for dealing with financial time series, such as jumps on returns to capture speculative movements and an effective way to deal with high dimensional assets portfolio log returns. Triantafyllopoulos (2014) uses a particle filter together with a Skew-T distribution for innovations in a multivariate approach. Lopes, McCulloch, and Tsay (2016) use a normal dynamic linear model but has the advantage of a sampling procedure fully based on Gibbs sampling. Nasri and Rémillard (2019) use a copula model for a multivariate approach from univariate dynamic models. The Dynamic dependence network models in West (2020) suffer from dimensionality issues due to the graphical approach. A multivariate extension for NGSVJ in Rego and Santos (2020) brings a model that includes a jump structure, heavy tails for log-returns, and a sampling procedure based on Gibbs sampler.

Rego and Santos (2020) present a univariate approach to estimating volatility from financial assets, including both heavy-tailed distribution and jump structure. Extending this model to the multivariate case is possible using the Matrix-Beta evolution structure, presented in Prado and West (2010). It has the advantage of an already developed retrospective analysis procedure, using a Wishart distribution, which is presented in this work, for sampling from the covariance matrix, Σ_t, posterior distribution, similar

to smoothing procedure of λ_t from Gamerman, Santos and Franco (2013). Although, this structure suffers from two constraints: the degree of freedom, h_t, for sampling from the Wishart distribution in the retrospective analysis must be such that $h_t > q-1$, where q is model dimensionality, the number of assets being jointly evaluated; and the discount factor β must be such that $\beta > \frac{q-2}{q-1}$. Such constraints are not a problem for low dimensionality applications, but, as q grows, the discount factor range becomes limited and it may be necessary to generate from singular Wishart distributions, which would increase the computational complexity of the algorithm.

The model with Beta-Bartlett evolution equation, present by Prado and West (2010), enables us to flexible the constraints regards the discount factor β and the model dimensionality, but lacks a retrospective analysis procedure, which is proposed in this thesis.

In this chapter, the Multivariate Stochastic Volatility Model with Jumps (MSVJ) is presented as an extension of the NGSVJ to the multivariate case. Two approaches are evaluated for the multivariate model: the Matrix-Beta and the Beta-Bartlett evolutions.

5.1 Multivariate Stochastic Volatility Model With Jumps and Matrix-Beta Evolution

The proposed Multivariate Stochastic Volatility Model with Jumps and Matrix-Beta Evolution (MSVJM) for extending the NGSVJ model, developed by Rego and Santos (2020) to the multivariate case, for an $q \times 1$ vector log-returns time series with observations $\underset{\sim}{y}_t$, is:

$$\underset{\sim}{y}_t = F_t'\underset{\sim}{\theta}_t + \underset{\sim}{J}_t + \underset{\sim}{v}_t, \qquad \text{where } \underset{\sim}{v}_t|\gamma_t \sim N_q(0, \gamma_t^{-1}\Sigma_t), \tag{5.1}$$

$$\underset{\sim}{\theta}_t = G_t\underset{\sim}{\theta}_{t-1} + \underset{\sim}{\omega}_t, \qquad \text{where } \underset{\sim}{\omega}_t \sim N_q(0, w_t\Sigma_t), \tag{5.2}$$

$$\Sigma_t^{-1} = \beta^{-1}\Sigma_{t-1}^{-1}\zeta_t, \qquad \text{where } \zeta_t|\mathcal{D}_{t-1}, \Phi \sim Beta_q\left(\frac{\beta h_{t-1}}{2}, \frac{(1-\beta)h_{t-1}}{2}\right), \tag{5.3}$$

$$\gamma_t \sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \tag{5.4}$$

where:

$$\Omega_t = \Sigma_t^{-1}, \qquad \underset{\sim}{J}_t = \underset{\sim}{\xi}_{t+1} \circ \underset{\sim}{N}_{t+1}, \qquad \xi_t \sim N_q(\mu_y, \Sigma_y),$$
$$P(N_{q_{t+1}} = 1) = \rho_q, \text{and } P(N_{q_{t+1}} = 0) = 1 - \rho_q$$

In this model $\underset{\sim}{y}_t$ follows the DLM defined by $\{F_t, G_t, \Sigma_t, \gamma_t, \omega_t\}$, with $F_t = I$ and $G_t = I$, for I being the identity matrix. $\underset{\sim}{y}_t$ denotes the t^{th} observed value of the $q \times 1$ vector time series of portfolio's assets log-returns, $\underset{\sim}{\theta}_t$ is the $q \times 1$ vector of latent states, Σ_t is the

covariance matrix at time t and the main interests lay on estimating its value over time, since it is the main variable on risk and stock options pricing. As Σ_t is positive definite, by defining a proper prior to Σ_t the full conditional posterior distribution is also proper. w_t and β are specified discount factors. Initialization for θ and Σ are $\theta_0|\mathcal{D}_0, \Sigma_0 \sim N_q(\underset{\sim}{m_0}, c_0)$ and $\Sigma_0|\mathcal{D}_0, \theta_0 \sim IW(n_0, D_0)$. γ_t is the variance mixture component and, depending on the choice of its prior distribution, the v_t unconditional posterior distribution can assume specific heavy-tailed distribution, e.g., by using the mixture component $\gamma_t \sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, the unconditional marginal distribution of errors assume a multivariate $t_\nu(\underset{\sim}{0}, I)$ distribution, for a specified ν, so we achieve the effect of a heavy-tailed distribution for the innovations.

The gamma mixture component γ_t was conceived as a scalar in order to preserve the correlations between log-returns in the covariance matrix Σ_t. Note that, as a scalar, we have that:

$$\gamma_t^{-1}\Sigma_t = \gamma_t^{-1}\begin{bmatrix} \sigma_{1,1,t} & \sigma_{1,2,t} & \dots & \sigma_{1,q,t} \\ \sigma_{1,2,t} & \sigma_{2,2,t} & \dots & \sigma_{2,q,t} \\ \vdots & & \ddots & \vdots \\ \sigma_{1,q,t} & \sigma_{2,q,t} & \dots & \sigma_{q,q,t} \end{bmatrix} = \begin{bmatrix} \gamma_t^{-1}\sigma_{1,1,t} & \gamma_t^{-1}\sigma_{1,2,t} & \dots & \gamma_t^{-1}\sigma_{1,q,t} \\ \gamma_t^{-1}\sigma_{1,2,t} & \gamma_t^{-1}\sigma_{2,2,t} & \dots & \gamma_t^{-1}\sigma_{2,q,t} \\ \vdots & & \ddots & \vdots \\ \gamma_t^{-1}\sigma_{1,q,t} & \gamma_t^{-1}\sigma_{2,q,t} & \dots & \gamma_t^{-1}\sigma_{q,q,t} \end{bmatrix} \tag{5.5}$$

So that the pearson correlation between two log return time series would be given by:

$$corr_{i,j,t} = \frac{\gamma_t^{-1}\sigma_{i,j,t}}{\sqrt{\gamma_t^{-1}\sigma_{i,i,t}\gamma_t^{-1}\sigma_{j,j,t}}} = \frac{\cancel{\gamma_t^{-1}}\sigma_{i,j,t}}{\cancel{\gamma_t^{-1}}\sqrt{\sigma_{i,i,t}\sigma_{j,j,t}}} = \frac{\sigma_{i,j,t}}{\sqrt{\sigma_{i,i,t}\sigma_{j,j,t}}} \tag{5.6}$$

where $\sigma_{i,j,t}$ is the correlation between log-returns for asset i and j on time t and $\sigma_{i,i,t}$ the variance of log-returns for asset i on time t. In this case, the correlation between log-returns for assets inside the portfolio is not affected by the mixture component γ_t, property that is not achieved by using a non-scalar mixture component that would lead to a matrix multiplication.

Each series on vector $\underset{\sim}{y_t}$ has a jump component, $J_{i,t}$, $i = 1, \dots, q$, composed by the element-wise product, denoted by operator $\circ$, of the jump indicator N_{t+1} and magnitude $\xi \sim N_q(\underset{\sim}{\mu_y}, \Sigma_y)$, in the same way proposed by Rego and Santos (2020). This jump structure gives extra flexibility to the model by allowing each vector of observations that compose $\underset{\sim}{y_t}$ to assume its own jump characteristics, which is particularly important when dealing with assets that have different intrinsic behavior.

5.2 Bayesian Inference

For ease notation, let $\Phi = (\underset{\sim}{\theta}, \underset{\sim}{J}, \gamma, \Sigma, \underset{\sim}{\mu_y}, \Sigma_y, \underset{\sim}{\rho}, \nu)$, excluding the parameter being evaluated, i.e. $\Phi_{[-\gamma]} = (\underset{\sim}{\theta}, \underset{\sim}{J}, \Sigma, \underset{\sim}{\mu_y}, \Sigma_y, \underset{\sim}{\rho}, \nu)$. Detailed calculations on how to find the posterior distributions henceforth mentioned can be found on Appendix A. Independent

priors were chosen so that proper posteriors are obtained, simplifying the sample procedure. When no initial information is available, non-informative priors can be chosen.

The latent state parameter, $\underset{\sim}{\theta}_t$, and covariance matrix, Σ_t, are jointly sampled using the Forward Filtering Backward Sampling algorithm proposed by Prado and West (2010). In this case, w_t is defined via a single discount factor ι. All distributions for states and observational covariance matrix in the sequential updating and retrospective smoothing are multivariate Normal Inverse Wishart, $NIW(\theta_t, \Sigma_t | \underset{\sim}{m}_t, c_t, n_t, D_t, \Phi_{[-(\theta,\Sigma)]})$, where $\underset{\sim}{m}_t$, c_t, n_t, D_t parameters are the mean, scale, degree of freedom and covariance matrix, respectively. Initialization for θ and Σ are $\theta_0|\mathcal{D}_0, \Sigma_0 \sim N_q(\underset{\sim}{m}_0, c_0)$ and $\Sigma_0|\mathcal{D}_0, \theta_0 \sim IW(n_0, D_0)$. The summary sequential updating equations at times $t-1$ to t are:

$$\underset{\sim}{m}_t = \underset{\sim}{m}_{t-1} + k_t e_t \quad \text{and} \quad c_t = r_t - k_t^2 q_t, \tag{5.7}$$

$$D_t = \beta D_{t-1} + \frac{\gamma_t}{c_t} e_t e_t^T \quad \text{and} \quad n_t = \beta n_{t-1} + 1, \tag{5.8}$$

where $\underset{\sim}{e}_t = \underset{\sim}{y}_t - \underset{\sim}{m}_{t-1} - J_t$, $r_t = c_{t-1} + w_t \equiv c_{t-1}/\iota$, $q_t = r_t + 1$ and $k_t = r_t/q_t$. The summary equations for the retrospective computations over $t = (T-1) : 1$ are, for the states, given by:

$$\underset{\sim}{z}_T(t-T) = (1-\iota)\underset{\sim}{m}_t + \iota \underset{\sim}{z}_T(t-T+1), \tag{5.9}$$

$$r_T(t-T) = (1-\iota)c_t + \iota^2 r_T(t-T+1). \tag{5.10}$$

The smoothing procedure for the covariance matrix Σ_t in MSVJM is analogous to the procedure for retrospective sampling from the volatility component, λ_t, in the NGSVJ model developed by Rego and Santos (2020). By using the Wishart distribution as a multivariate generalization of the Gamma distribution, or analogously, the Inverse Wishart distribution as a generalization of the Inverse Gamma distribution, a very similar procedure for retrospective sampling is obtained.

Proposition 1 *A draw from the conditional posterior distribution* $p(\Sigma_t|\Sigma_{t+1}, \mathcal{D}_t)$ *can be obtained by setting* $\Sigma_t^{-1} = \Omega_t$ *and recursively sampling from:*

$$(\Omega_t - \beta\Omega_{t+1}|\mathcal{D}_t) \sim W((1-\beta)h_t, D_t^{-1}) \tag{5.11}$$

so that $(\Sigma_t - \beta\Sigma_{t+1}|\mathcal{D}_t) \sim IW((1-\beta)h_t, D_t)$.

A more detailed approach on the proof of Proposition 1 can be found in Appendix B.

Retrospective simulation of the covariance matrix, Σ_t sequence is obtained as in Proposition 1 and the algorithm for sampling from its full conditional posterior distribution proceeds precisely as defined in Appendix B.2.

Note that the Matrix-Beta evolution is a generalization of the evolution Beta in the NGSVJ model, developed by Rego and Santos (2020). By evaluating a single return time series, Ω_t in MGSVJM is equivalent to λ_t in NGSVJ and the Matrix-Beta evolution is equivalent to the evolution Beta procedure.

For the mixture component γ_t, a gamma distribution $G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ is defined, which, when mixed as γ_t^{-1}, resulting in Inverse-Gamma, leads to a multivariate Student-t with ν degrees of freedom to the innovations. The full conditional posterior distribution is:

$$p(\gamma_t|\mathcal{D}_n, \Phi_{[-\gamma_t]}) \sim G\left(\frac{\nu+q}{2} - 1, \frac{\nu^*}{2}\right), \tag{5.12}$$

$$\nu^* = \nu + tr\left((\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)^T(\Sigma_t)^{-1}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)\right). \tag{5.13}$$

The parameter ν will be specified, since its posterior distribution does not have closed-form, leading to a Metropolis step. Since the main objective of this work is defining the methodology and a general procedure for extending the NGSVJ, we preserve its initial structure and address further developments to future works. Recall that our focus in this chapter is on expanding the applications of the model to a portfolio of assets, presenting the MSVJM as an alternative to the multivariate models commonly used on literature.

The jump sizes ξ_{t+1} follow a $N_q(\mu_y, \Sigma_y)$, for $\mu_y = (\mu_{y_1}, \mu_{y_2}, \ldots, \mu_{y_q})$ and $\Sigma_y = diag(\sigma^2_{y_1}, \sigma^2_{y_2}, \ldots, \sigma^2_{y_q})$, q is the dimension of $\underset{\sim}{y}_t$. For the mean μ_y a non-informative prior $\mu_{y_q} \sim N(m, v)$ is set, resulting in a full conditional posterior:

$$p(\mu_{y_q}|\mathcal{D}_n, \Phi_{[-\mu_y]}) \sim N\left(\frac{m\sigma^2_{y_q} + vn_q\bar{\xi_q}}{\sigma^2_{y_q} + n_q v}, \frac{v\sigma^2_{y_q}}{\sigma^2_{y_q} + n_q v}\right). \tag{5.14}$$

For the covariance matrix Σ_y, a prior $\sigma^2_{y_q} \sim IG(\alpha, \beta)$ is assumed for each of the diagonal elements, resulting in the full conditional posterior:

$$p(\sigma^2_{y_q}|\mathcal{D}_n, \Phi_{[-\sigma^2_y]}) \sim IG\left(\alpha + \frac{n_q}{2}, \beta + \frac{\sum_{\substack{i=1 \\ J_i \neq 0}}^{n} (\xi_{q,i+1} - \mu_{y_q})^2}{2}\right). \tag{5.15}$$

In both cases, n_q is the number of times that the jump is observed on dimension q, and $\bar{\xi}$ the mean of jump sizes $\xi_{q,t}$.

As the prior of jump sizes are assumed to be Multivariate Normal, the full conditional posterior is also Multivariate Normal, given by:

$$p(\xi_{t+1}|\mathcal{D}_n, \Phi_{[-\xi]}) \sim N_q\left(m^*_\xi, v^*_\xi\right). \tag{5.16}$$

where:

$$m^*_\xi = \left(\Sigma_y^{-1} + (\gamma^{-1}\Sigma_t)^{-1}\right)^{-1}\left(\Sigma_y^{-1}\mu_y + (\gamma^{-1}\Sigma_t)^{-1}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t)\right) \tag{5.17}$$

$$v^*_\xi = \left(\Sigma_y^{-1} + (\gamma^{-1}\Sigma_t)^{-1}\right)^{-1} \tag{5.18}$$

For jump probabilities $\rho = (\rho_1, \rho_2, \ldots, \rho_q)$, a prior $Beta(\alpha, \beta)$ is set. The full conditional posterior, for a sample size n, is given by:

$$p(\rho_q|\mathcal{D}_n, \Phi_{[-\rho]}) \quad \sim \quad Beta\left(\alpha + \sum_{i=0}^{n} N_{q,i}, \beta + n - \sum_{i=0}^{n} N_{q,i}\right) \tag{5.19}$$

Since the jump indicator N^y can assume only two values, 0 or 1. The probability of observation at $t+1$ be a jump is given by:

$$P(N_{q,t+1} = 1|\mathcal{D}_{t+1}, \Phi_{[-N]}) \quad \propto \quad \rho_q P(\underset{\sim}{y}_{t+1}|N_{q,t+1} = 1, \Phi_{[-N]}). \tag{5.20}$$

which is easy to calculate, since $P(\underset{\sim}{y}_{t+1}|N_{q,t+1} = 1, \Phi_{[-N]})$ is a Multivariate Normal distribution. If this probability is greater then a threshold α, then $N_{q,t+1} = 1$. The threshold α is chosen such that the number of jumps identified corresponds to the estimate of the jump intensity ρ_q.

5.3 General Procedure

Here we present the Gibbs Sampler algorithm to sample from the MSVJM model's parameters.

Let $\underset{\sim}{Y}_n = \{\underset{\sim}{y}_t\}_{t=1}^n$, $\underset{\sim}{\theta} = \{\underset{\sim}{\theta}_t\}_{t=1}^n$, $\Sigma = \{\Sigma_t\}_{t=1}^n$, $\underset{\sim}{J} = \{\underset{\sim}{J}_t\}_{t=1}^n = \{\xi_{t+1}N_{t+1}\}_{t=1}^n$, $\gamma = \{\gamma_t\}_{t=1}^n$, $\xi = \{\xi_{t+1}\}_{t=1}^n$, $N = \{N_{t+1}\}_{t=1}^n$ and prior probability density $p(\gamma)$, $p(\mu_y)$,$p(\Sigma_y)$, $p(\xi)$,$p(\underset{\sim}{\rho})$ are set for $\gamma, \mu_y, \Sigma_y, \xi, \underset{\sim}{\rho}$. Then, a sample of size M from the joint posterior distribution $p(\underset{\sim}{\theta}, \Sigma, \gamma, \mu_y, \Sigma_y, \underset{\sim}{J}, \underset{\sim}{\rho}|\underset{\sim}{Y}_n)$ is drawn via Gibbs Sampler, whose follows:

1. Initialize $\underset{\sim}{\theta}^{(0)}, \Sigma^{(0)}, \gamma^{(0)}, \mu_y^{(0)}, \Sigma_y^{(0)}, \xi^{(0)}, N^{(0)}$ and $\underset{\sim}{\rho}^{(0)}$.
2. Set $j = 1$.
3. Block sample $\underset{\sim}{\theta}^{(j)}, \Sigma^{(j)}|\underset{\sim}{Y}_n, \gamma^{(j-1)}, J^{(j-1)}$ using the FFBS algorithm presented in Section 5.2 by:
   a) Sample $\{\Sigma_t^{-1}\}_{t=1}^n$ as in the algorithm in Appendix B.2 using as parameters values in Eq. (5.8).
   b) Sample $\underset{\sim}{\theta}_t \sim N_q(a_t, r_t\Sigma_t)$ using the smoothed parameters in Eq. (5.9) and Eq. (5.10).
4. Block sample $\gamma^{(j)}|\underset{\sim}{Y}_n, \theta^{(j)}, \Sigma^{(j)}, \underset{\sim}{J}^{(j-1)}$ as in Eq. (5.12).
5. Sample $\mu_y^{(j)}|\xi^{(j-1)}, \Sigma_y^{(j-1)}$ as in Eq. (5.14).
6. Sample $\Sigma_y^{(j)}|\xi^{(j-1)}, \mu_y^{(j)}$ as in Eq. (5.15).
7. Block sample $\underset{\sim}{J}^{(j)}|\underset{\sim}{Y}_n, \underset{\sim}{\theta}^{(j)}, \Sigma^{(j)}, \gamma^{(j)}, \mu_y^{(j)}, \Sigma_y^{(j)}$ by

a) Block sample $\xi^{(j)}|\underset{\sim}{Y_n}, \underset{\sim}{\theta}^{(j)}, \Sigma^{(j)}, \gamma^{(j)}, \mu_y^{(j)}, \Sigma_y^{(j)}$ as in Eq. (5.16).

b) Block sample $N^{(j)}|\underset{\sim}{Y_n}, \underset{\sim}{\theta}^{(j)}, \Sigma^{(j)}, \gamma^{(j)}, \xi^{(j)}$ as in Eq. (5.20).

8. Sample $\underset{\sim}{\rho}^{(j)}|\underset{\sim}{J}^{(j)}$ as in Eq. (5.19).

9. Set $j = j + 1$.

10. If $j \leq M$, go to step 3, otherwise stop.

Since all full conditional posterior distribution has closed-form, only Gibbs Sampler steps are used. The MSVJM structure is similar to the NGSVJ, but allowing the estimation of the covariance matrix for a portfolio of assets, instead of a single asset at time, which is specially relevant for financial market applications.

5.4 Simulation

To illustrate the performance of the MSVJM, we adapt the method to synthetic data from the model proposed by Warty, Lopes and Polson (2018) to the multivariate case. To generate from the covariance matrix, Σ_t, we generate the diagonal values $i = 1, \ldots, q$ from:

$$v_{i,i,t} = v_{i,i,t-1} + \kappa(\theta - v_{i,i,t-1})\Delta + \rho\sigma_v\sqrt{v_{i,i,t-1}\Delta}\epsilon_{1,t} + \sigma_v\sqrt{(1-\rho^2)v_{i,i,t-1}\Delta}\epsilon_{2,t} \tag{5.21}$$

where $\epsilon_{1,t}$ and $\epsilon_{2,t} \sim N(0,1)$. The covariance matrix off-diagonal values for each time t are obtained by making:

$$\sigma_{i,j,t} = \rho_{\sigma,i,j} \times \sqrt{v_{i,i,t} \times v_{j,j,t}}, \quad i \neq j \tag{5.22}$$

for $i = 1, \ldots, q$, $j = 1, \ldots, q$, $t = 1, \ldots, n$ and $\rho_{\sigma,i,j}$ the correlation between $v_{i,i,t}$ and $v_{j,j,t}$.

Synthetic data for returns is then generated from:

$$\underset{\sim}{y_t} = N_q(\mu + \underset{\sim}{J_t}, \gamma_t^{-1}\Sigma_t), \tag{5.23}$$

$$\underset{\sim}{J_t} = \xi_{t+1}N_{t+1}, \tag{5.24}$$

where the jump times, $N_{q,t}$ are generated from a Bernoulli(ρ_q), jump sizes ξ_t from $N_q(\mu_y, \Sigma_y)$, and γ_t from $G(\frac{\nu}{2}, \frac{\nu}{2})$. Setup of parameters was: log-returns mean $\mu = (0.05, 0.01, -0.04)$; jump probability $\rho_q = (0.015, 0.02, 0.009)$; jump magnitude mean $\mu_y = (-2.5, -1.0, 2.0)$ and covariance matrix $\Sigma_y = diag(16, 4, 9)$; variance mixture degrees of freedom $\nu = 30$; Returns covariance matrix diagonal components $\Delta = 1$, $\theta = 0.8$, $\kappa = 0.015$, $\sigma_v = 0.1$, $\rho = 0.4$, same used by Rego and Santos (2020). Correlations were defined as $\rho_{\sigma,1,2} = 0.8$, $\rho_{\sigma,1,3} = -0.4$, $\rho_{\sigma,2,3} = 0.01$. 30 replications were maid to evaluate model performance.

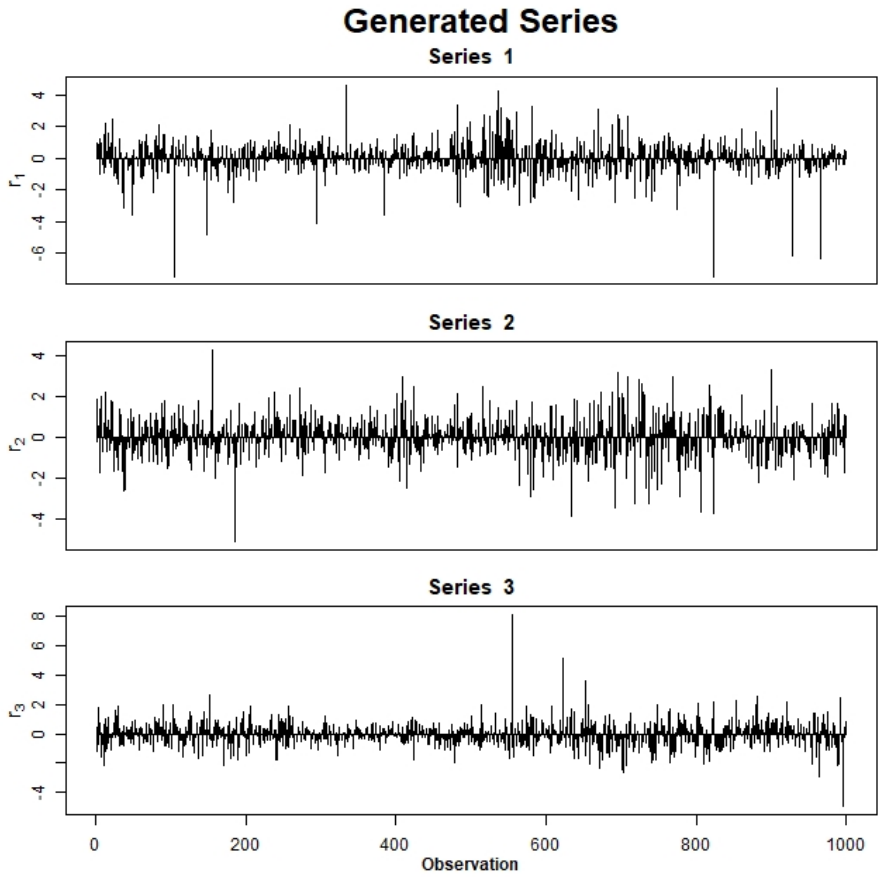


Figure 28 – Simulation study: One simulated realization (n = 1,000).

The simulated portfolio consists of approximately 4 years of daily data (n = 1,000). Figure 28 shows one realization generated from the model. All codes for the model estimation were written in R software, using package Rcpp, available at The Comprehensive R Archive Network (CRAN). Machine specifications are Intel Core i7-8700K 3.70 GHz processor, 16GB RAM, using a 64-bit Windows 10 Operating System. MCMC retained 2,020 samples after 40,000 iterations, burn-in of 20,000 and lag of 10 iterations.

Figure 29 shows the time series of true instantaneous volatility $v_{i,t}$, together with the estimated volatility $\sigma^{-1}_{i,i,t}$, for $i = 1, \ldots, 3$. The MSVJM is able to closely track the latent state. Almost every point from the true volatility is inside the 95% credibility interval, even when the estimate mean slightly deviates from the true value. Figure 30 shows the estimated correlation between the time series. True value ρ_σ is shown in dashed line.

Figure 5.4 shows the time series of true instantaneous jumps J_t, together with the estimated jumps J_t. Recall that the jumps represent moments of punctual abnormal returns, caused by the market's speculative movements. The MSVJ is able to catch most of the simulated jumps, together with its magnitudes.

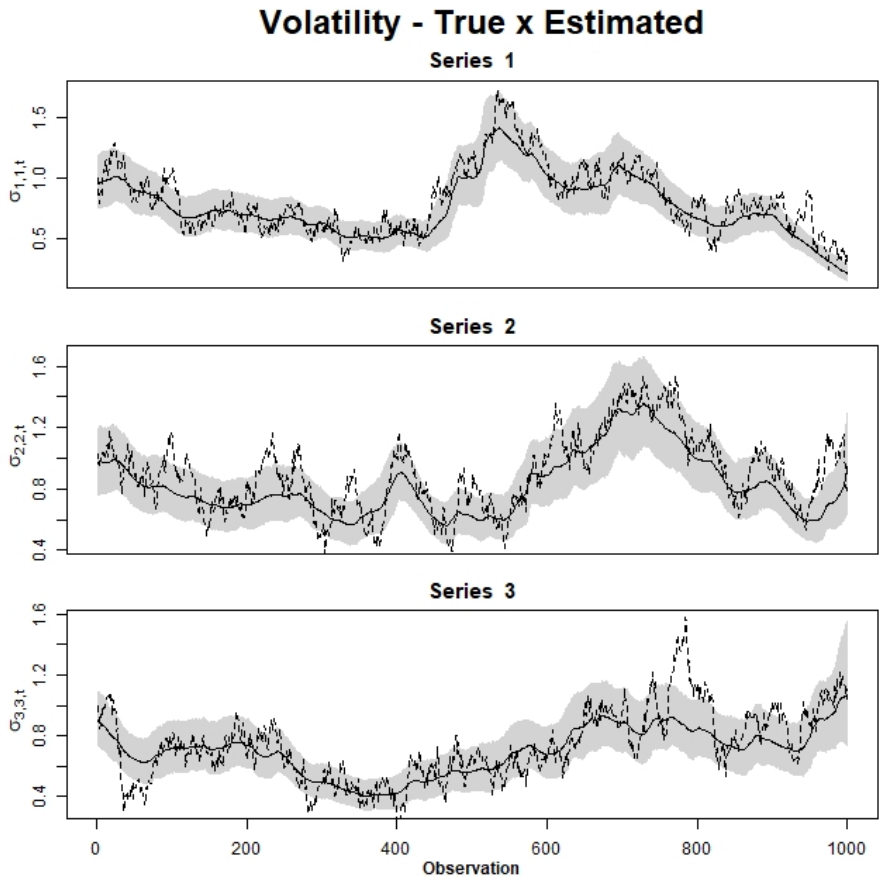


Figure 29 – Simulation study: Posterior mean estimates of instantaneous volatility $v_{i,t}$ for simulated data. True volatility series shown in dashed line; posterior mean estimates $\sigma_{i,i,t}$ in solid line; and 95% credibility interval is the light gray area.

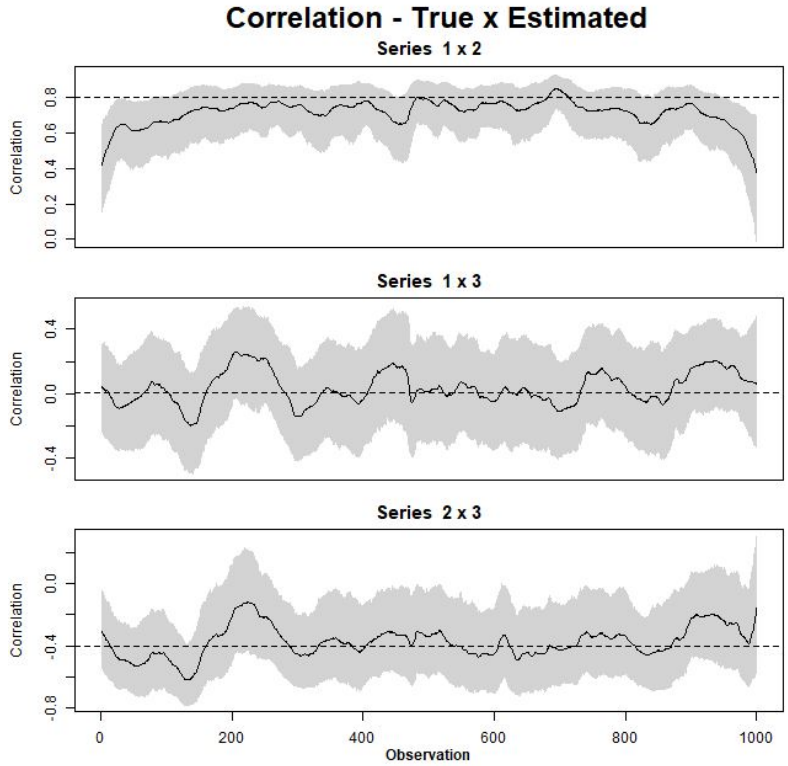


Figure 30 – Simulation study: Posterior estimates of correlation coefficient ρ_σ for simulated data. True value is shown in dashed line; posterior mean estimates in solid line; and 95% credibility interval is the light gray area.

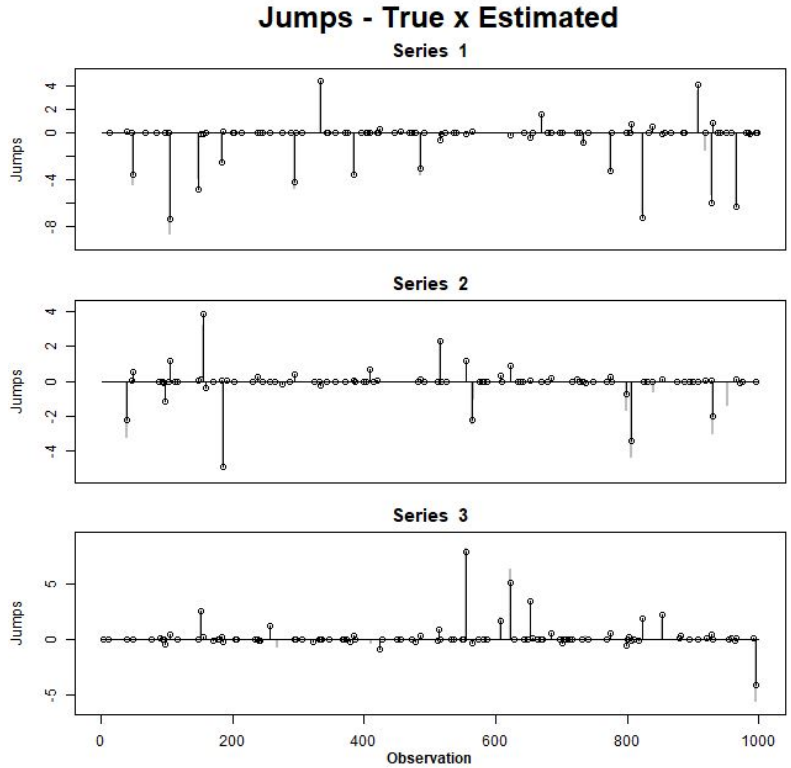


Figure 31 – Simulation study: Posterior estimates of instantaneous jumps J_t for simulated data. True values of the time series shown in gray; posterior mean estimates J_t in black.

Table 16 presents true values and the summary information for the posterior estimates over the 30 replications. As jump related parameters are estimated based only on values recognized by the model as jumps, they tend to have higher variance and root mean square error when compared to true values. Despite that, estimations obtained by the model are close to real parameters.

	True	MCMC Mean	SD	RMSE
ρ_{y_1}	0.015	0.0537	0.0068	0.0015
ρ_{y_2}	0.020	0.0392	0.0060	0.0004
ρ_{y_3}	0.009	0.0594	0.0101	0.0026
μ_{y_1}	-2.5	-3.5700	0.4836	1.3715
μ_{y_2}	-1.0	-1.1340	0.3311	0.8562
μ_{y_3}	2.0	0.35545	0.3932	0.5654
σ_{y_1}	4.0	4.2088	0.3352	0.1524
σ_{y_2}	2.0	1.7006	0.3908	0.2376
σ_{y_3}	3.0	3.6459	0.2950	0.5015
log L	-		-3119	
BIC	-		**6405**	
AICc	-		**6289**	

Table 16 – Posterior estimates descriptive estatistics of MSVJM mean posterior static parameters for simulated daily returns (n = 1,000), over 30 replications.

5.5 Model application

The MSVJM model is applied to a dataset containing nine exchange rates against U.S. dollar: Brazilian Reals (BZUS), Canadian Dollar (CAUS), Chinese Yuan (CHUS), Danish Kroner (DNUS), Hong Kong Dollar (HKUS), Indian Rupees (INUS), Japanese Yen (JPUS), South Korean Won (KOUS), Malaysian Ringgit (MAUS), obtained from Federal Reserve of St. Louis website. The dataset contains 2,020 daily observations from December 2010 to January 2019. Table 17 provides descriptive statistic for the log-returns, multiplied by 100. They will be referred in the model as time series $y_1, y_2, \ldots, y_9$ in the same order mentioned above.

With exception of HKUS, all assets present excess of returns, as observed by extreme values on minimum and maximum observations. Mean returns for all assets are close to 0%, since currency exchange rates tends to be stable on the long run, except on extreme cases when countries go bankrupt, get involved on long duration wars, etc.

Exchange rates are closely linked to the countries economy healthy and constant declines or increases may indicate permanent changes on its economic structure.

Desc. Stats	BZUS (y_1)	CAUS (y_2)	CHUS (y_3)	DNUS (y_4)	HKUS (y_5)
Mean	0.0385	0.0140	0.0015	0.0073	0.0003
St. Dev.	0.9397	0.4967	0.1805	0.5533	0.0338
Min	-5.2991	-2.9001	-1.2417	-3.0555	-0.4422
Max	8.6670	3.3678	1.8161	2.6461	0.3345
Desc. Stats	INUS (y_6)	JPUS (y_7)	KOUS (y_8)	MAUS (y_9)	
Mean	0.0214	0.0126	-0.0010	0.0136	
St. Dev.	0.5007	0.5884	0.5396	0.4507	
Min	-3.7560	-3.4977	-2.9630	-3.6571	
Max	3.7919	3.3428	2.8113	2.7750	

Table 17 – Exhange rates log-returns [×100%] descriptive statistics.

5.5.1 Parameters Specification

For a portfolio of $q = 9$ log-return series, the FFBS related parameters setup was $\theta_0|\mathcal{D}_0, \Sigma_0 \sim N_9(0, 100)$ and $\Sigma_0|\mathcal{D}_0, \theta_0 \sim IW(20, I_9)$, with discount parameters $\iota = 0.99$ and $\beta = 0.95$, as suggested in Prado and West (2010). For the mixture component γ_t a $G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ distribution, with $\nu = 30$, was specified, following Rego and Santos (2020). For jump related components, μ_{y_q}, N(0,100) priors were specified, for $\sigma^2_{y_q}$ a IG(0.1,0.1) prior, an for ρ_q a $Beta(2, 40)$, for $q = 1, \ldots, 9$, since it is known from previous works in literature that jump events are uncommon so that, by setting an informative prior, better results for convergence can be achieved. The threshold was fixed at $\alpha = 0.7$, following Rego and Santos (2020).

The results were obtained with a 40,000 iteration chain, a burn-in of 20,000 observations, with a lag of 10 observations, resulting in 2,500 samples. MCMC chain convergence was verified through graphical methods. All the coding was done in the R software, using the RcppArmadillo package.

Convergence of MCMC chains was verified trough graphical methods. Residual analysis results does not show evidence of strong violations of the model hypotheses.

5.5.2 Results

Figures 32, 33 and 34 shows the posterior mean estimates of instantaneous standard deviation, or volatility, $diag(\Sigma_t)$, for exchange rates log-returns. Figures 35 to 40 shows the correlation between different currencies exchange rates. The MSVJM model is capable of estimating not only the volatility of log-returns but also the correlation between the time series. This feature is especially relevant when dealing with an investment portfolio

that contains several assets, as an investor would like to diversify his portfolio by choosing negative or non-correlated assets to reduce the portfolio risk.

Higher values of volatility are observed on BZUS, peaking 1.5% in 2016. The period of higher volatility for Brazilian currency matches a moment of political instability in the country, which caused the currency to plunge in front of the United States dollar. On the other hand, the smallest values of volatility are observed on HKUS, mostly under 0.1%. The reason is that, according to Ranganathan (2018) on Reuters, the Hong Kong dollar was pegged at a fixed rate of 7.8 to the U.S. dollar in October 1983 and, since May 2005, it has been allowed to move between 7.75 and 7.85, a narrow range, thus restricting its volatility. Another example is an increase on volatility observed on early 2013 for JPUS, which matches political changes in Japan after elections on December 2012. An economic analysis of volatility causes is out of the scope of this work, with major events being cited only with illustration purposes.

The increase on estimated volatility on the first periods of CHUS and HKUS series that contrasts with the subsequent period is caused by the effect of prior specification that is the same for every time series, since $\Sigma_0|\mathcal{D}_0, \theta_0 \sim IW(20, I_9)$. This disadvantage is easily overcome by increasing the analysis period beyond the desired study time interval and ignoring the estimates for the first observations in applications where the behavior of assets volatility is considerable different from one another.

Another point of attention is the increase of volatility in late 2011 and early 2012 for most exchange rates. In this period, the eurozone debt crisis dramatically worsened, with Portugal requiring a bailout package to stabilize its public finances and Spain, Ireland, and Greece with rising debt, leading to a substantial downgrade of those countries' credit rating, affecting all European Union. In such a scenario, investors tend to look towards the United States dollar as a safe currency for keeping their capital, causing all exchange rates to fall in front of the U.S. dollar. This effect can be seen on Figures 35 to 40 as an increase in correlation between exchange currencies, with exception of JPUS.

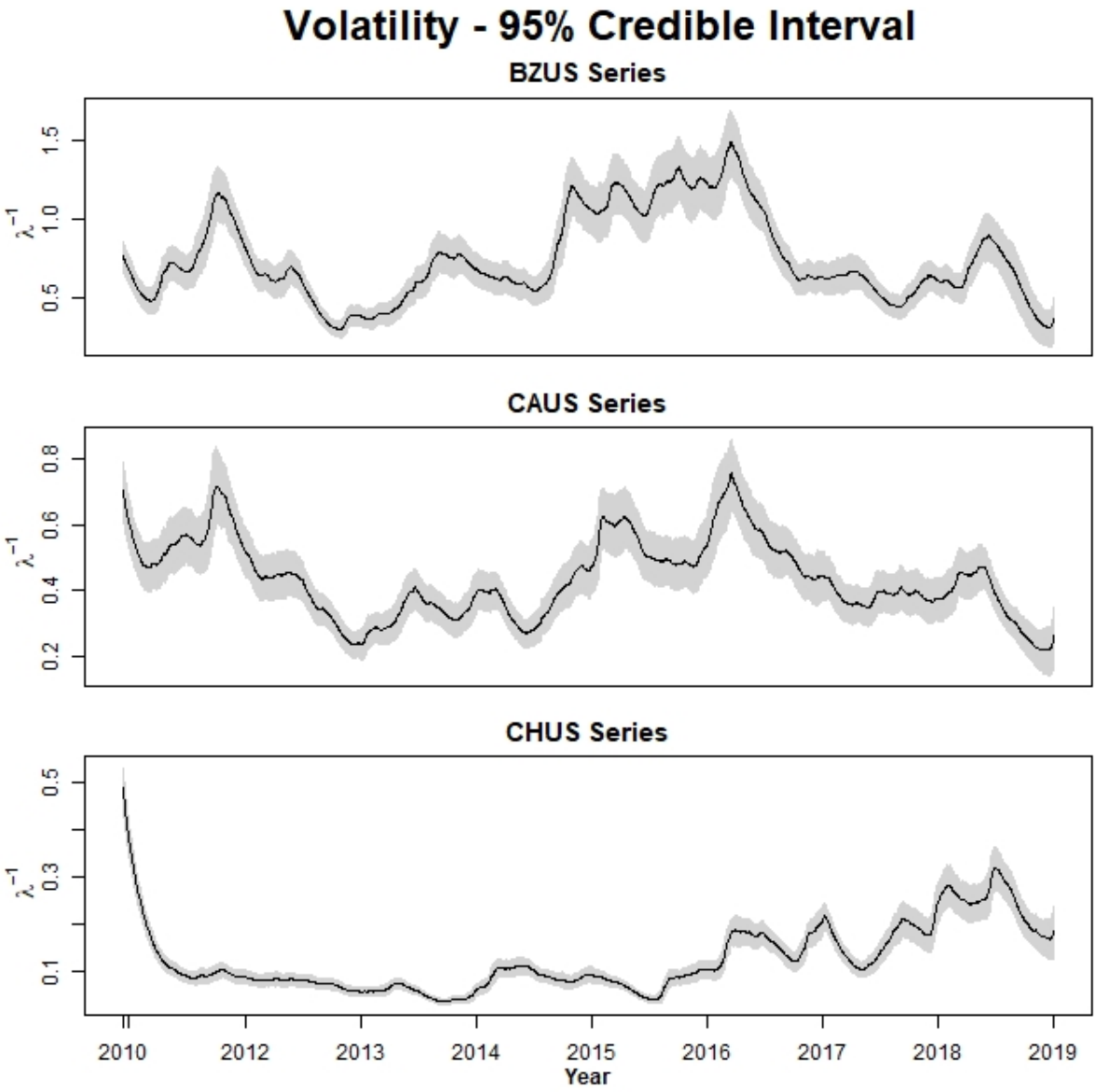


Figure 32 – Posterior mean estimates of instantaneous volatility for Exchange Rates data. Mean estimates in solid line; and 95% credibility interval is the gray area.

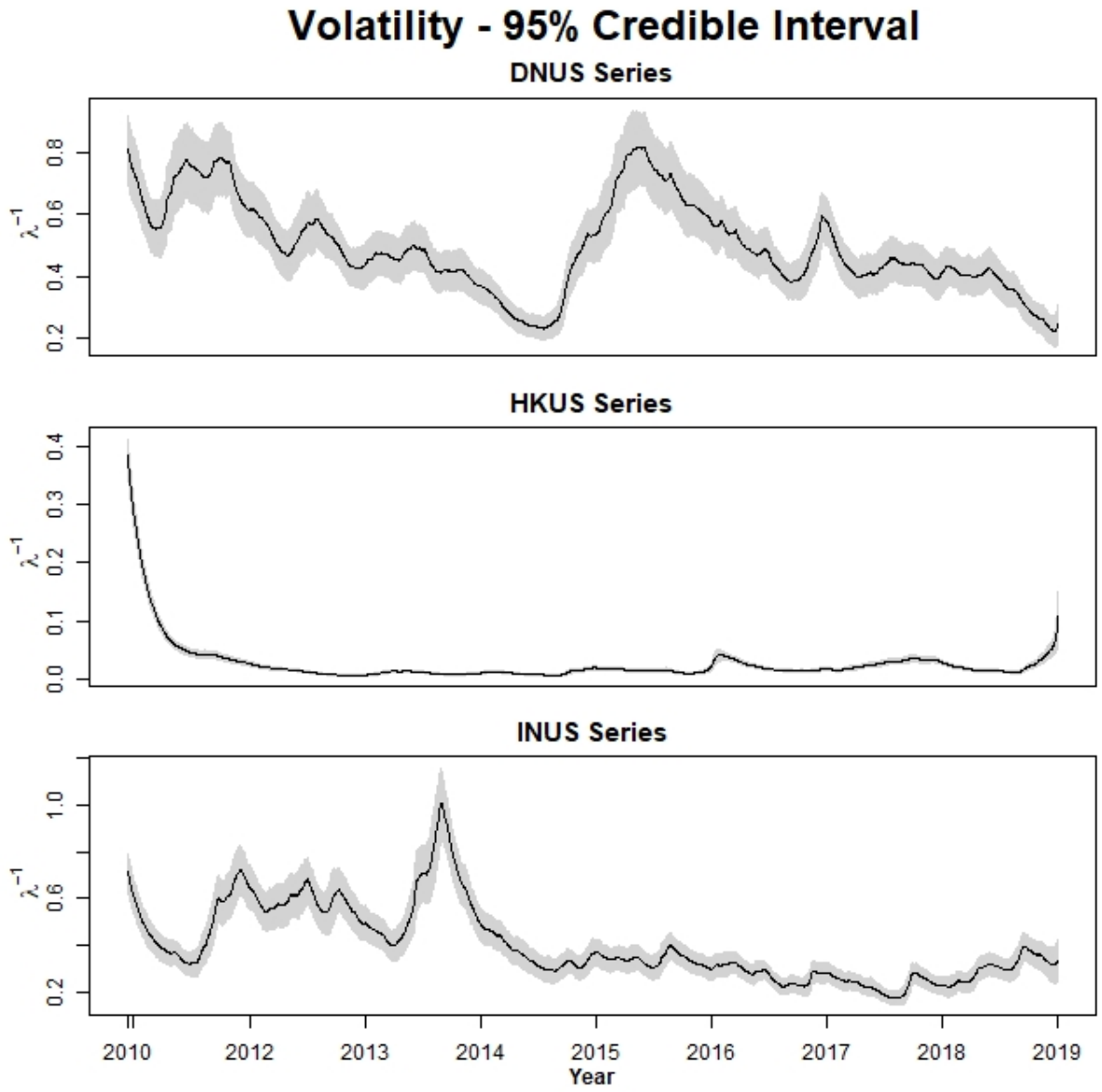


Figure 33 – Posterior mean estimates of instantaneous volatility for Exchange Rates data. Mean estimates in solid line; and 95% credibility interval is the gray area.

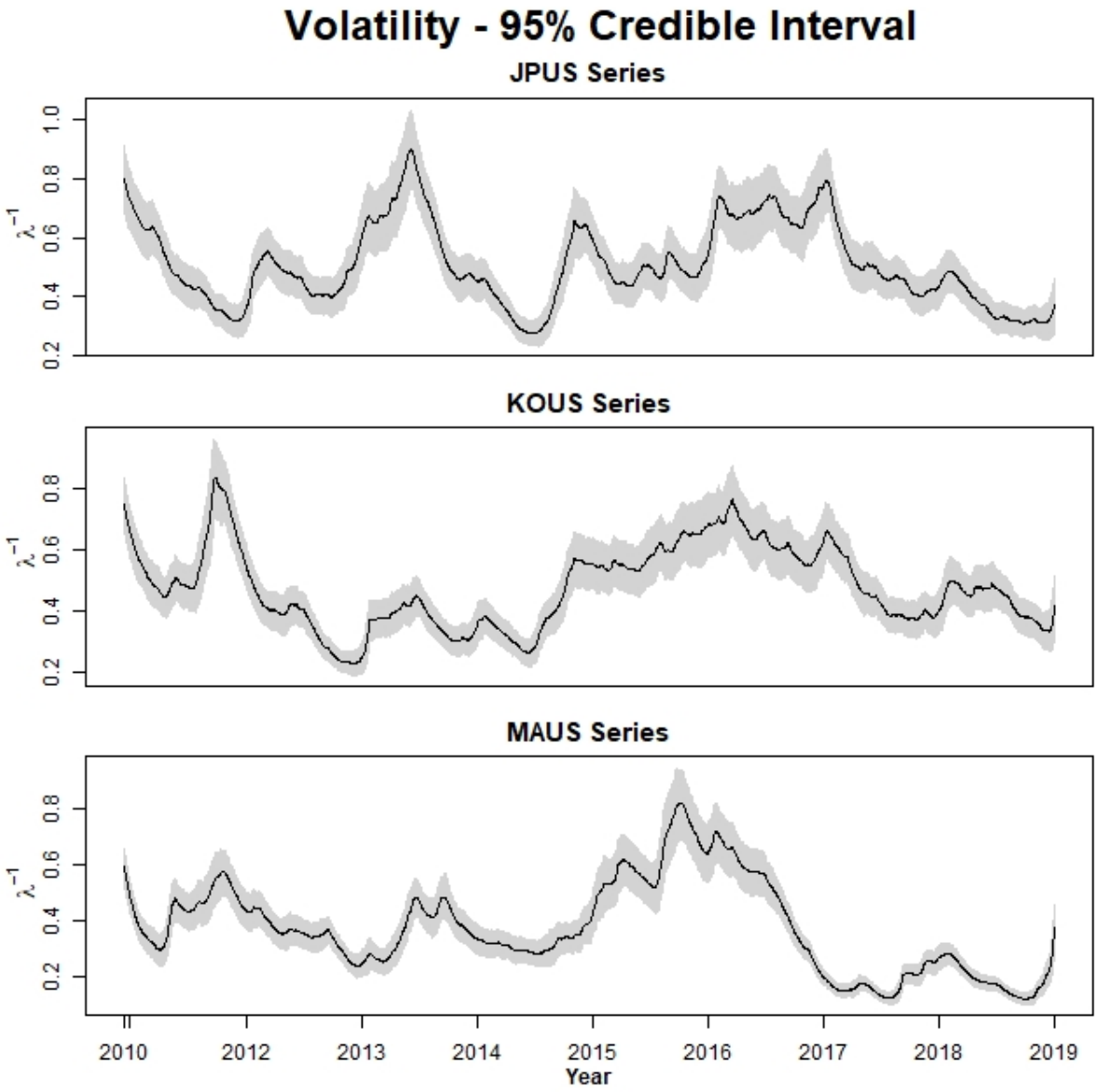


Figure 34 – Posterior mean estimates of instantaneous volatility for Exchange Rates data. Mean estimates in solid line; and 95% credibility interval is the gray area.

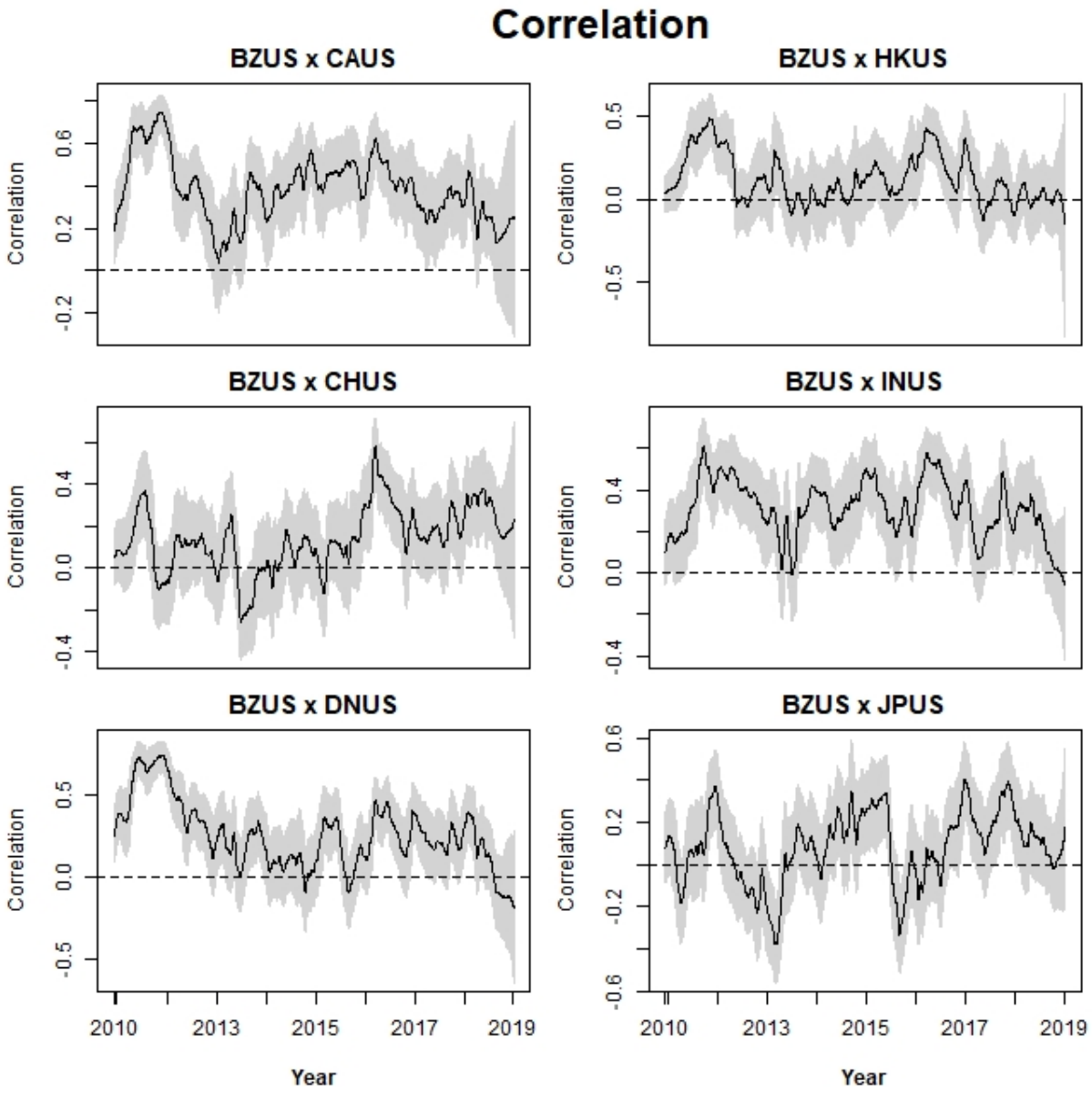


Figure 35 – Posterior mean estimates of instantaneous correlation between Exchange Rates time series. Mean estimates in solid line; and 95% credibility interval is the gray area.

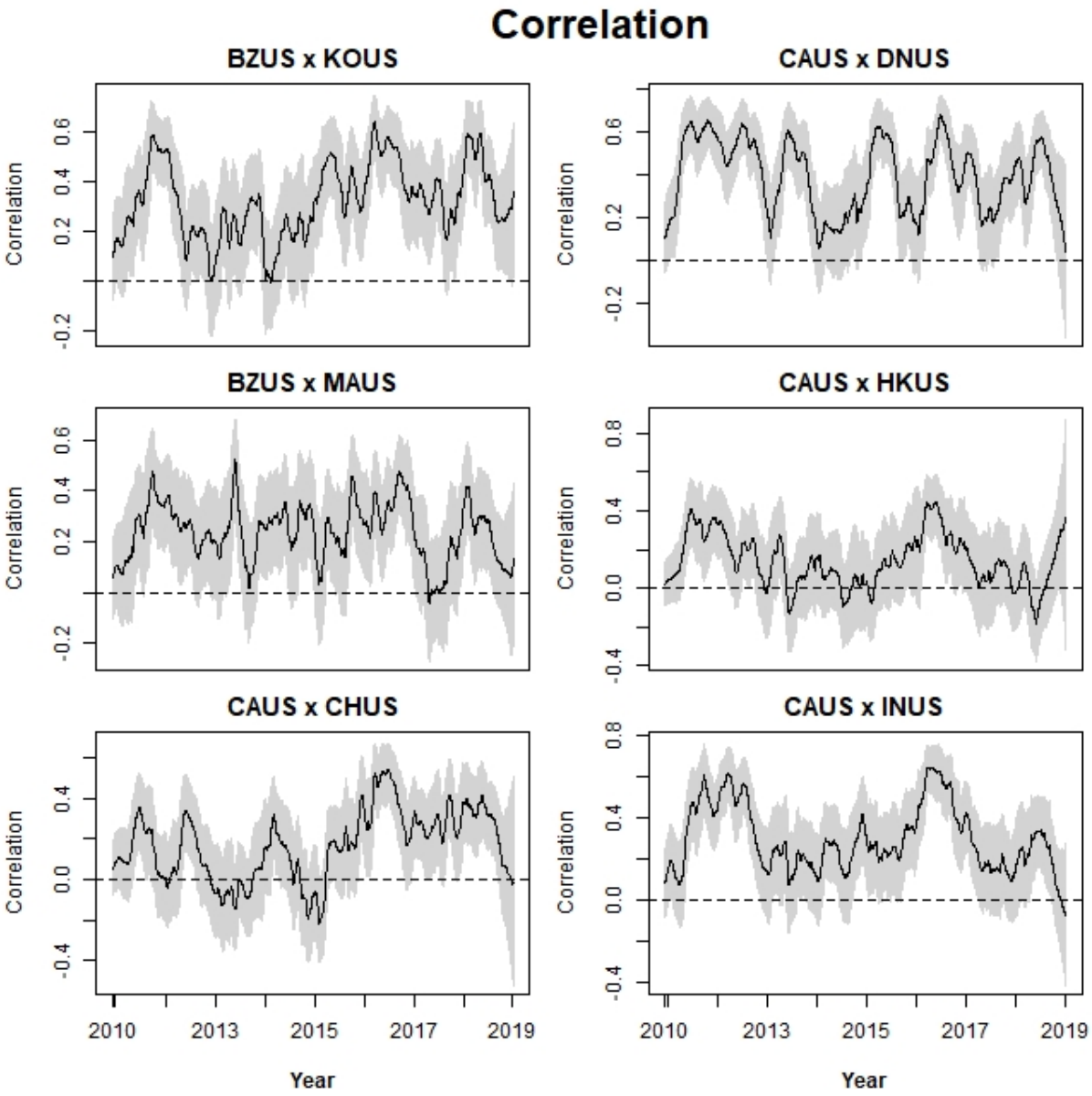


Figure 36 – Posterior mean estimates of instantaneous correlation between Exchange Rates time series. Mean estimates in solid line; and 95% credibility interval is the gray area.

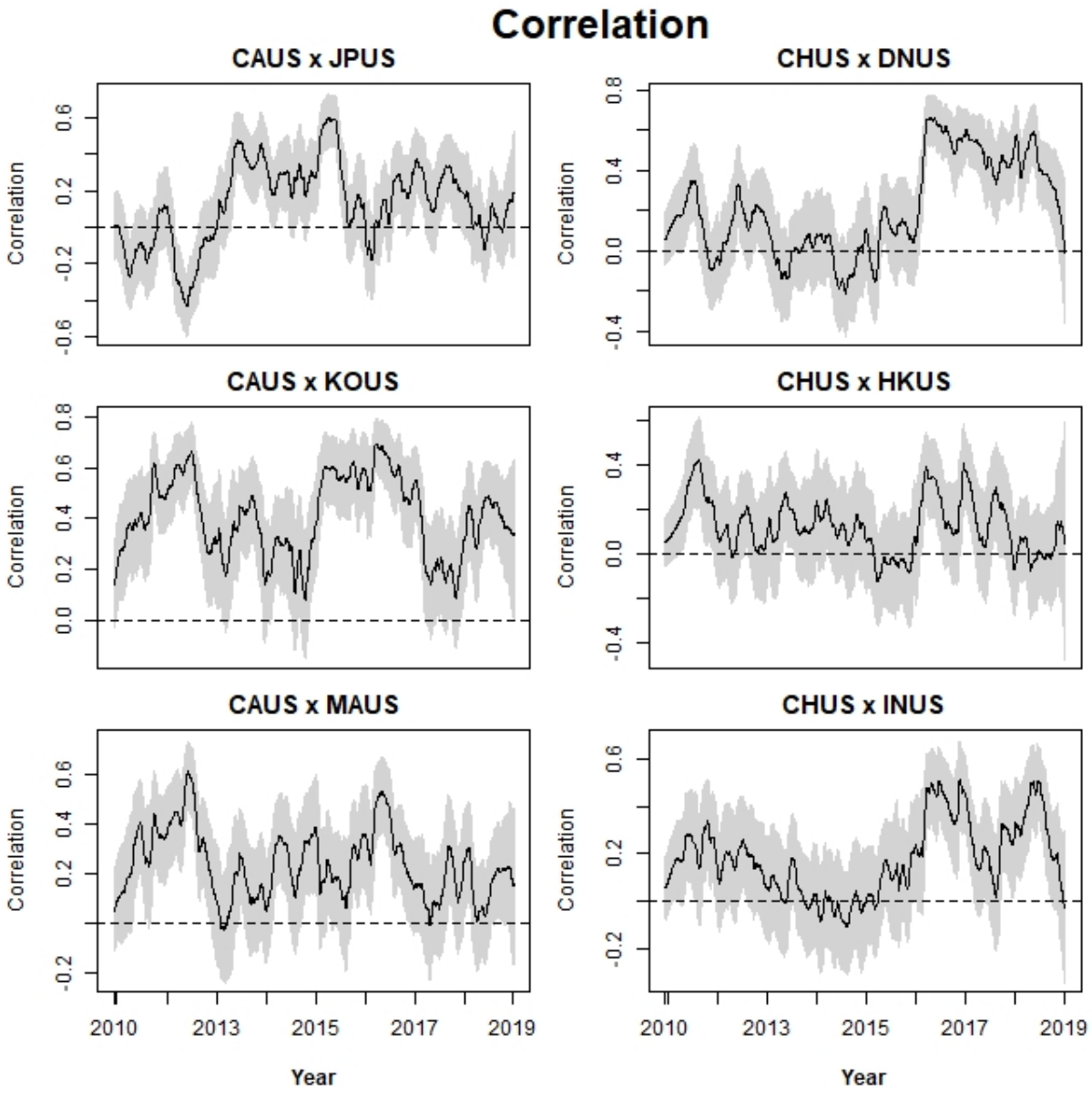


Figure 37 – Posterior mean estimates of instantaneous correlation between Exchange Rates time series. Mean estimates in solid line; and 95% credibility interval is the gray area.

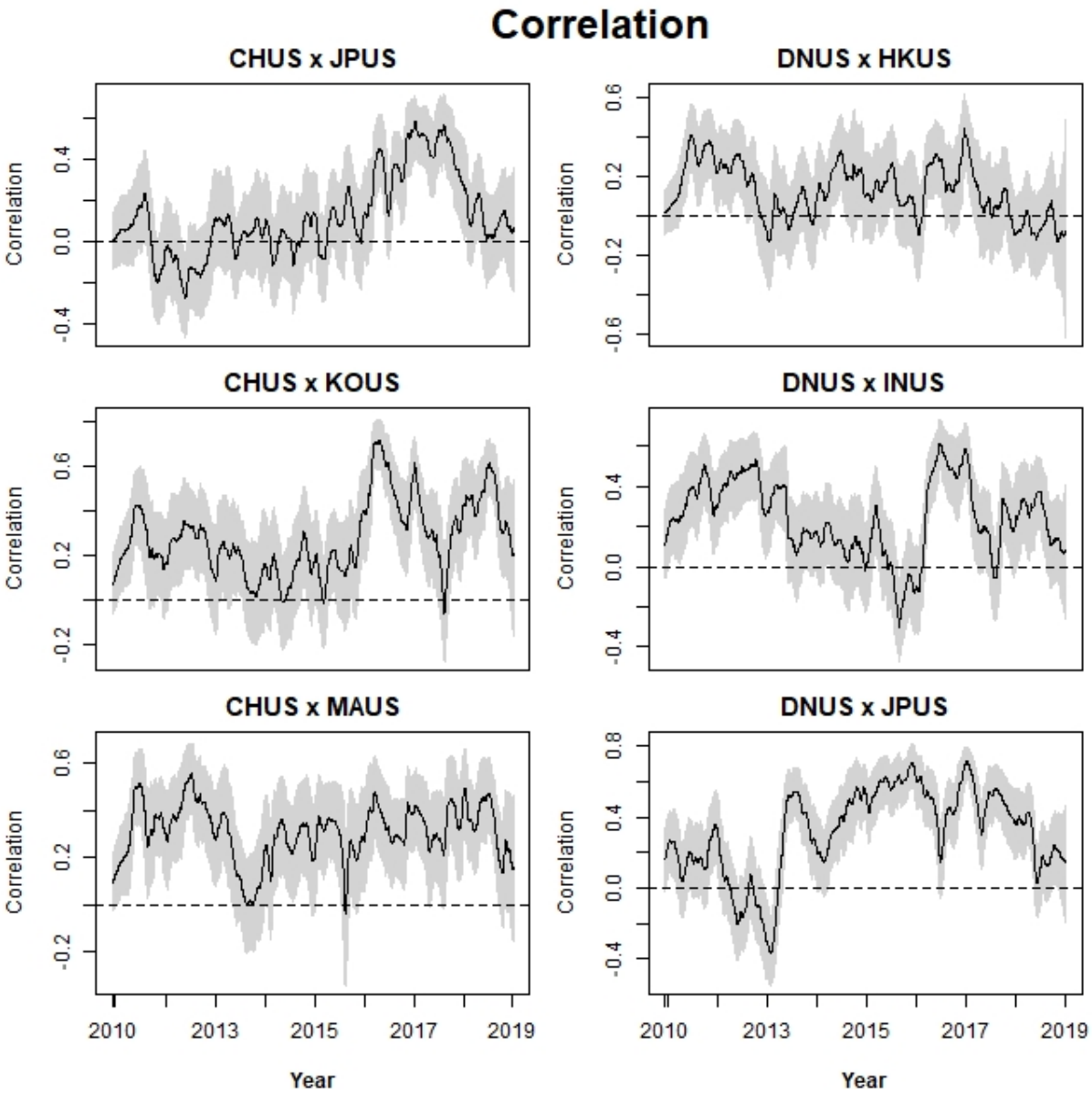


Figure 38 – Posterior mean estimates of instantaneous correlation between Exchange Rates time series. Mean estimates in solid line; and 95% credibility interval is the gray area.

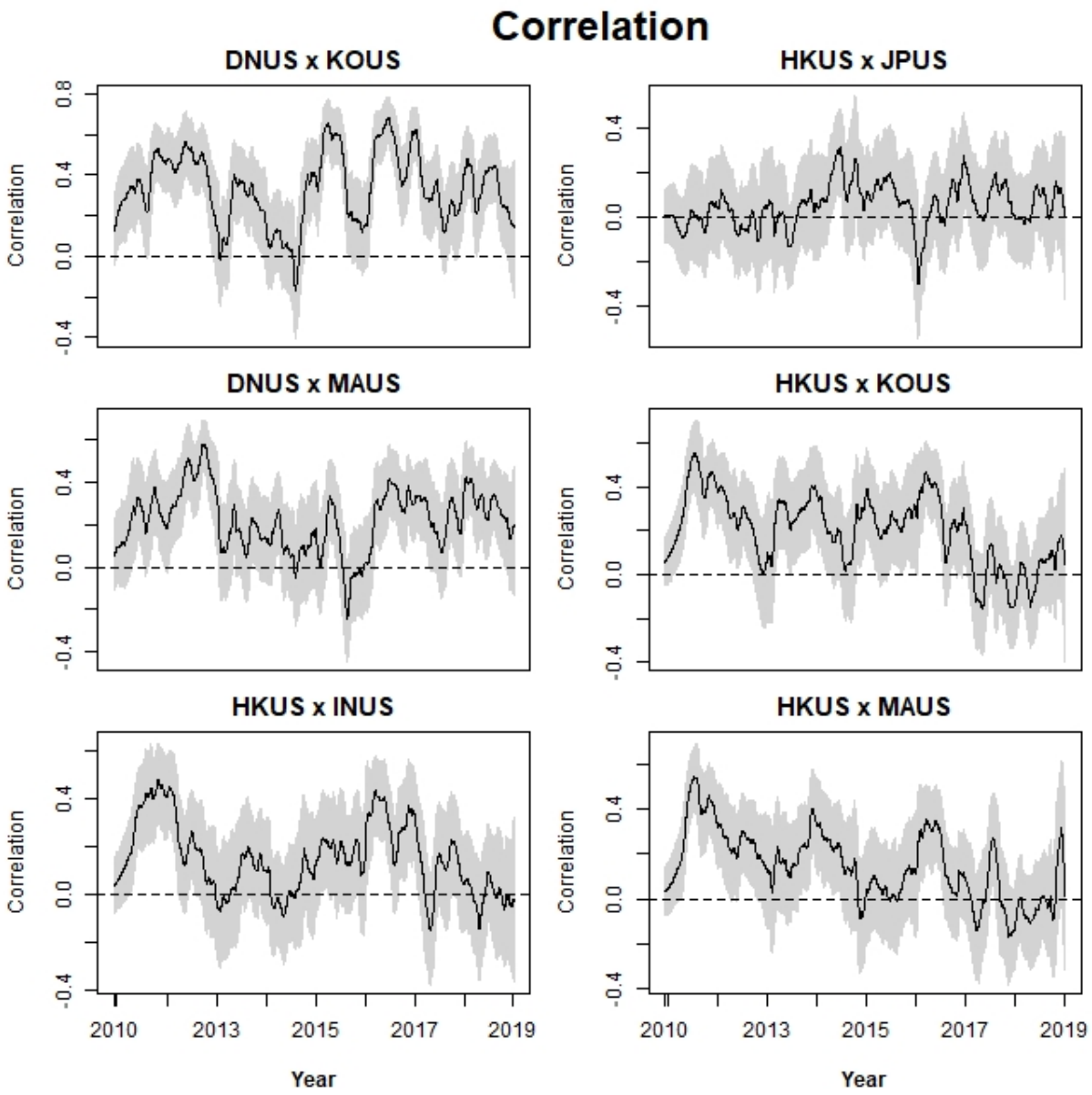


Figure 39 – Posterior mean estimates of instantaneous correlation between Exchange Rates time series. Mean estimates in solid line; and 95% credibility interval is the gray area.

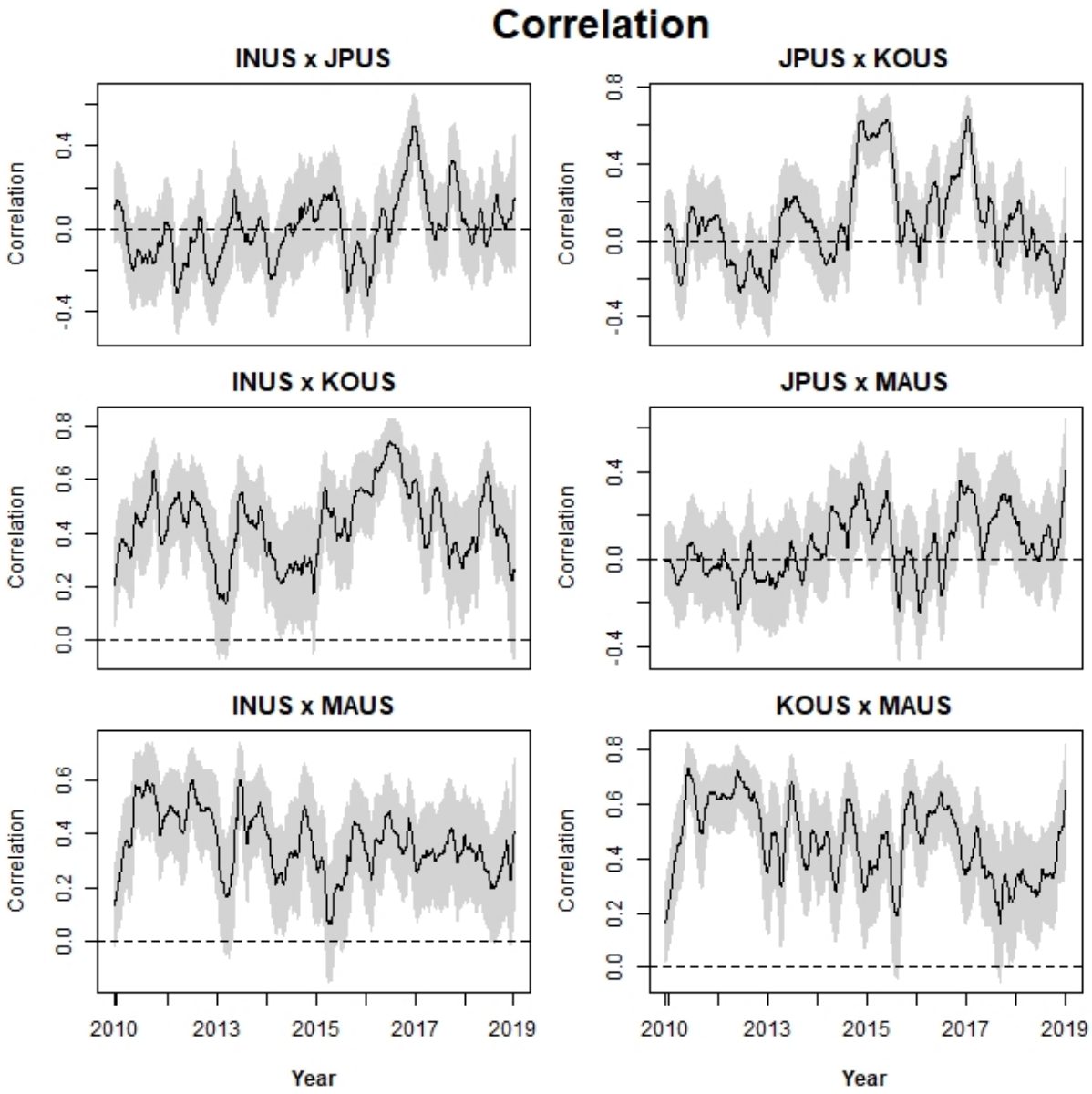


Figure 40 – Posterior mean estimates of instantaneous correlation between Exchange Rates time series. Mean estimates in solid line; and 95% credibility interval is the gray area.

Figures 41, 42 and 43 provides jump sizes for each observation. Jumps usually precede moments of higher volatility, as they are a way of capturing speculative movements at the market. As the higher the number of jumps, more speculative the currency can be considered, thus more vulnerable to external shocks it is. HKUS presents only six jumps on the period of analysis, as expected by the characteristics of this particular currency mentioned before.

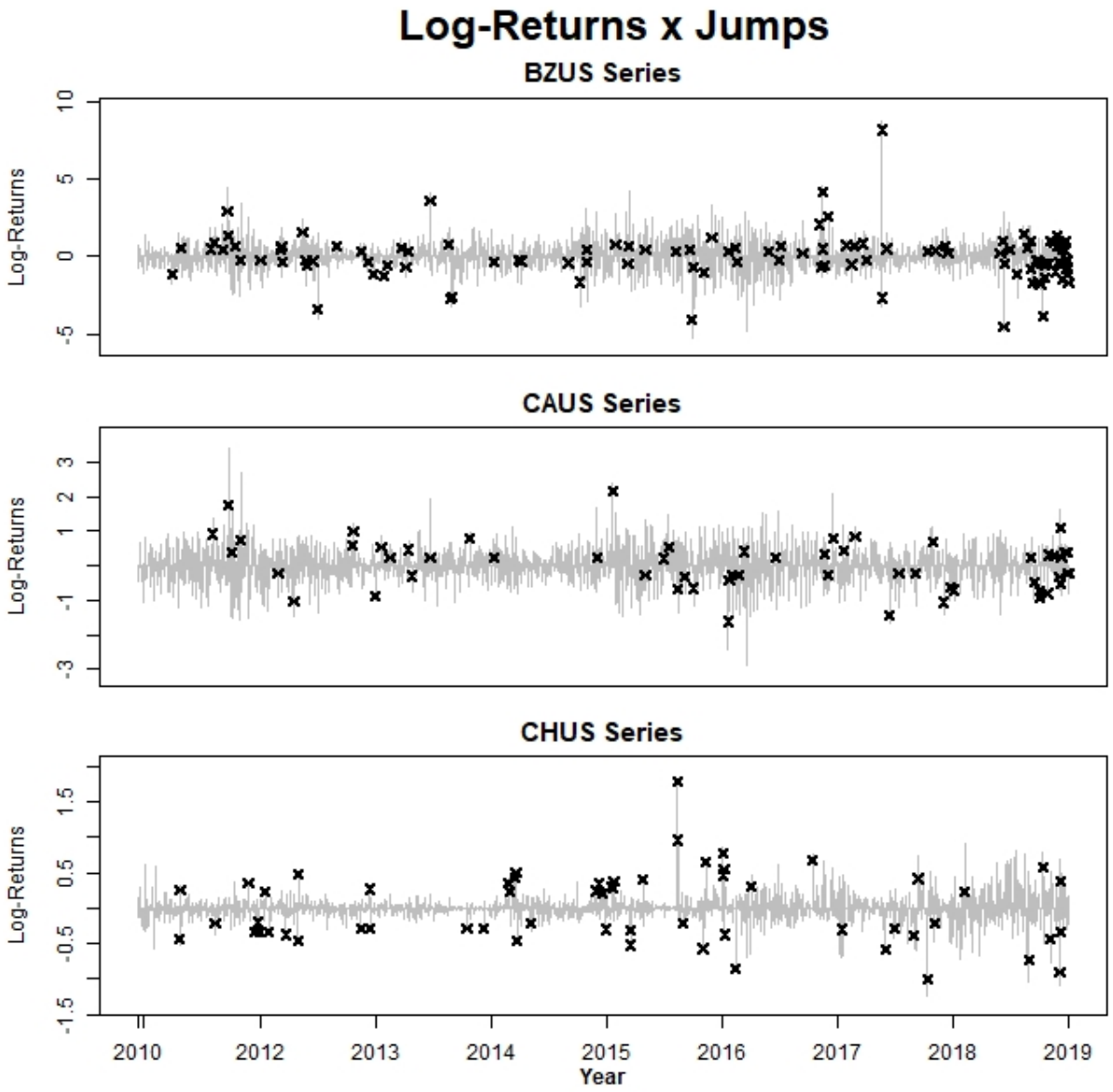


Figure 41 – Posterior mean estimates of instantaneous jumps J_t. Jumps are represented by black crosses and log-returns the light gray lines.

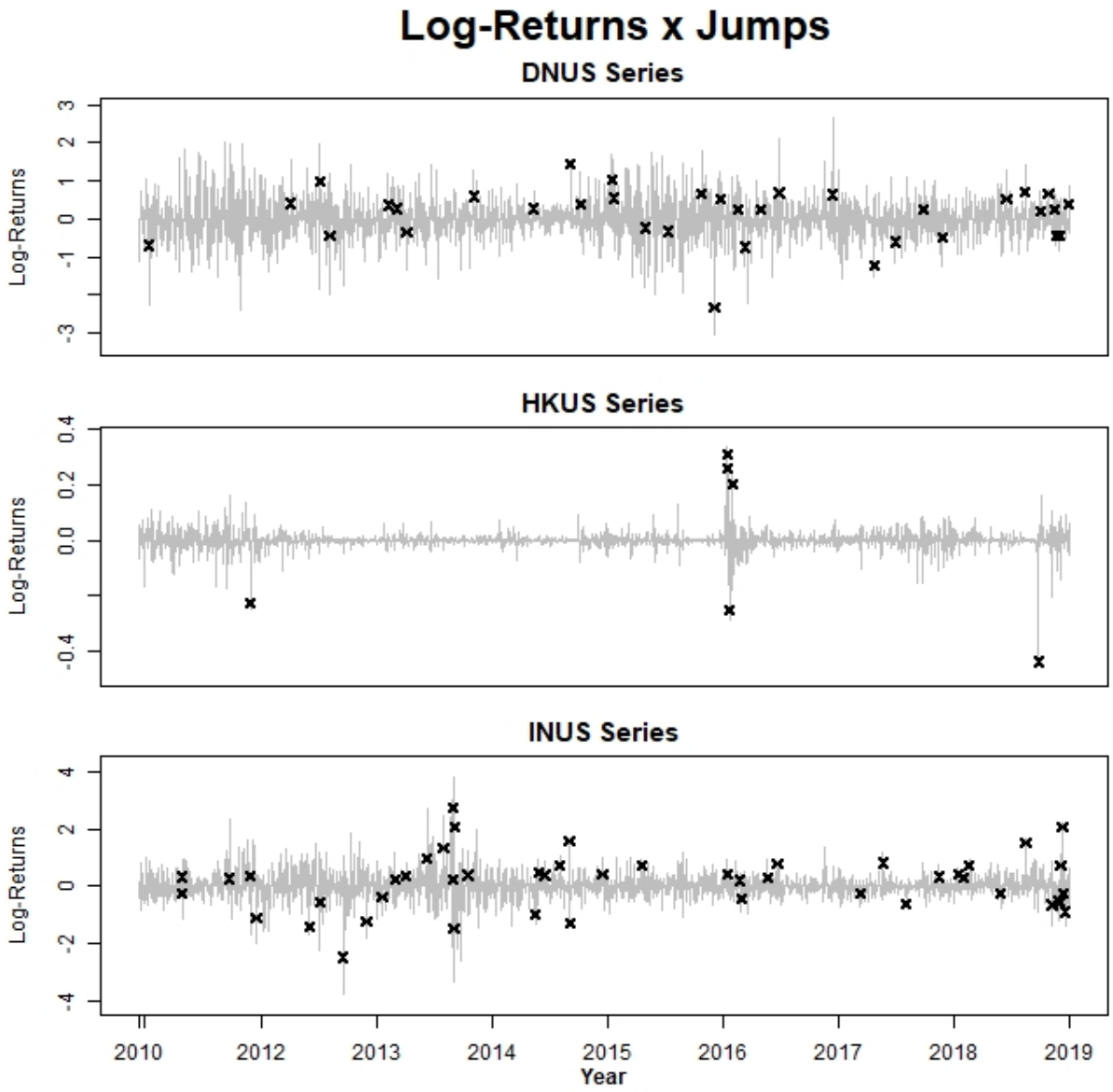


Figure 42 – Posterior mean estimates of instantaneous jumps J_t. Jumps are represented by black crosses and log-returns the light gray lines.

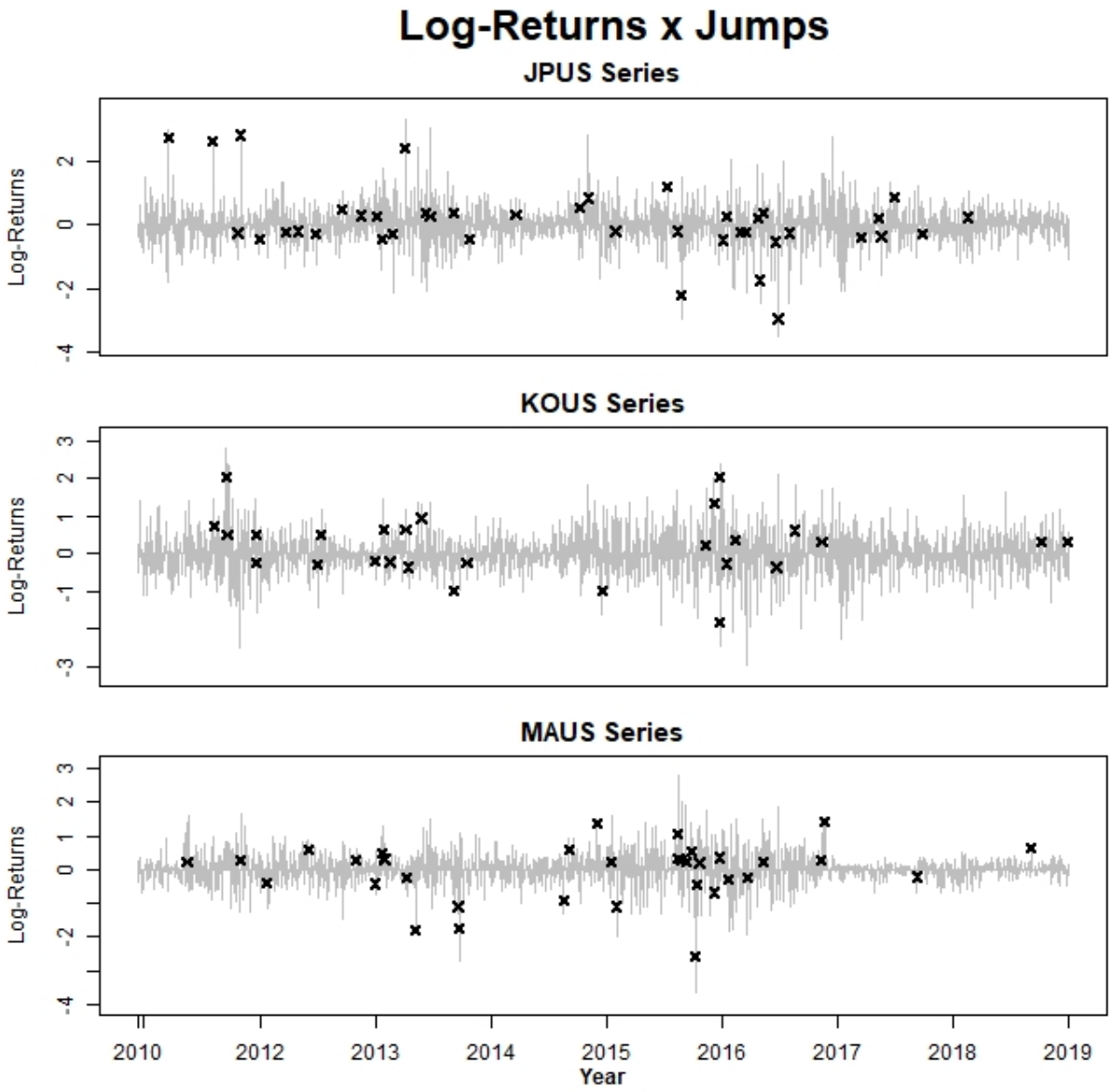


Figure 43 – Posterior mean estimates of instantaneous jumps J_t. Jumps are represented by black crosses and log-returns the light gray lines.

Table 18 shows model estimates for each of the static parameters. For most currencies, jumps are infrequent but have a large magnitude, as can be seen by the large standard deviation. The computational time was 87.44988 minutes.

	Mean	SD		Mean	SD		Mean	SD
μ_{y_1}	-0.0443	0.3380	μ_{y_2}	0.0465	0.2041	μ_{y_3}	-0.0006	0.0519
μ_{y_4}	0.1850	0.2759	μ_{y_5}	-0.0084	0.0127	μ_{y_6}	0.1985	0.2556
μ_{y_7}	0.1985	0.4266	μ_{y_8}	0.3214	0.3655	μ_{y_9}	-0.1586	0.3150
σ_{y_1}	2.3321	0.2853	σ_{y_2}	1.1035	0.1453	σ_{y_3}	0.4639	0.0411
σ_{y_4}	1.1814	0.2099	σ_{y_5}	0.1163	0.0106	σ_{y_6}	1.3881	0.2106
σ_{y_7}	1.8184	0.3136	σ_{y_8}	1.3599	0.2762	σ_{y_9}	1.3398	0.2476
ρ_{y_1}	0.0349	0.0085	ρ_{y_2}	0.0233	0.0052	ρ_{y_3}	0.0482	0.0068
ρ_{y_4}	0.0159	0.0046	ρ_{y_5}	0.0472	0.0072	ρ_{y_6}	0.0208	0.0045
ρ_{y_7}	0.0140	0.0038	ρ_{y_8}	0.0117	0.0033	ρ_{y_9}	0.0135	0.0036
log L	647		BIC	-1,233		AICc	-1,278	

Table 18 – Posterior inference of static parameters for MSVJM model for currency exchange rates daily log-returns.

6 Multivariate Stochastic Model with Jumps and Beta-Bartlett Evolution

According to Prado and West, the Beta-Bartlett evolution for the inverse of covariance matrix, $\Omega_t = \Sigma_t^{-1}$, is based on Markov evolution distributions $p(\Omega_t|\Omega_{t-1})$ in which random innovations are applied to elements of the Bartlett decomposition of Ω_{t-1}. The Bartlett decomposition is detailed on Appendix C. Under the Wishart distribution, for all information available until time $t-1$, represented by $\mathcal{D}_{t-1}$, $(\Omega_{t-1}|\mathcal{D}_{t-1}) \sim W(h_{t-1}, D_{t-1}^{-1})$ at time $t-1$, let P_{t-1} be the upper triangular Cholesky component of D_{t-1}^{-1}, so that the Bartlett decomposition is:

$$\Omega_{t-1} \quad = \quad P_{t-1}^T U_{t-1}^T U_{t-1} P_{t-1}, \tag{6.1}$$

where U_{t-1} is a upper triangular matrix of random elements $u_{t-1,i,j}$ such that:

1. $u_{t-1,i,i}^2 \sim G\left(\frac{h_{t-1}-i+1}{2}, \frac{1}{2}\right)$, for $i = 1:q$;
2. above the diagonal, $u_{t-1,i,j} \sim N(0,1)$, for $j = (i+1):q$ and $i = 1:q$;
3. the $u_{t-1,i,j}$ are mutually independent.

According to Quintana, Lourdes, Aguilar, and Liu (2003), the Markovian evolution becomes:

$$\Omega_t \quad = \quad \beta^{-1} P_{t-1}^T \widetilde{U}_t^T \widetilde{U}_t P_{t-1}, \tag{6.2}$$

for a discount factor $\beta \in (0,1)$ and where the new random matrix $\widetilde{U}_t$ is constructed from U_{t-1} using univariate beta-gamma stochastic volatility models, as it was done in the MSVJM. $\widetilde{U}_t$ is upper triangular, has the same upper off-diagonal elements $u_{t-1,i,j}$ as U_{t-1}, but has diagonal elements $\widetilde{u}_{t,i,j} = u_{t-1,i,j}\gamma_t^{1/2}$, where $\gamma_{t,i}$ are independent beta random quantiles, all independent of the off-diagonal $u_{t-1,i,j}$. For $i = 1:q$:

$$(\gamma_{t,i}|\mathcal{D}_{t-1}) \quad \sim \quad Beta\left(\frac{\beta_{t,i}(h_{t-1}-i+1)}{2}, \frac{(1-\beta_{t,i})(h_{t-1}-i+1)}{2}\right), \tag{6.3}$$

where, for each i, $\beta_{t,i} = (\beta h_{t-1} - i + 1)/(h_{t-1} - i + 1)$. It follows that

$$(\widetilde{u}_{t,i,j}^2|\mathcal{D}_{t-1}) \quad \sim \quad G\left(\frac{\beta h_{t-1}-i+1}{2}, \frac{1}{2}\right). \tag{6.4}$$

The result of this construction is that the evolution in Eq. (6.2) implies:

$$(\Omega_t|\mathcal{D}_{t-1}) \quad \sim \quad W(\beta h_{t-1}, (\beta D_{t-1})^{-1}). \tag{6.5}$$

that is, the same discounting structure of the Matrix-Beta model, but relaxing the constrains on β, so it can assume any value $\beta \in (0, 1)$.

Another construction cited by Prado and West (2010) relaxes the dimensionality supported by the model, without appealing to generating from singular Wisharts, that is, any degrees of freedom can be generated for the evolved precision or covariance matrix from the Beta-Bartlett model. In this case, the Eq (6.2) is modified to:

$$\Omega_t \quad = \quad b_t^{-1} P_{t-1}^T \tilde{U}_t^T \tilde{U}_t P_{t-1}, \tag{6.6}$$

for a constant b_t to be chosen, and with the beta distributions of the shocks having $\beta_{t,i} = (\beta n_{t-1} + q - i)/(n_{t-1} + q - i)$, for $i = 1 : q$ and $n_{t-1} = h_{t-1} - q + 1$. With this modification in the evolution equation, Eq.(6.5), the $W(n_{t-1} + q - 1, D_{t-1}^{-1})$ distribution for Ω_{t-1} evolves to:

$$(\Omega_t | \mathcal{D}_{t-1}) \quad \sim \quad W(\beta n_{t-1} + q - 1, (b_t D_{t-1})^{-1}) \tag{6.7}$$

and by taking $b_t = (\beta n_{t-1} + q - 1)/(n_{t-1} + q - 1)$ ensures that the mean of the precision matrix is unchanged trough the evolution, being fixed at $(n_{t-1}+q-1)D_{t-1}^{-1} = h_{t-1}D_{t-1}^{-1}$, but now it is possible to generate from from Wishart distribution, without any dimensionality issues on degrees of freedom parameter, for assets portfolio.

However, despite all advantages above mentioned, Prado and West (2010) states that there is no retrospective analysis procedure developed for the Beta-Bartlett evolution structure.

6.1 Retrospective Analysis Procedure

The main challenge on developing a retrospective analysis procedure for the Beta-Bartlet evolution is that there is no closed-form for the distribution of $\Omega_{t-1}|\Omega_t$, once the evolution, in this case, is more complex than the Matrix-Beta evolution, to avoid sampling the precision matrix directly from a Wishart distribution, thus avoiding issues related to data dimensionality. However, a smoothed sample can be obtained using the smoothing procedure presented bellow, taking advantage of Wishart distribution properties to build an effective algorithm for the retrospective analysis.

According to Leondes *et. al.* (1970), the smoothing estimation problem is concerned with finding the best estimate of the value of a time-varying parameter, using information from related observations taken over a period of time before or after the instant of analysis t. Smoothing techniques are especially useful when the estimated parameter is noisy, in order to remove the noisiness of estimation, leading to more precise estimates.

According to Hunter (1986), the exponentially weighted moving average (EWMA) technique, originated on the early work of econometricians and spread to different areas,

including engineering, as an effective tool for smoothing. Hunter (1986) states that the EWMA is a statistic with the characteristic that it gives less and less weight to data as they get older and older. Hawkins and Maboudou-Tchao (2008), Harris and Yilmaz (2010) and Shen , Tsung, and Zou (2014) extends applications of the EWMA technique on smoothing time-varying covariance matrixes.

According to Hawkins and Maboudou-Tchao (2008), multivariate processes measurements benefit from the use of inherently multivariate methods rather than a collection of univariate methods applied to each of its components, and present the Multivariate Exponentially Weighted Moving Covariance Matrix (MEWMC), which is an extension of the EWMA for covariance matrixes. Harris and Yilmaz (2010) apply the EWMA for a covariance matrix on a financial markets approach on smoothing the covariance matrix for exchange rates returns.

The EWMA technique can also be used for smoothing, as in Zhou and Lawson (2008) and Prado and West (2010) uses the exponentially weighted moving average method for smoothing moments to build a retrospective simulation method for sampling from states parameters in a space state model. The procedure consists, for a covariance matrix, of using a single decay factor δ_0 so that for a covariance matrix time series, S_t, $t = 1, \ldots, n$ we have that:

$$S_{S_t} = \delta_0 \, S_{S_{t+1}} + (1 - \delta_0) \, S_t \tag{6.8}$$

or equivalently, using the element wise notation for $i = 1, \ldots, q$ and $j = 1, \ldots, q$, as in Harris and Yilmaz (2010):

$$s_{S_{i,j,t}} = \delta_0 \, s_{S_{i,j,t+1}} + (1 - \delta_0) \, s_{i,j,t} \tag{6.9}$$

where the underscript S denotes the smoothed parameter and, for convention, we assume $S_{S_n} = S_n$.

The same procedure can be applied to the precision matrix Ω_t for obtaining a punctual estimative and the respective credibility interval for the smoothed expectation for Ω_t using the Beta-Bartlet evolution, so we have that:

$$\Omega_{S_t} = \beta \, \Omega_{S_{t+1}} + (1 - \beta) \, \Omega_t \tag{6.10}$$

where $\beta \in (0, 1)$ is the same discount factor in Eq.(6.5), therefore, the same discount factor used in Matrix-Beta evolution. By adopting this procedure, the smoothed sampled obtained has the same expected value from that obtained in the procedure described in Appendix B.2, applied on MSVJM. The proof follows directly from Wishart distribution properties, as described bellow.

Based on Wishart distribution properties available in Eaton (1983), for $\Omega \sim W(h, D^{-1})$, we have that $E(\Omega) = h \times D^{-1}$, and for $M \sim W(h, I)$ then $\Omega \overset{\mathrm{D}}{=} P^T M P$, so

that P is a decomposition of D^{-1} such that $P^T P = D^{-1}$. So we have that, for any given time t, the expected value of $\Omega_{S,t}$ in Eq. (6.10), $E(\Omega_{S,t})$, is given by:

$$\begin{aligned} E(\Omega_{S_t}|\mathcal{D}_n) &= \beta\, E(\Omega_{S_{t+1}}) + (1-\beta)\, E(\Omega_t) && (6.11)\\ &= \beta\, E(\beta\Omega_{S_{t+2}} + (1-\beta)\Omega_{t+1}) + (1-\beta) \times h_t \times D_t^{-1} && (6.12)\\ &= \beta^2\, E(\Omega_{S_{t+2}}) + \beta(1-\beta) \times h_{t+1} \times D_{t+1}^{-1} && (6.13)\\ &\quad +(1-\beta) \times h_t \times D_t^{-1} && (6.14)\\ &= \beta^{n-t} \times h_n \times D_n^{-1} + \sum_{i=0}^{n-t-1} \beta^i (1-\beta) \times h_{t+i} \times D_{t+i}^{-1}. && (6.15) \end{aligned}$$

Note that, directly from Wishart property mentioned before, it follows that for a positive constant $(1-\beta)$ and a Wishart distributed random variable $\Upsilon \sim W((1-\beta)h, D^{-1})$, the expected value of Υ is given by $E(\Upsilon) = (1-\beta) \times h \times D^{-1}$. The smoothed sample obtained trough the procedure available in Appendix B.2 has, for any given time t, the expected value $E(\Omega_{S_t})$ given by:

$$\begin{aligned} E(\Omega_{S_t}|\mathcal{D}_n) &= \beta\, E(\Omega_{S_{t+1}}) + E(\Upsilon_t) && (6.16)\\ &= \beta\, E(\beta\Omega_{S_{t+2}} + \Upsilon_{t+1}) + (1-\beta) \times h_t \times D_t^{-1} && (6.17)\\ &= \beta^2\, E(\Omega_{S_{t+2}}) + \beta(1-\beta) \times h_{t+1} \times D_{t+1}^{-1} && (6.18)\\ &\quad +(1-\beta) \times h_t \times D_t^{-1} && (6.19)\\ &= \beta^{n-t} \times h_n \times D_n^{-1} + \sum_{i=0}^{n-t-1} \beta^i (1-\beta) \times h_{t+i} \times D_{t+i}^{-1}. && (6.20) \end{aligned}$$

The proposed method for covariance matrix smoothing the in the Beta-Bartlett evolution model consists of using the EWMA as in Eq. (6.10), using $\beta \in (0,1)$ is a smoothing constant used to tune the procedure to different sizes of change, with a small value of β being used for detecting small shifts, whereas large values of β are used if large shifts are interested, as in Hawkins and Maboudou-Tchao (2008).

With the proposed procedure it is possible to obtain a smoothed estimative for the expectation of Ω_t, since a closed form for the smoothed distribution of Ω_t is not available using the Beta-Bartlett approach. Since the Beta-Bartlett structure is more flexible on dealing with high dimensional portfolio applications than the Matrix-Beta, it is still a good choice of model despite the lack of a smoothing procedure as in Matrix-Beta, that allows sampling directly from the smoothed distribution. Nevertheless, obtaining estimates for the smoothed average for precision matrix, with respective intervals, is enough for practical purposes in financial applications, so that the usage of the proposed method is justified.

6.2 Simulation Study

To evaluate the performance of the MSVJB, using Beta-Bartlett evolution, and the proposed smoothing procedure, we applied the model to the same synthetic data and model specifications presented in Section 5.4. Results are similar to those found in the previous section, but now the model relaxes its constrains on dimensionality and range of discount factor β, giving extra flexibility to the MSVJ.

Figure 44 shows the time series of true instantaneous volatility $v_{i,t}$, together with the estimated volatility $\sigma_{i,i,t}^{-1}$, for $i = 1, \ldots, 3$ and Figure 45 shows the estimated correlation between the time series. True value ρ_σ is shown in the dashed line. The proposed smoothing procedure was effective in reducing the noise on the estimation of the correlation matrix parameter, Σ_t, while precisely estimating volatility and correlation values. The MSVJ using Beta Bartlett evolution is able to closely track the latent state. Almost every point from the true volatility is inside the 95% credibility interval.

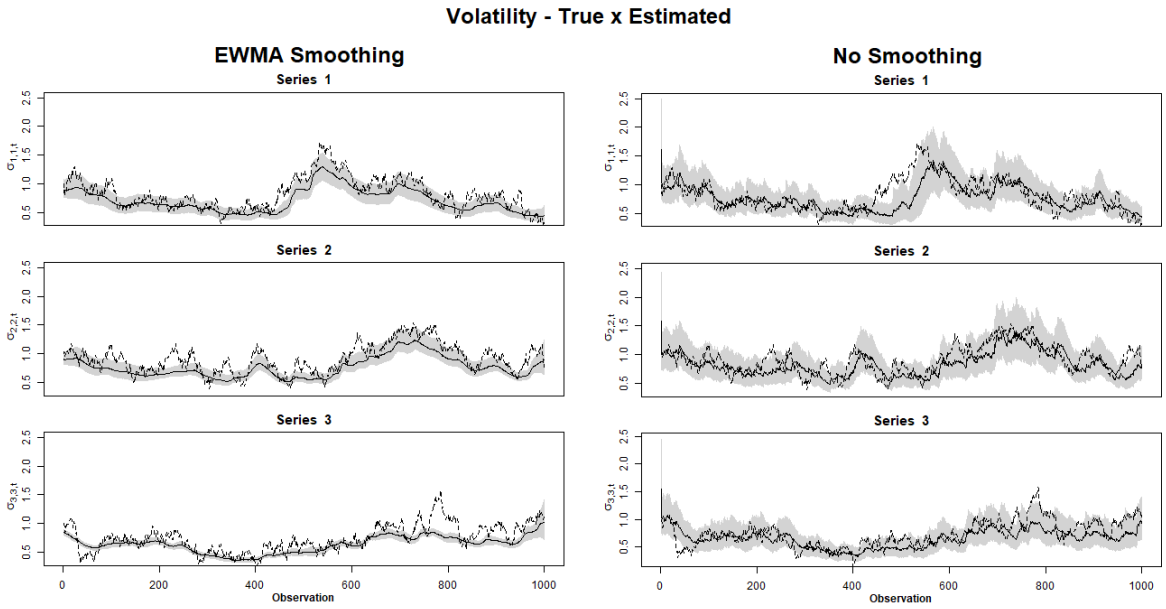


Figure 44 – Simulation study: Posterior mean estimates of instantaneous volatility $v_{i,t}$ for simulated data, using Beta Bartlett evolution for Σ_t. True volatility series shown in dashed line; posterior mean estimates $\sigma_{i,i,t}$ in solid line; and 95% credibility interval is the light gray area.

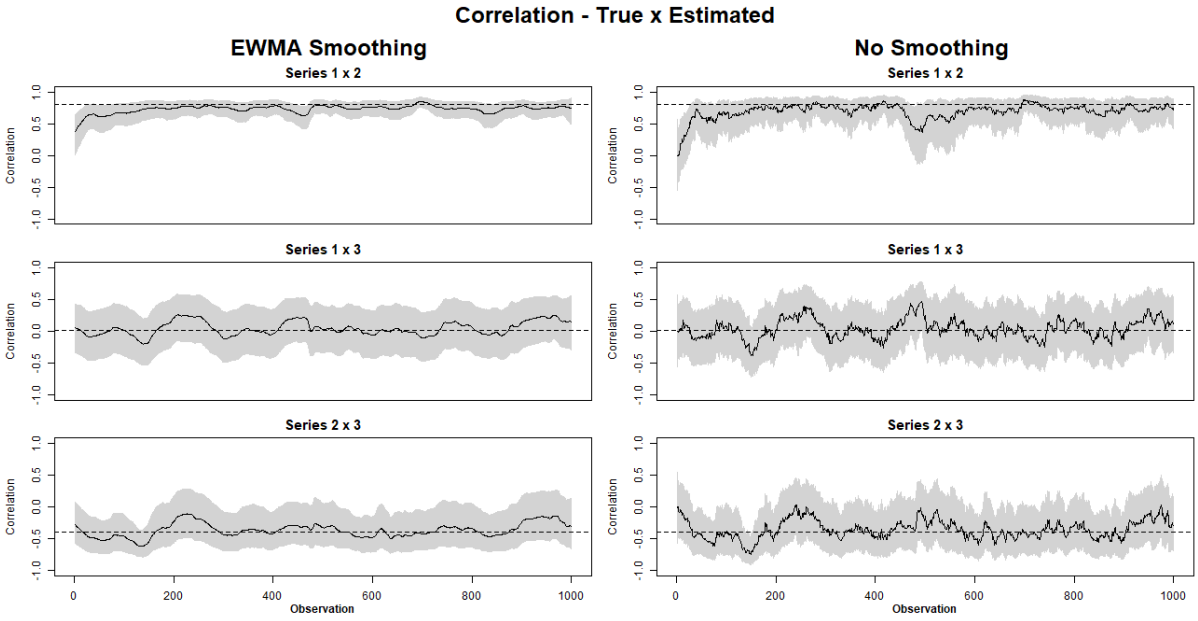


Figure 45 – Simulation study: Posterior mean estimates of correlation coefficient ρ_σ for simulated data, using Beta Bartlett evolution for Σ_t. True value is shown in dashed line; posterior mean estimates in solid line; and 95% credibility interval is the light gray area.

Table 19 presents true values and the summary information for the posterior estimates over the 30 replications using MSVJB. As in MSVJM, the jump related parameters are estimated based only on values recognized by the model as jumps, they tend to have higher variance and root mean square error when compared to true values. When compared to MSVJM results in Table 16 the MSVJB can estimate better jump related parameters, except for the jump probability, that is overestimated. Model selection criteria favors MSVJB over MSVJM as the former has the advantage of sampling from the covariance matrix trough univariate distribution, mitigating numerical issues. Also, the main interest relies on estimating the covariance matrix over time, with statical parameters being used to capture outlier points.

Table 20 shows a comparison between MSVJM, using Matrix-Beta evolution, and MSVJB, using Beta-Bartlett evolution, over a simulation scenario of 1,000 observations. Results are similar for both models, with model selection criteria, pointing a slight advantage towards the MSVJB model. The similarity between both models is expected, since the only difference between MSVJM and MSVJB resides on the method applied for the precision matrix evolution.

Figure 46 shows posterior mean estimates of instantaneous volatility and correlation for simulated data, using MSVJM and MSVJB. For volatility models present similar results, but MSVJB model presents narrower credibility intervals. This phenomena occurs due

	True	MCMC Mean	SD	RMSE
ρ_{y_1}	0.015	0.0607	0.0072	0.0021
ρ_{y_2}	0.020	0.0460	0.0062	0.0007
ρ_{y_3}	0.009	0.0783	0.0119	0.0049
μ_{y_1}	-2.5	-2.9598	1.3203	1.9002
μ_{y_2}	-1.0	-0.9567	0.3007	1.1762
μ_{y_3}	2.0	0.2577	0.3164	0.6480
σ_{y_1}	4.0	4.1544	0.3124	0.1184
σ_{y_2}	2.0	1.7338	0.3212	0.1708
σ_{y_3}	3.0	3.3451	0.2658	0.1876
log L	-		-3047	
BIC	-		**6261**	
AICc	-		**6144**	

Table 19 – Posterior estimates descriptive estatistics of MSVJB mean posterior static parameters for simulated daily returns (n = 1,000), over 30 replications.

to the procedure for sampling in the Beta Bartlett evolution, that samples univariate variables to compose the matrix, instead of sampling the entire matrix as in MSVJM, leading to a smaller computational variability on results. The difference, however, does not affect the results substantially for practical purposes. The estimates for correlation are very similar in both mean and credibility intervals.

	True	MSVJM	MSVJB
ρ_{y_1}	0.04	0.05838 (0.00830)	0.06813 (0.01011)
ρ_{y_2}	0.02	0.03836 (0.00713)	0.04322 (0.00940)
ρ_{y_3}	0.09	0.06753 (0.00899)	0.08119 (0.01518)
μ_{y_1}	5.0	4.1481 (0.7056)	3.5717(0.6680)
μ_{y_2}	-2.0	-1.2437 (0.3397)	-1.1070 (0.3531)
μ_{y_3}	0.5	0.6323 (0.5074)	0.4119 (0.4466)
σ_{y_1}	4.0	4.8262 (0.4831)	4.7248 (0.4288)
σ_{y_2}	2.0	1.7205 (0.2677)	1.7410 (0.2440)
σ_{y_3}	3.0	3.9694 (0.3708)	3.7262 (0.3851)
log L	-	-3196	**-3151**
BIC	-	6557	**6467**
AICc	-	6440	**6351**

Table 20 – Posterior mean estimates of MSVJM and MSVJB static parameters for one simulated daily returns (n = 1,000) scenario. Standard deviations are in parenthesis.

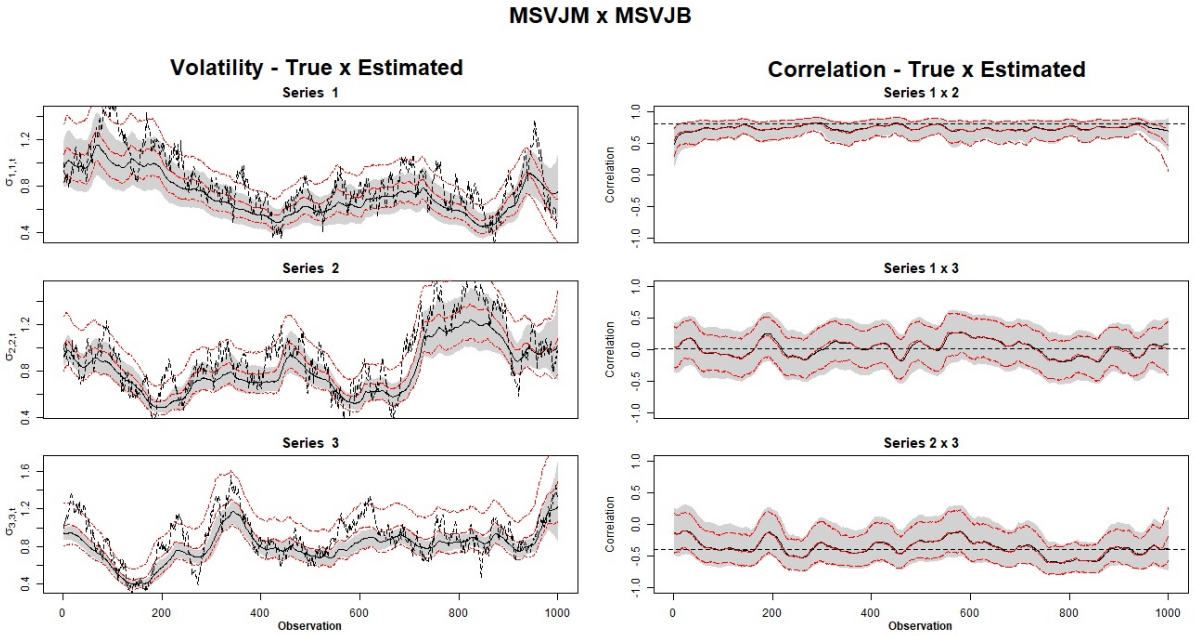


Figure 46 – Simulation study: Posterior mean estimates of instantaneous volatility and correlation for simulated data, using MSVJM and MSVJB. True volatility series shown in dashed line; posterior mean estimates for MSVJB in solid line; and 95% credibility interval is the light gray area; posterior mean estimates for MSVJM in long dashed red line and 95% credibility interval is delimited by the dot dashed red lines.

6.3 Model application

The MSVJB model is applied to the exchange rates dataset presented on Section 1.1.1, on which model MSVJM was applied in Section 5.5.

For a portfolio of $q = 9$ log-return series, the FFBS related parameters setup was $\theta_0|\mathcal{D}_0, \Sigma_0 \sim N_9(0, 100)$ and $\Sigma_0|\mathcal{D}_0, \theta_0$ now follows the Beta-Bartlett evolution structure, using the relaxed discount factor b_t as in Eq. 6.6, with discount parameters $\iota = 0.99$ and $\beta = 0.95$, as suggested in Prado and West (2010). For the mixture component γ_t a prior $G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, with $\nu = 30$, was specified, following Rego and Santos (2020). For jump related components, μ_{y_q}, N(0,100) priors were specified, for $\sigma^2_{y_q}$ a IG(0.1,0.1) prior, an for ρ_q a $Beta(2, 40)$, for $q = 1, \ldots, 9$, as known from previous works in literature and to keep results comparable to MSVJM, using Matriz-Beta evolution, in Section 5.5. The threshold was fixed at $\alpha = 0.7$, following Rego and Santos (2020).

The results were obtained with a 40,000 iteration chain, a burn-in of 20,000 observations, with a lag of 10 observations, resulting in 2,500 samples. MCMC chain convergence was verified through graphical methods. All the coding was done in the R software, using the RcppArmadillo package.

Convergence of MCMC chains was verified trough graphical methods. Residual analysis results does not show evidence of strong violations of the model hypotheses.

Figures 47, 48 and 49 shows the posterior mean estimates of instantaneous standard deviation, or volatility, $diag(\Sigma_t)$, for exchange rates log-returns, using MSVJB model. Figures 50 to 55 shows the correlation between different currencies exchange rates. The MSVJB model results is coherent with the results previously obtained on MSVJM model. As the results are similar for both models.

MSVJB model results present smaller values for volatility estimates since the dimensionality correction by using the b_t discount factor in Eq. 6.6 mitigates sampling issues on Wishart distribution. Volatility for HKUS Series is close to zero, as expected by the government control on the exchange rate. This also affects correlation estimative for this specific time series. For all other estimatives, MSVJB follows the exact same pattern of MSVJM, with slightly smaller values for volatility.

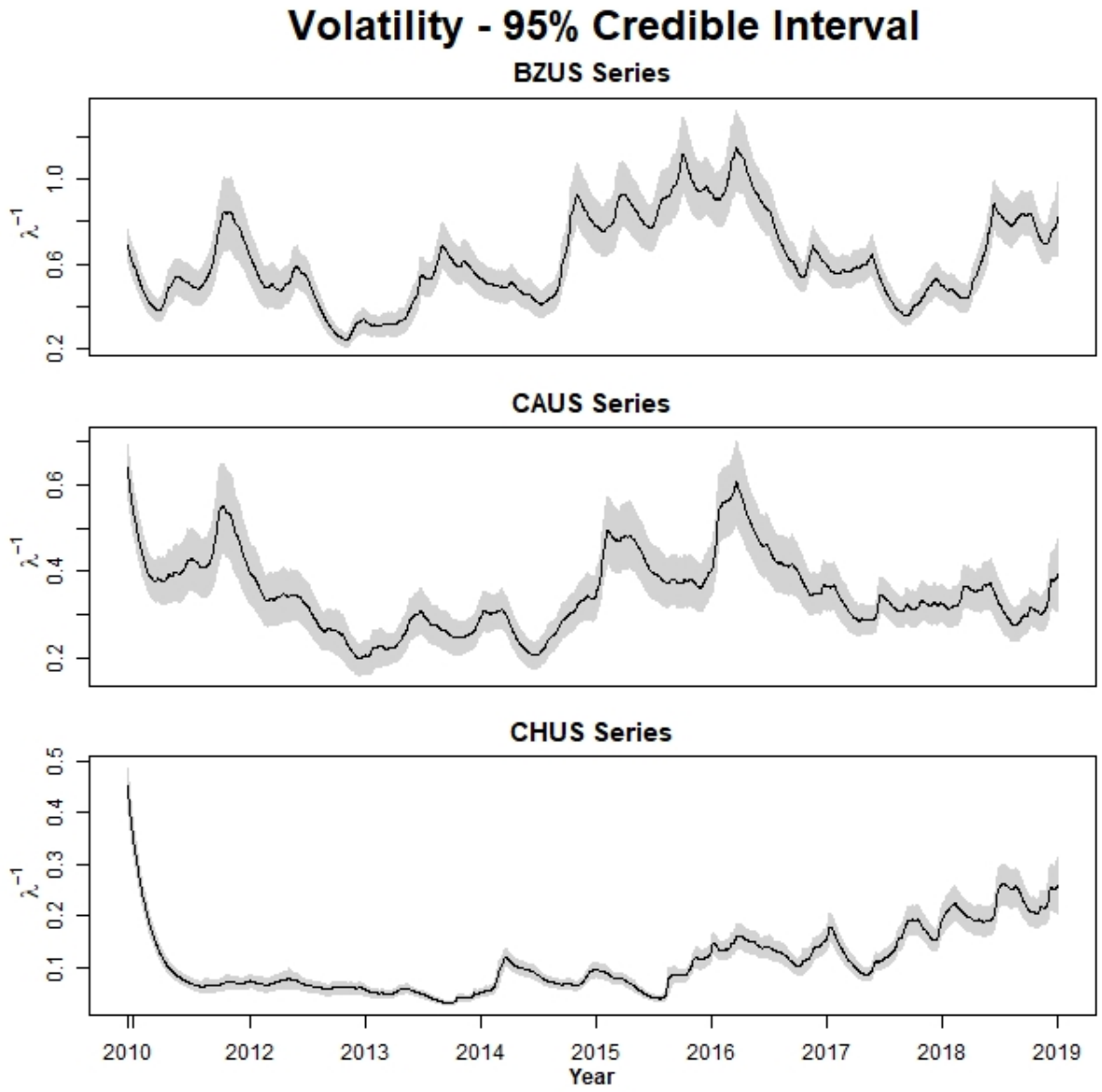


Figure 47 – Posterior mean estimates of instantaneous volatility for Exchange Rates data. Mean estimates in solid line; and 95% credibility interval is the gray area.

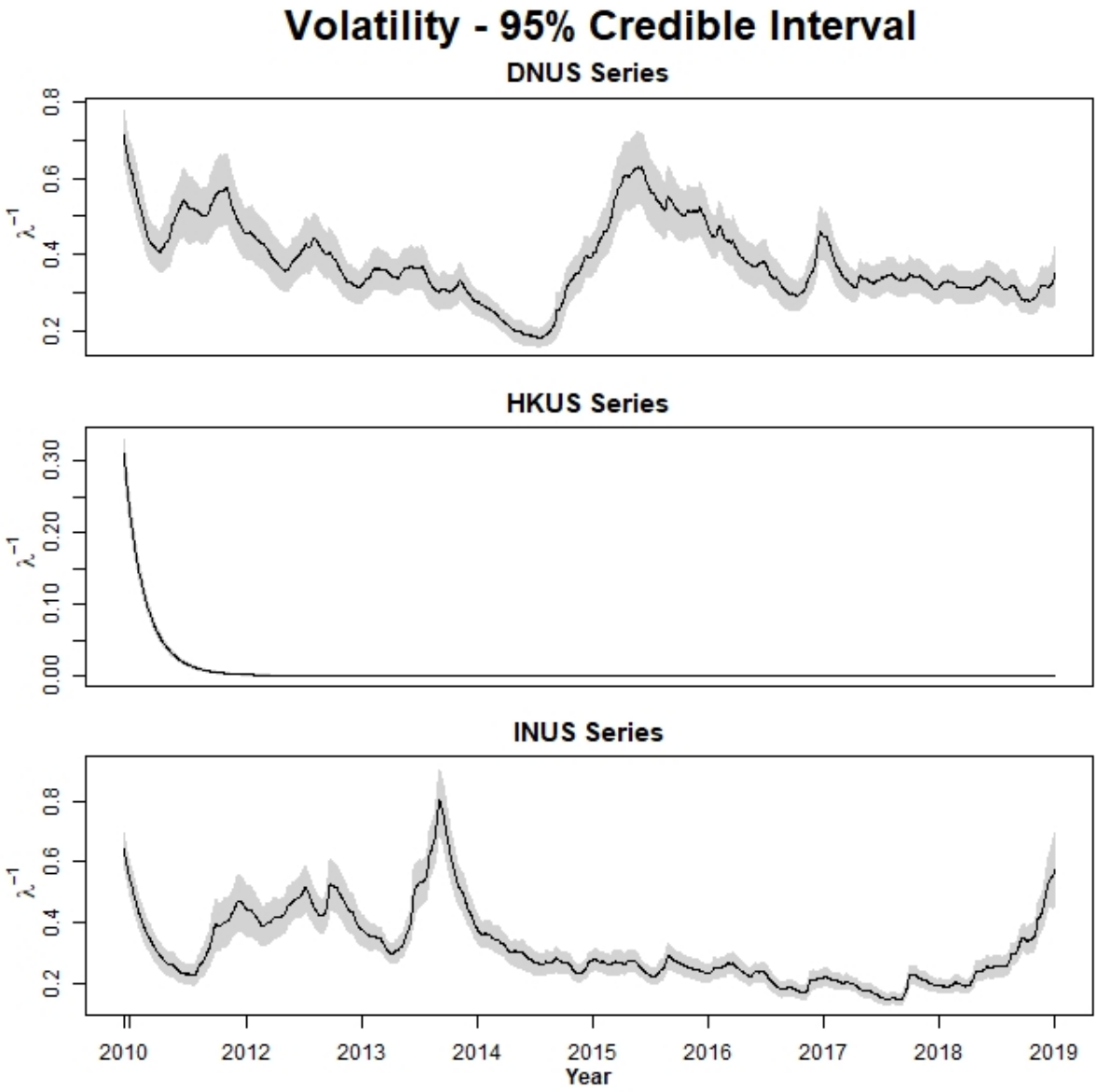


Figure 48 – Posterior mean estimates of instantaneous volatility for Exchange Rates data. Mean estimates in solid line; and 95% credibility interval is the gray area.

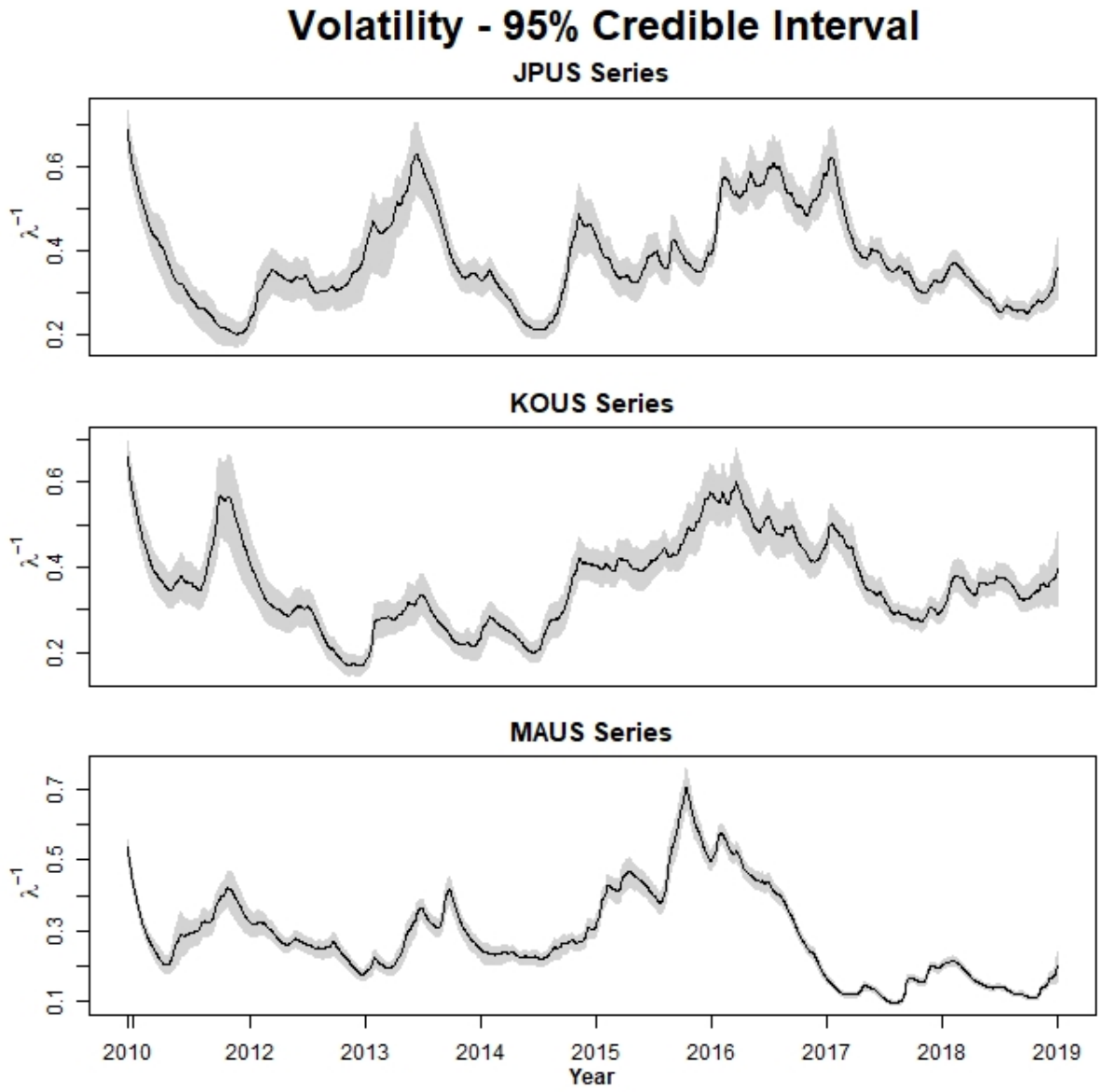


Figure 49 – Posterior mean estimates of instantaneous volatility for Exchange Rates data. Mean estimates in solid line; and 95% credibility interval is the gray area.

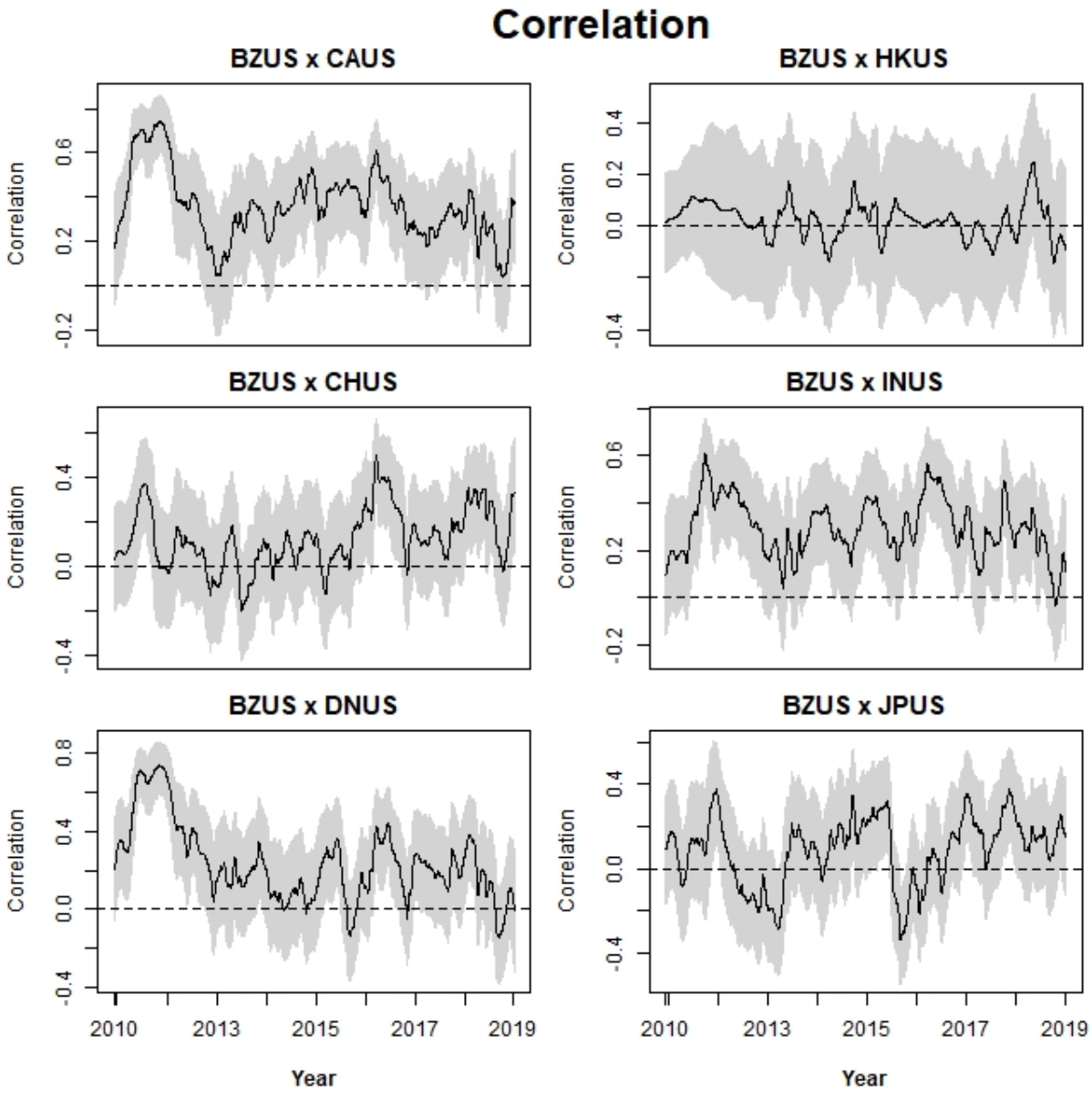


Figure 50 – Posterior mean estimates of instantaneous correlation between Exchange Rates time series. Mean estimates in solid line; and 95% credibility interval is the gray area.

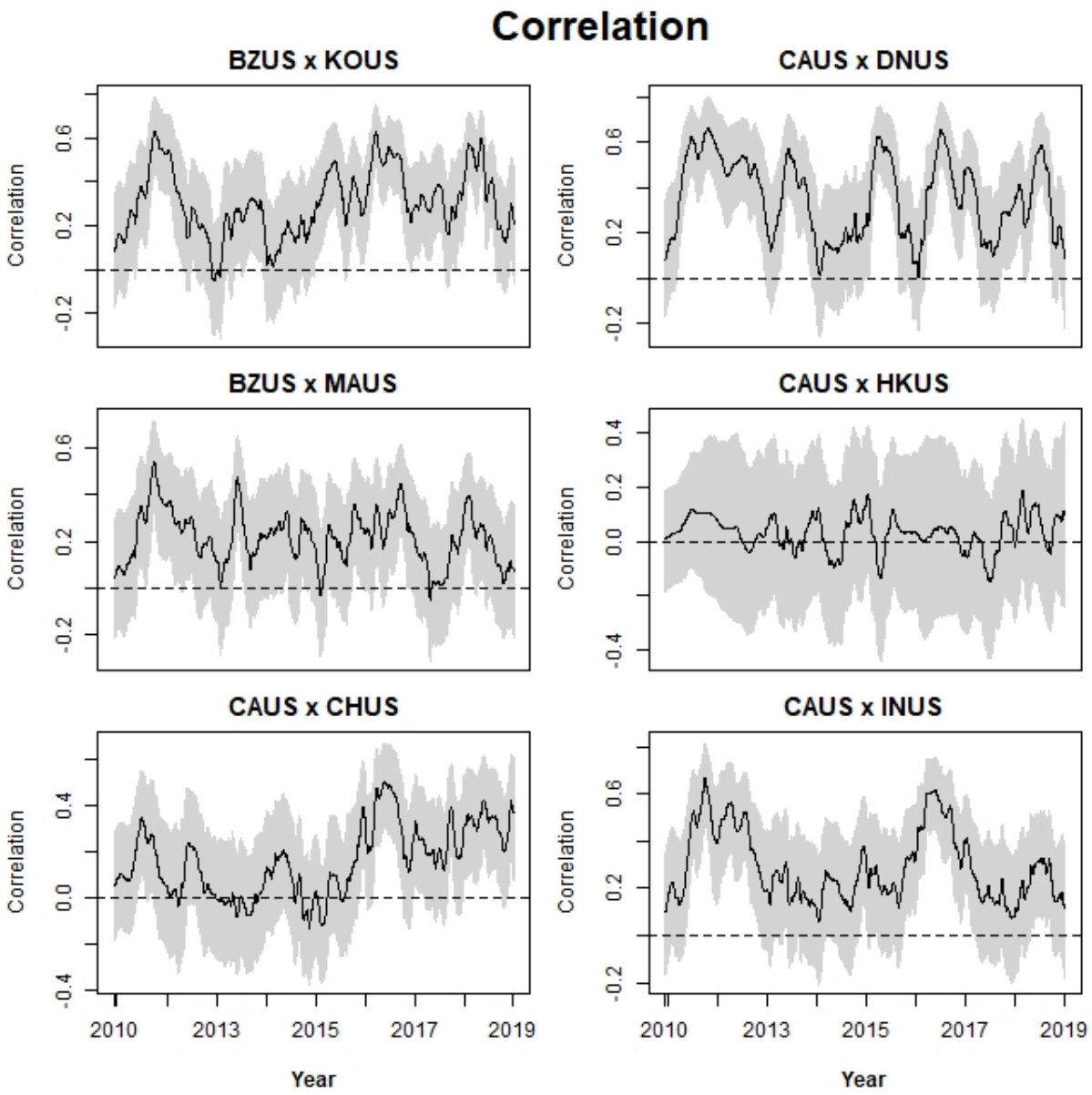


Figure 51 – Posterior mean estimates of instantaneous correlation between Exchange Rates time series. Mean estimates in solid line; and 95% credibility interval is the gray area.

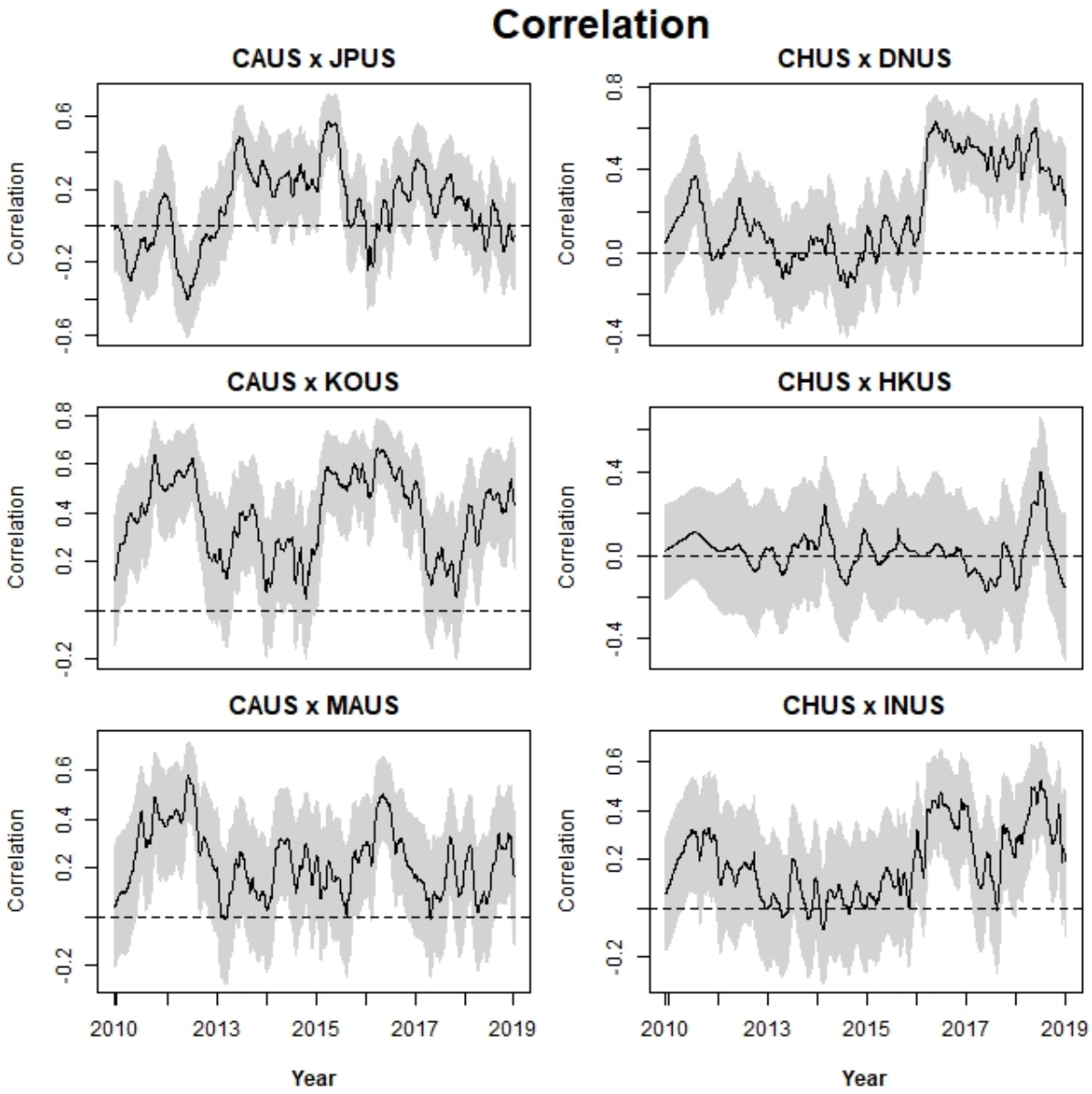


Figure 52 – Posterior mean estimates of instantaneous correlation between Exchange Rates time series. Mean estimates in solid line; and 95% credibility interval is the gray area.

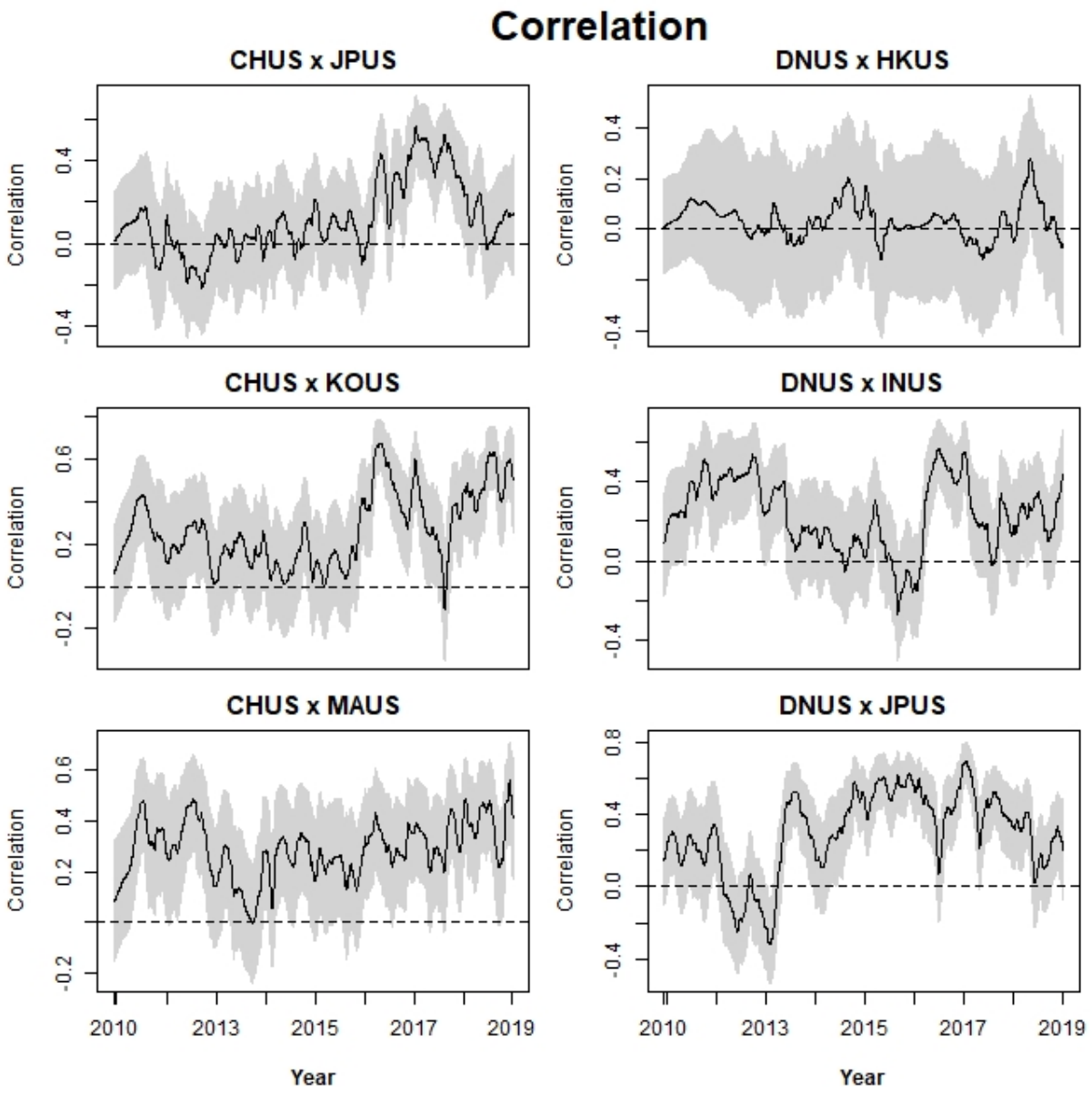


Figure 53 – Posterior mean estimates of instantaneous correlation between Exchange Rates time series. Mean estimates in solid line; and 95% credibility interval is the gray area.

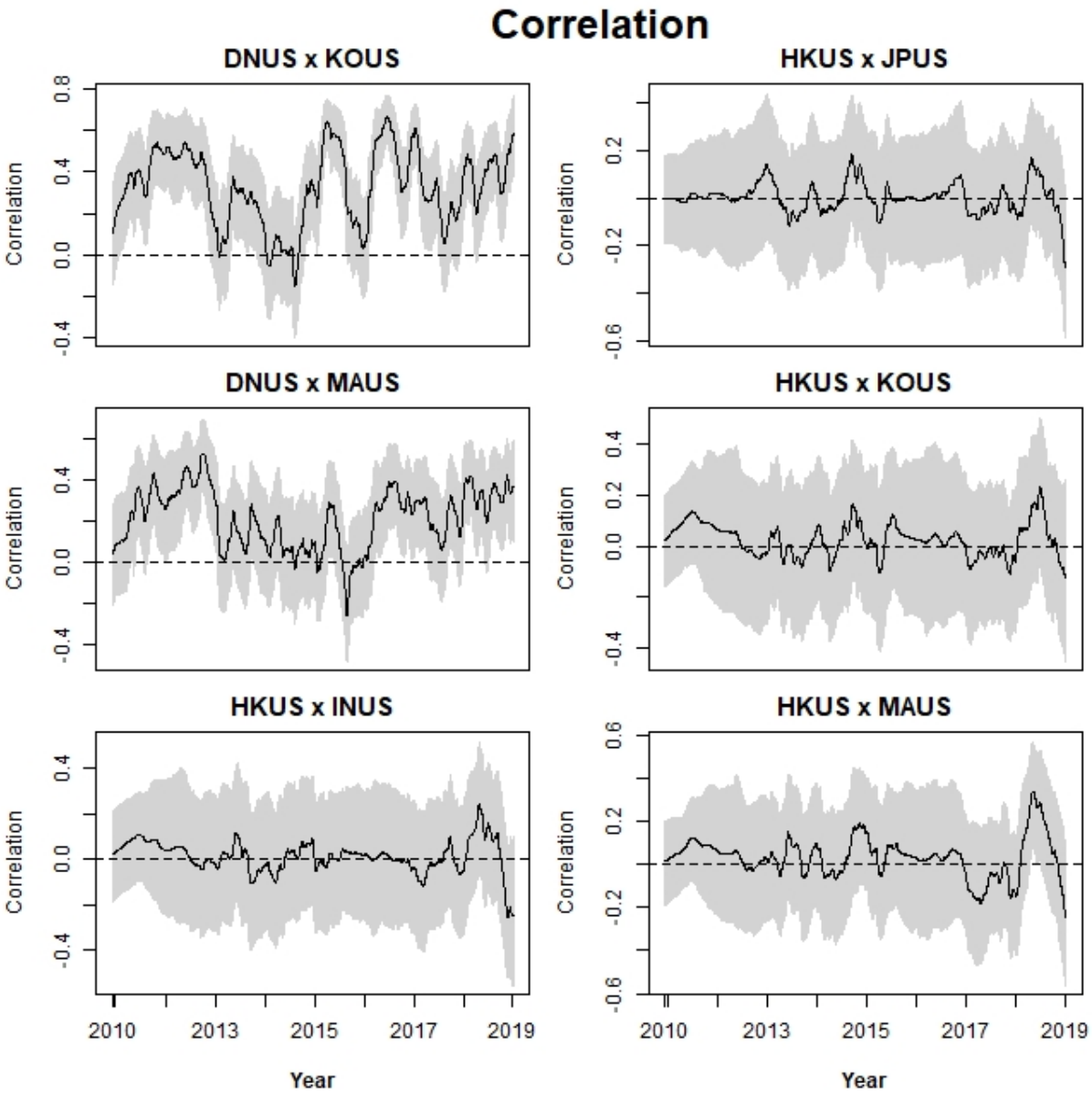


Figure 54 – Posterior mean estimates of instantaneous correlation between Exchange Rates time series. Mean estimates in solid line; and 95% credibility interval is the gray area.

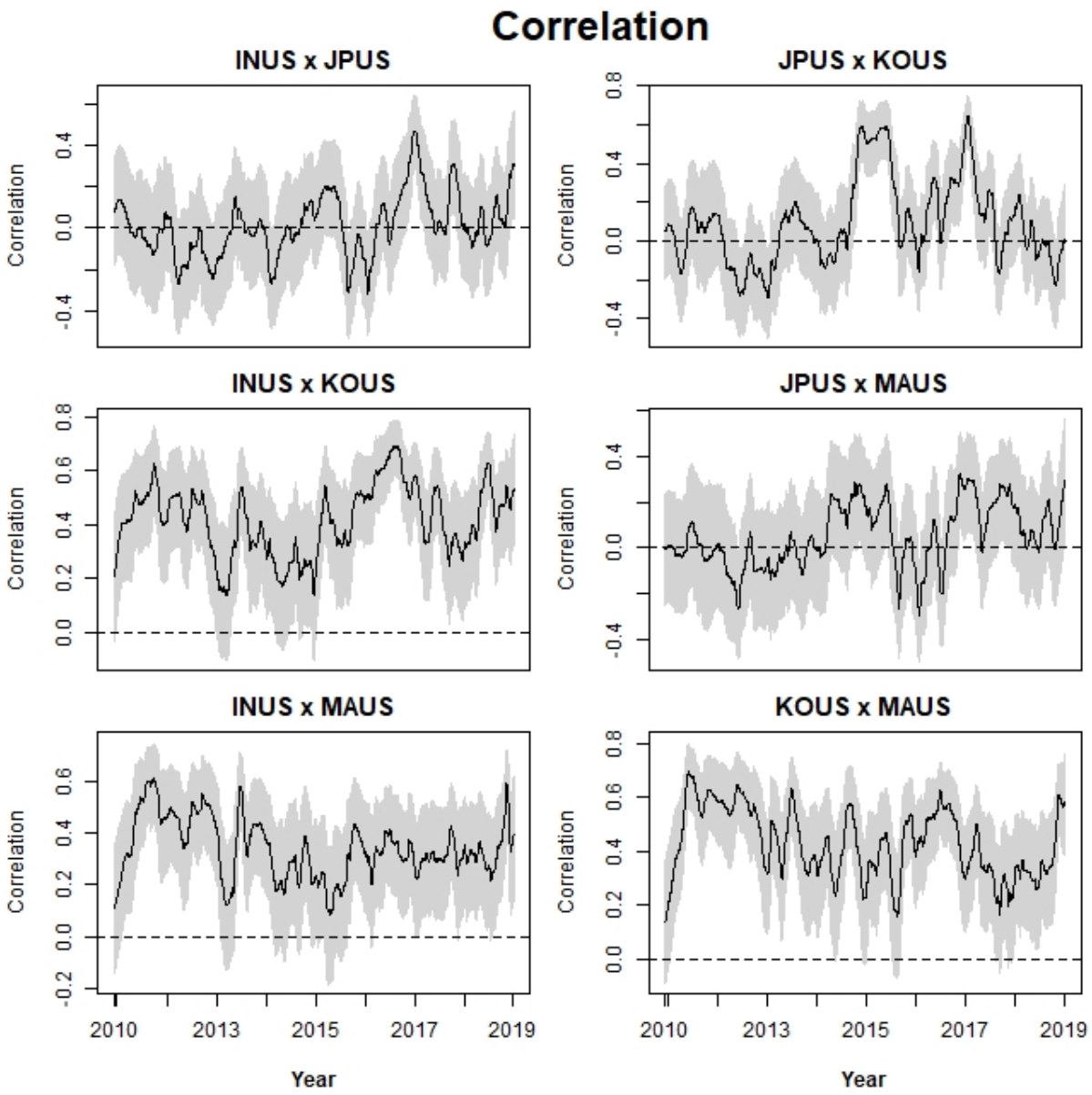


Figure 55 – Posterior mean estimates of instantaneous correlation between Exchange Rates time series. Mean estimates in solid line; and 95% credibility interval is the gray area.

Figures 56, 57 and 58 provides jump sizes for each observation. Jumps usually precede moments of higher volatility, as they are a way of capturing speculative movements at the market. As the higher the number of jumps, more speculative the currency can be considered, thus more vulnerable to external shocks it is. The MSVJB model presents more jumps than MSVJM, since estimatives for volatility are smaller, so that the jump structure is able to capture the excess of returns.

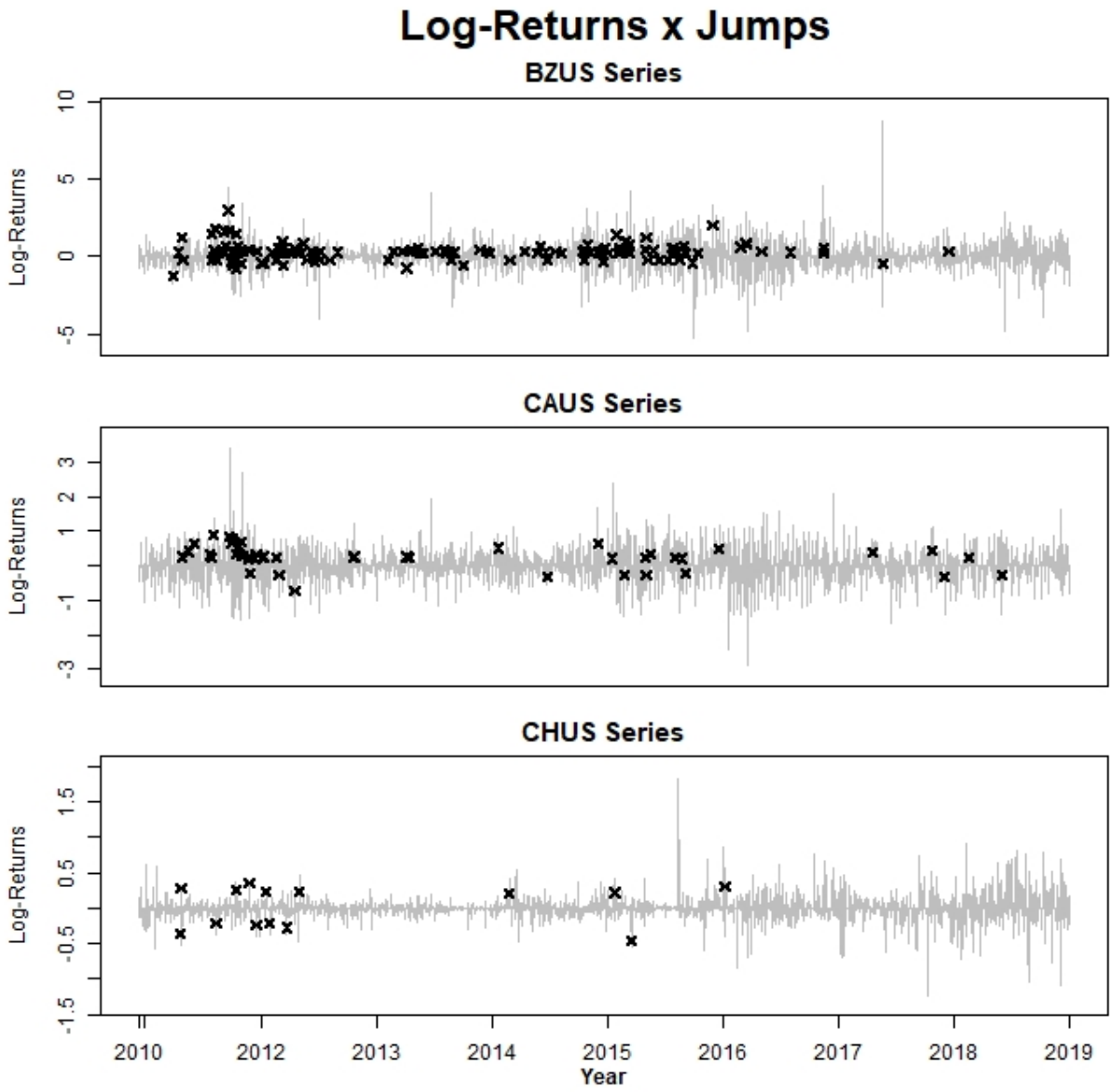


Figure 56 – Posterior mean estimates of instantaneous jumps J_t. Jumps are represented by black crosses and log-returns the light gray lines.

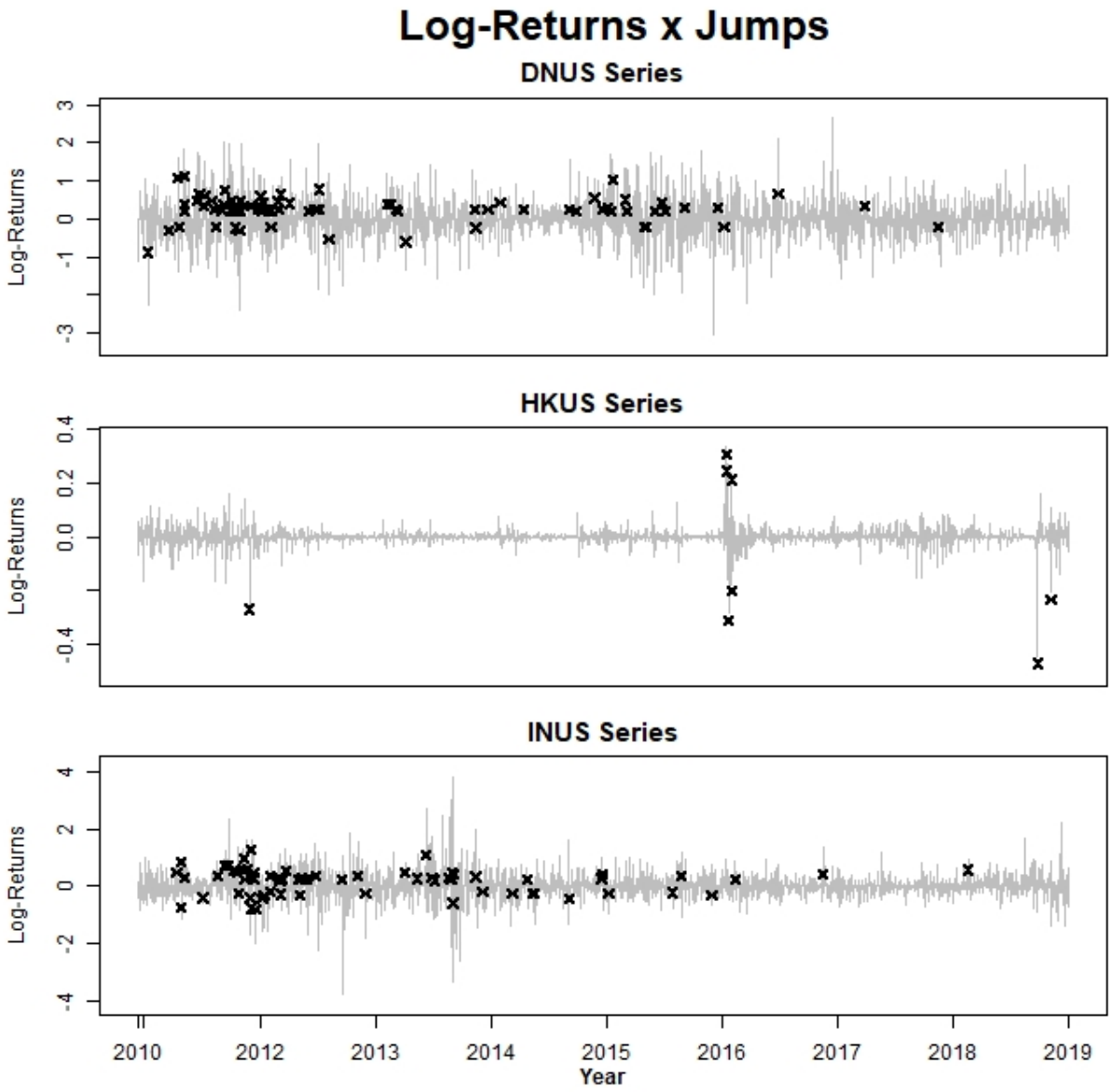


Figure 57 – Posterior mean estimates of instantaneous jumps J_t. Jumps are represented by black crosses and log-returns the light gray lines.

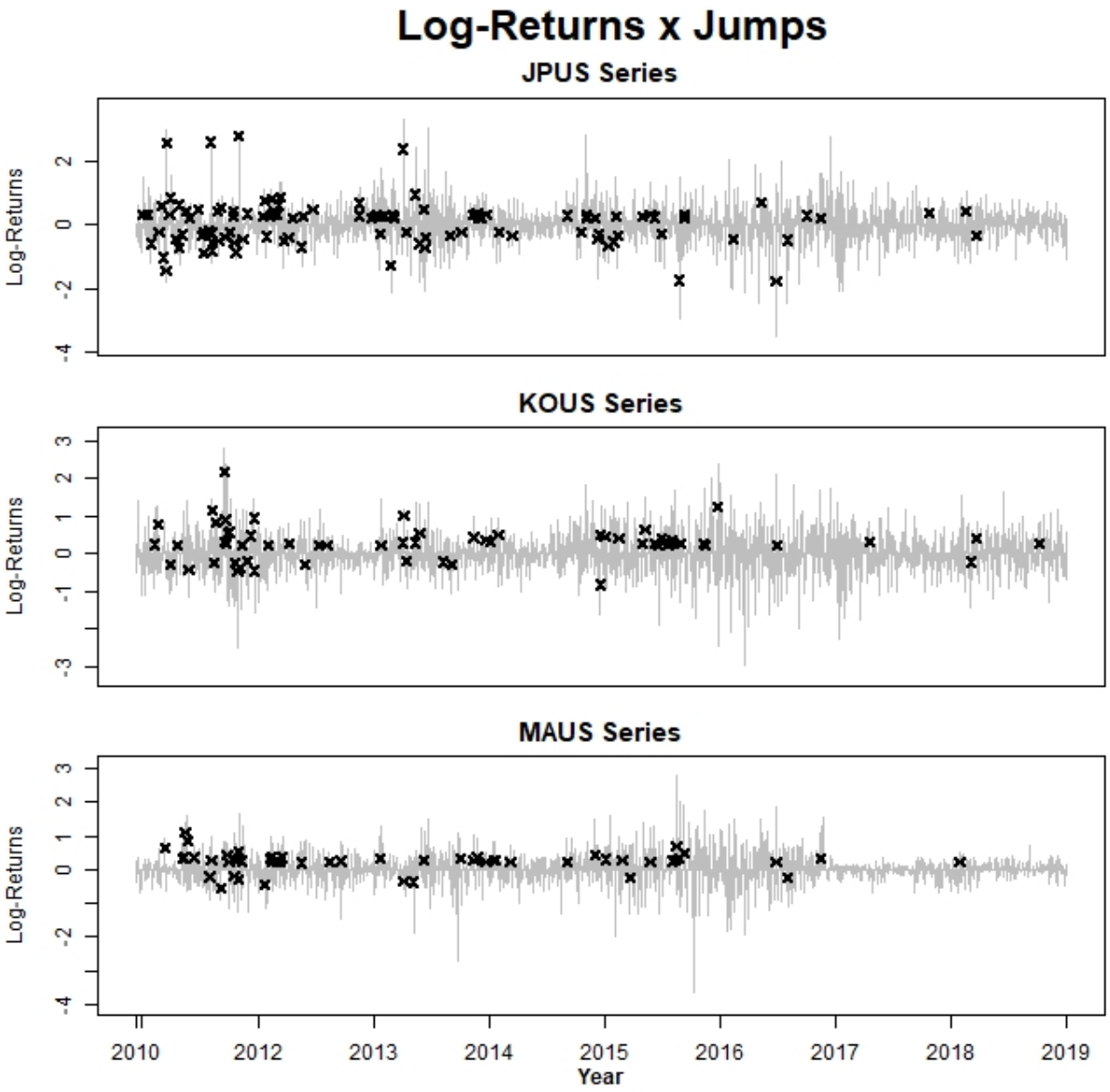


Figure 58 – Posterior mean estimates of instantaneous jumps J_t. Jumps are represented by black crosses and log-returns the light gray lines.

Table 21 shows model estimates for each of the static parameters. For most currencies, jumps are infrequent but have a large magnitude, as can be seen by the large standard deviation. The computational time was 91.91418 minutes. With a higher jump probability, more observations are considered as jumps, leading to a more precise estimation for jump related parameters. For HKUS time series jump probability is extremely high, since volatility estimation is close to zero. Yet, jump magnitude is close to zero, as expected by the government control of this specific exchange rate, leading to distorted statistics for this specific time series.

The model selection criteria BIC and AICc strongly favors MSVJB model over

MSVJM. The MSVJM model has limitations when dealing with higher dimensionality portfolios, due to the Matrix Beta structure. These limitations are overcome on MSVJB model, by the usage of the discount factor b_t and the Beta-Bartlett evolution structure. Also, by sampling from the Wishart distribution trough univariate distributions, computational limitations of directly sampling the entire matrix are mitigated.

On the other hand, a higher computational power is required for using MSVJB model. As results are similar for practical purposes, it is up to the user to balance pros and cons for each model and choose the one that better suits its application. If portfolio dimensionality is small and available computational resourses are limited, MSVJM should be preferred. For higher dimensionality portfolios MSVJB should be preferred.

	Mean	SD		Mean	SD		Mean	SD
μ_{y_1}	0.4375	0.4931	μ_{y_2}	0.2345	0.2693	μ_{y_3}	0.0335	0.0716
μ_{y_4}	0.2864	0.3266	μ_{y_5}	-0.0252	0.0021	μ_{y_6}	0.2074	0.3056
μ_{y_7}	0.0851	0.2950	μ_{y_8}	0.2927	0.2632	μ_{y_9}	0.2242	0.2514
σ_{y_1}	1.0612	0.2807	σ_{y_2}	0.5501	0.1868	σ_{y_3}	0.2033	0.0510
σ_{y_4}	0.6522	0.1780	σ_{y_5}	0.0341	0.0006	σ_{y_6}	0.6506	0.1557
σ_{y_7}	0.9289	0.1899	σ_{y_8}	0.6449	0.1763	σ_{y_9}	0.5216	0.1370
ρ_{y_1}	0.06404	0.04096	ρ_{y_2}	0.06274	0.04502	ρ_{y_3}	0.06288	0.03802
ρ_{y_4}	0.06460	0.04528	ρ_{y_5}	0.97830	0.00337	ρ_{y_6}	0.06089	0.04188
ρ_{y_7}	0.07558	0.04299	ρ_{y_8}	0.06348	0.04024	ρ_{y_9}	0.06270	0.04383
log L	5,818		BIC	-11,453		AICc	-11,587	

Table 21 – Posterior inference of static parameters for MSVJB model for currency exchange rates daily log-returns.

7 Conclusion

This thesis brings major developments for the DM class of models, especially for applications on market risk and portfolio analysis. The NGSVJ, developed by Rego and Santos (2020), has implementation and flexibility advantages when compared to the SV models commonly used on literature, but its application was restricted to univariate data series.

The inclusion of a procedure for estimating the degrees of freedom of the Student-t distribution in the gamma mixture for NGSVJ brought extra flexibility to the model by removing the need to do a grid analysis in order to fix the degree of freedom parameter ν. The usage of a prior proposed by Kwok-Wah (2012) gave satisfactory results on both simulation and real data application scenarios.

Another advance that increases the applicability of NGSVJ on practical situations is the inclusion, inside the model, of a Markov classifier structure, using an HMM. The NGSVJ-HMM is able to simultaneously estimate the volatility of a financial returns time series and classify the market regime. Thus, the user can get an interpretable result on whether the volatility is high or low, related to the overall behavior of the returns of this specific asset. This feature is especially relevant when the user is developing trading strategies on speculative markets. Moments of high volatility imply that changes in price are happening faster then it usually does, so those are good moments for arbitrage operations.

The major contribution of this thesis is extending the NGSVJ to the multivariate case, the MSVJM or MSVJB, to be applied when the user has several asset portfolio candidates and needs to estimate both volatilities for individual assets and correlation between different assets to choose the combination that will reduce risk and increase the expected return from the portfolio. The MSVJM with evolution Beta matrix has a simpler implementation structure and is a natural extension of the NGSVJ. However, it suffers from constraints related to model dimensionality and the discount factor β. As the dimensionality of the portfolio (q) grows large, a smaller range of discount factors can be used, and model functionality gets compromised. The model MSVJB with the Beta-Bartlett evolution releases the MSVJM from those constraints, but its implementation structure is more complex and no smoothing procedure was yet developed to this structure so that estimates for volatility were noisy and counter-intuitive. We proposed an EWMA structure, based on Hawkins and Maboudou-Tchao (2008), to obtain a smoothed estimative of the covariance matrix Σ_t. The results are similar to those found when using the Beta-Bartlett evolution, for the same synthetic data generated on simulation, successfully removing the noise from estimation and the constrains on model dimensionality.

We now offer a complete set of DM or space-state class models to be used on market risk and portfolio analysis. They have the advantage of a simpler implementation structure; faster convergence, since are mostly based on Gibbs Sampler steps; the flexibility to implement new structures, given the SSM structure; and highly interpretable results. Even so, there are some developments left for future works: automatic estimation of discount factors, so that the user does not need to specify a specific value; inclusion of a jump structure for volatility; estimation of the degrees of freedom for the multivariate model; estimating optimal weights for portfolio assets based on MSVJM or MSVJB results; etc.

With this work, we expect to open whole new usage possibilities for DM class models in finance, but also to provide the basis for future applications in other areas of interest, where methods here presented can be applied.

Bibliography

Alexander, L., Das, S. R., Ives, Z., Jagadish, H. V., and Monteleoni, C. (2017). *Research challenges in financial data modeling and analysis*. Big data, 5(3), 177-188.

Anscombe, F. J. (1967). Topics in the investigation of linear relations fitted by the method of least squares. Journal of the Royal Statistical Society: Series B (Methodological), 29(1), 1-29.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate GARCH models: a survey. Journal of applied econometrics, 21(1), 79-109.

Benyacoub, B., ElBernoussi, S., Zoglat, A. and Ismail, E.M. (2014). *Classification with Hidden Markov Model*. Applied Mathematical Sciences, 8(50), pp.2483-2496.

Brooks, C. and Prokopczuk, M. (2011). The Dynamics of Commodity Proces. *ICMA Centre Discussion Papers in Finance*, DP2011-09.

Carriero, A., Clark, T. E., and Marcellino, M. (2019). Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. Journal of Econometrics, 212(1), 137-154.

Chaim, P., and Laurini, M. P. (2018). Volatility and return jumps in bitcoin. Economics Letters, 173, 158-163.

Chib, S. Nardari, F. and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108:281–316.

Chib, S. Nardari, F. and Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2), 341-371.

Chib, S., Omori, Y., and Asai, M. (2009). Multivariate stochastic volatility. *Handbook of Financial Time Series*, Springer, Berlin, Heidelberg.

Dendramis, Y., Kapetanios, G., and Tzavalis, E. (2015). Shifts in volatility driven by large stock market shocks. Journal of Economic Dynamics and Control, 55, 130-147.

Dias, A., and Embrechts, P. (2004). Dynamic copula models for multivariate high-frequency data in finance. Manuscript, ETH Zurich, 81.

Diez, M., Burget, L. and Matejka, P. (2018). *Speaker Diarization based on Bayesian HMM with Eigenvoice Priors*. In Proc. Odyssey 2018 The Speaker and Language Recognition Workshop (pp. 147-154).

Ding, P. (2014). Bayesian robust inference of sample selection using selection-t models. Journal of Multivariate Analysis, 124, 451-464.

Drake, P. P., and Fabozzi, F. J. (2010). The Basics of Finance: An Introduction to Financial Markets, Business Finance, and Portfolio Management. *John Wiley and Sons.*

Duffie, D. and Pan, J. (1997). An Overview of Value at Risk. *The Journal of Derivatives*, 3:7–49.

Eaton, M. L. (1983). *Multivariate statistics: a vector space approach*. JOHN WILEY & SONS, INC., 605 THIRD AVE., NEW YORK, NY 10158, USA, 1983, 512.

Eddelbuettel, Dirk, et al. "Rcpp: Seamless R and C++ integration." Journal of Statistical Software 40.8 (2011): 1-18.

Engle, R. F., and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric theory*, 11(1), 122-150.

Engle, R. F., and Siriwardane, E. N. (2018). Structural GARCH: the volatility-leverage connection. The Review of Financial Studies, 31(2), 449-492.

Eraker, B. Johannes, M. and Polson, N. (2003). The Impact of Jumps in Volatility and Returns. *The Journal of Finance*, 58:1269–1299.

Fengler, M. R., Herwartz, H., and Raters, F. H. C. (2017). Multivariate volatility models. In Applied Quantitative Finance (pp. 25-37). Springer, Berlin, Heidelberg.

Feunou, B., and Tédongap, R. (2012). A stochastic volatility model with conditional skewness. Journal of Business & economic statistics, 30(4), 576-591.

Fonseca, T. C., Ferreira, M. A., and Migon, H. S. (2008). *Objective Bayesian analysis for the Student-t regression model*. Biometrika, 95(2), 325-333.

Federal Reserve Bank of St. Louis., 2019, accessed 3 March 2019, <https://fred.stlouisfed.org>.

Gamerman, D. Santos, T. and Franco, G. (2013). A Non-Gaussian Family of State-Space Models With Exact Marginal Likelihood. *Journal of Time Series Analysis*, 34:625–645.

Gelman, A., and Hill, J. (2006). Data analysis using regression and multilevel/hierarchical models. Cambridge university press.

Gong, X., and Zhuang, X. (2016). Option pricing for stochastic volatility model with infinite activity Lévy jumps. Physica A: Statistical Mechanics and its Applications, 455, 1-10.

Harris, R. D., and Yilmaz, F. (2010). Estimation of the conditional variance–covariance matrix of returns using the intraday range. International Journal of Forecasting, 26(1), 180-194.

Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, 61(2), 247-264.

Hassan, M. R., and Nath, B. (2005). Stock market forecasting using hidden Markov model: a new approach. In 5th International Conference on Intelligent Systems Design and Applications (ISDA'05) (pp. 192-196). IEEE.

Hassan, M. R., Nath, B., and Kirley, M. (2007). A fusion model of HMM, ANN and GA for stock market forecasting. Expert systems with Applications, 33(1), 171-180.

Hawkins, D. M., and Maboudou-Tchao, E. M. (2008). *Multivariate exponentially weighted moving covariance matrix*. Technometrics, 50(2), 155-166.

Hendricks, D. (1995). Evaluation of Value-at-Risk Models Using Historical Data. *Working paper*, Federal Reserve Bank of New York.

Hunter, J. S. (1986). *The exponentially weighted moving average*. Journal of quality technology, 18(4), 203-210.

Izzeldin, M., Tsionas, M. G., and Michaelides, P. G. (2019). Multivariate stochastic volatility with large and moderate shocks. Journal of the Royal Statistical Society: Series A (Statistics in Society), 182(3), 887-917.

Jacquier, E., Polson, N. G., and Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. Journal of Econometrics, 122(1), 185-212.

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

Ji, T. (2018). *A Bayesian Hidden Markov Model for Detecting Differentially Methylated Regions*. Biometrics.

Jondeau, E., and Rockinger, M. (2003). Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. Journal of Economic dynamics and Control, 27(10), 1699-1737.

Juárez, M. A., and Steel, M. F. (2006). Model-based clustering of non-Gaussian panel data. MPRA Paper, 880. Available at: <https://mpra.ub.uni-muenchen.de/880/>.

Juárez, M. A., and Steel, M. F. (2010). Model-based clustering of non-Gaussian panel data based on skew-t distributions. Journal of Business and Economic Statistics, 28(1), 52-66.

Kirkby, J. L., Nguyen, D., and Cui, Z. (2017). A unified approach to Bermudan and barrier options under stochastic volatility models with jumps. Journal of Economic Dynamics and Control, 80, 75-100.

Kshirsagar, A. M. (1959). Bartlett decomposition and Wishart distribution. The Annals of Mathematical Statistics, 30(1), 239-241.

Kwok-Wah, Ho (2012). The use of Jeffreys priors for the Student-t distribution. *Journal of Statistical Computation and Simulation*, 82:7, 1015-1021, DOI: 10.1080/00949655.2011.563239.

Kroner, K. F., and Ng, V. K. (1998). odeling asymmetric comovements of asset returns. *The review of financial studies*, 11(4), 817-844.

Leondes, C. T., Peller, J. B., and Stear, E. B. (1970). *Nonlinear smoothing theory*. IEEE Transactions on Systems Science and Cybernetics, 6(1), 63-71.

Lee, T. H., and Long, X. (2009). Copula-based multivariate GARCH model with uncorrelated dependent errors. Journal of Econometrics, 150(2), 207-218.

Lin, C. H., and Shen, S. S. (2006). Can the student-t distribution provide accurate value at risk?. The Journal of Risk Finance.

Liu, C. (2009). *cuHMM: a CUDA implementation of hidden Markov model training and classification*. The Chronicle of Higher Education, pp.1-13.

Lopes, H.F. and Carvalho, C.M. (2007). *Factor stochastic volatility with time varying loadings and Markov switching regimes*. Journal of Statistical Planning and Inference, 137(10), pp.3082-3091.

Lopes, H. F., McCulloch, R. E., and Tsay, R. S. (2016). Parsimony inducing priors for large scale state-space models. Bayesian Anal.

Luciano, E., and Schoutens, W. (2006). A multivariate jump-driven financial asset model. Quantitative finance, 6(5), 385-402.

Marin, Jean-Michel et al. (2005). *Bayesian modelling and inference on mixtures of distributions.* Handbook of Statistics, 25, 459-507.

McAleer, M. (2005). Automated inference and learning in modeling financial volatility. Econometric Theory, 21(1), 232-261.

McLachlan, G. and Krishnan, T. (2007). *The EM algorithm and extensions* (Vol. 382). John Wiley & Sons.

Meeden, G. and Vardeman, S. (2000). *A simple hidden Markov model for Bayesian modeling with time dependent data*. Communications in Statistics-Theory and Methods, 29(8), pp.1801-1826.

Merener, N. (2015). Concentrated Production and Conditional Heavy Tails in Commodity Returns. *The Journal of Futures Markets*, XX:1–20.

Nakajima, J. and Omori, Y. (2007). Leverage, Heavy-Tails and Correlated Jumps in Stochastic Volatility Models. *CARF Working Paper*, CARF-F-107.

Nakajima, J., and Omori, Y. (2009). Leverage, heavy-tails and correlated jumps in stochastic volatility models. Computational Statistics and Data Analysis, 53(6), 2335-2353.

Nasri, B. R., and Rémillard, B. N. (2019). Copula-based dynamic models for multivariate time series. Journal of Multivariate Analysis, 172, 107-121.

Nkemnole, E. B., and Wulu, J. T. (2017). Modeling of stock indices with HMM-SV models. Theoretical and Applied Economics, 24(2).

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. Journal of Econometrics, 140(2), 425-449.

Ozturk, S. S., and Richard, J. F. (2015). Stochastic volatility and leverage: Application to a panel of S&P500 stocks. Finance Research Letters, 12, 67-76.

Patton, A. (2013). Copula methods for forecasting multivariate time series. In Handbook of economic forecasting (Vol. 2, pp. 899-960). Elsevier.

Pohle, J., Langrock, R., van Beest, F.M. and Schmidt, N.M. (2017). *Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement*. Journal of Agricultural, Biological and Environmental Statistics, 22(3), pp.270-293.

Prado, R. and West, M. (2010). Time Series Modeling, Computing and Inference. *CRC Press*, Boca Raton.

Quintana, J. M., V. Lourdes, O. Aguilar, and J. Liu (2003). Global gambling. *Bayesian Statistics VII, Oxford*, pp. 349-368.

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

Rego, A.T. and Santos, T.R. (2020). *Non-Gaussian Stochastic Volatility Model with Jumps via Gibbs Sampler*. Statistics and Its Interfaces, 13 (2), pp.209-219.

Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.

Riley, W. J. (2008). Algorithms for frequency jump detection. *Metrologia*, 45:S154–S161.

Santos, T.R. (2018). *A Bayesian GED-Gamma stochastic volatility model for return data: a marginal likelihood approach*. arXiv preprint arXiv:1809.01489.

Sebastian, T., Jeyaseelan, V., Jeyaseelan, L., Anandan, S., George, S. and Bangdiwala, S.I. (2018). *Decoding and modelling of time series count data using Poisson hidden Markov model and Markov ordinal logistic regression models*. Statistical methods in medical research, p.0962280218766964.

Shen, X., Tsung, F., and Zou, C. (2014). A new multivariate EWMA scheme for monitoring covariance matrices. International Journal of Production Research, 52(10), 2834-2850.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and SØrbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. Statistical science, 32(1), 1-28.

Smith, D. R. (2002). Markov-switching and stochastic volatility diffusion models of short-term interest rates. Journal of Business & Economic Statistics, 20(2), 183-197.

Stoyanov, J., and Lin, G. D. (2011). Mixtures of power series distributions: identifiability via uniqueness in problems of moments. Annals of the Institute of Statistical Mathematics, 63(2), 291-303.

Subrahmanyam, A. (2019). *Big data in finance: Evidence and challenges*. Borsa Istanbul Review.

Sun, C. (2018). *Implementing Trade Strategy with HMM Model: A Practice on Some Telecommunication Companies*. Open Journal of Social Sciences, 6(03), p.12.

The R foundation for Statistical Computing.(2015). R version 3.2.1 [Computer Software]. Retrieved from *http://www.r-project.org/*.

Ranganathan, V. Why is the Hong Kong dollar probing the weak end of its band?. Reuters. March, 9, 2018. Retrieved from: <https://www.reuters.com/article/us-hongkong-dollar-explainer/why-is-the-hong-kong-dollar-probing-the-weak-end-of-its-band-idUSKCN1GL0KT>

Triantafyllopoulos, K. (2008). Multivariate stochastic volatility with Bayesian dynamic linear models. *Journal of Statistical Planning and Inference*, 138:1021–1037.

Triantafyllopoulos, K. (2014). Multivariate Stochastic Volatility Estimation Using Particle Filters. *In Topics in Nonparametric Statistics*, (pp. 335-345). Springer, New York, NY.

Villa, C. and Walker, J.(2014). Objective Prior for the Number of Degrees of Freedom of a t Distribution. *Bayesian Analysis*, 9, pp. 197-220.

Warty, S. P., Lopes, H. F., and Polson, N. G. (2018). *Sequential Bayesian learning for stochastic volatility with variance-gamma jumps in returns*. Applied Stochastic Models in Business and Industry, 34(4), 460-479.

West, M. and Harrison, J. (1997). Bayesian Forecasting and Dynamic Models. *Springer*, New York.

West, M. (2020). Bayesian forecasting of multivariate time series: scalability, structure uncertainty and decisions. Annals of the Institute of Statistical Mathematics, 72(1), 1-31.

Xu, L. and Jordan, M.I. (1996). *On convergence properties of the EM algorithm for Gaussian mixtures*. Neural computation, 8(1), pp.129-151.

Xu, Y., and Jasra, A. (2019). Particle filters for inference of high-dimensional multivariate stochastic volatility models with cross-leverage effects. Foundations of Data Science, 1(1), 61-85.

Yu, J., and Meyer, R. (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. Econometric Reviews, 25(2-3), 361-384.

Zhou, H., and Lawson, A. B. (2008). EWMA smoothing and Bayesian spatial modeling for health surveillance. Statistics in Medicine, 27(28), 5907-5928.

Appendix

APPENDIX A – Full Conditional Posterior Distributions for MSVJ Model

This appendix presents the proof of the conditional posterior distributions for the variables MSVJ model. For simplifying notation, let $\Phi = (Y_t, \theta_t, J_t, \gamma_t, \Sigma_t, \mu_y, \sigma_y^2, \nu)$ be model parameters, and $\Phi_{[-\bullet]}$ be all parameters excluding that on the brackets.

A.1 Full Conditional Posterior Distribution for γ_t

As the prior

$$
\begin{aligned}
p(\gamma_t) &\sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \text{thus} \\
p(\gamma_t|\mathcal{D}_n, \Phi_{[-\gamma_t]}) &\propto (\gamma_t)^{\frac{\nu}{2}-1} e^{-\gamma_t \frac{\nu}{2}} |\gamma_t^{-1}\Sigma_t|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)^T(\gamma_t^{-1}\Sigma_t)^{-1}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)\right), \\
p(\gamma_t|\mathcal{D}_n, \Phi_{[-\gamma_t]}) &\propto (\gamma_t)^{\frac{\nu}{2}-1+\frac{q}{2}} \exp\left(-\gamma_t\left[\frac{\nu}{2} + \frac{tr\left((\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)^T(\Sigma_t)^{-1}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)\right)}{2}\right]\right), \\
p(\gamma_t|\mathcal{D}_n, \Phi_{[-\gamma_t]}) &\sim G\left(\frac{\nu+q}{2} - 1, \frac{\nu}{2} + \frac{tr\left((\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)^T(\Sigma_t)^{-1}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)\right)}{2}\right).
\end{aligned}
$$

A.2 Full Conditional Posterior Distribution for μ_y

Let $\underset{\sim}{\mu_y}$ be $\mu_y = (\mu_{y_1}, \mu_{y_2}, \ldots, \mu_{y_p})$. Then, for each μ_{y_p}, $p = 1, \ldots, q$, the full conditional posterior distribution is given by:

As the prior:

$$
\begin{aligned}
p(\mu_{y_p}) &\sim N(m, v), \text{thus} \\
p(\mu_{y_p}|\mathcal{D}_n, \Phi_{[-\mu_y]}) &\propto \exp\left(-\frac{1}{2v}(\mu_{y_p} - m)^2\right) \exp\left(-\frac{1}{2\sigma^2_{y_p}} \sum_{\substack{i=1 \\ J_{p,i}\neq 0}}^{t} (\xi_{p,i+1} - \mu_{y_p})^2\right), \\
p(\mu_{y_p}|\mathcal{D}_n, \Phi_{[-\mu_y]}) &\propto \exp\left(-\frac{1}{2v}(\mu_{y_p} - m)^2 - \frac{1}{2\sigma^2_{y_p}} \left(\sum_{\substack{i=1 \\ J_{p,i}\neq 0}}^{t} (\xi_{p,i+1})^2 - 2\mu_{y_p} n_p \bar{\xi_p} + n_p \mu^2_{y_p}\right)\right), \\
p(\mu_{y_p}|\mathcal{D}_n, \Phi_{[-\mu_y]}) &\propto \exp\left(-\frac{1}{2v\sigma^2_{y_p}} \left[(\sigma^2_{y_p} + n_p v)\mu^2_{y_p} - 2\mu_{y_p}(m\sigma^2_{y_p} + v n_p \bar{\xi_p})\right]\right) \\
p(\mu_{y_p}|\mathcal{D}_n, \Phi_{[-\mu_y]}) &\propto \exp\left(-\frac{1}{2\frac{v\sigma^2_{y_p}}{\sigma^2_{y_p} + n_p v}} \left[\mu_{y_p} - \frac{m\sigma^2_{y_p} + v n_p \bar{\xi_p}}{\sigma^2_{y_p} + n_p v}\right]^2\right), \\
p(\mu_{y_p}|\mathcal{D}_n, \Phi_{[-\mu_y]}) &\sim N\left(\frac{m\sigma^2_{y_p} + v n_p \bar{\xi_p}}{\sigma^2_{y_p} + n_p v}, \frac{v\sigma^2_{y_p}}{\sigma^2_{y_p} + n_p v}\right).
\end{aligned}
$$

where n_p is number of times so that $J_{p,i} \neq 0$ and $\bar{\xi_p} = \frac{\sum_{\substack{i=1 \\ J_{p,i}\neq 0}}^{t} \xi_{p,i+1}}{n_p}$.

A.3 Full Conditional Posterior Distribution for σ_y^2

Let $\underset{\sim}{\sigma_y^2}$ be $\sigma_y^2 = (\sigma_{y_1}^2, \sigma_{y_2}^2, \ldots, \sigma_{y_p}^2)$. Then, for each $\sigma_{y_p}^2$, $p = 1, \ldots, q$, the full conditional posterior distribution is given by:

As the prior:

$$
\begin{aligned}
p(\sigma_{y_p}^2) &\sim IG(\alpha, \beta), \text{thus} \\
p(\sigma_{y_p}^2 | \mathcal{D}_n, \Phi_{[-\sigma_y^2]}) &\propto \left(\frac{1}{\sigma_{y_p}^2}\right)^{\alpha+1} e^{-\frac{\beta}{\sigma_{y_p}^2}} \left(\frac{1}{\sigma_{y_p}^2}\right)^{\frac{n_p}{2}} \exp\left(-\frac{1}{2\sigma_{y_p}^2} \sum_{\substack{i=1 \\ J_i \neq 0}}^{t} (\xi_{p,i+1} - \mu_{y_p})^2\right), \\
p(\sigma_{y_p}^2 | \mathcal{D}_n, \Phi_{[-\sigma_y^2]}) &\propto (\sigma_{y_p}^2)^{-\left(\alpha + \frac{n_p}{2} + 1\right)} \exp\left(-\frac{1}{\sigma_{y_p}^2}\left[\beta + \frac{\sum_{\substack{i=1 \\ J_i \neq 0}}^{t} (\xi_{p,i+1} - \mu_{y_p})^2}{2}\right]\right), \\
p(\sigma_{y_p}^2 | \mathcal{D}_n, \Phi_{[-\sigma_y^2]}) &\sim IG\left(\alpha + \frac{n_p}{2}, \beta + \frac{\sum_{\substack{i=1 \\ J_i \neq 0}}^{t} (\xi_{p,i+1} - \mu_{y_p})^2}{2}\right).
\end{aligned}
$$

A.4 Full Conditional Posterior Distribution for ξ_{t+1}

Let $\Sigma_y = diag(\sigma_y^2)$, that is, $\Sigma_y = diag(\sigma_{y_1}^2, \sigma_{y_2}^2, \ldots, \sigma_{y_p}^2)$ and $\mu_y = (\mu_{y_1}, \mu_{y_2}, \ldots, \mu_{y_p})$, $p = 1, \ldots, q$.

As the prior

$$
\begin{aligned}
p(\xi_{t+1}) &\sim N(\mu_y, \Sigma_y), \text{thus} \\
p(\xi_{t+1}|\mathcal{D}_n, \Phi_{[-\xi_t]}) &\propto \exp\left(-\frac{1}{2}(\xi_{t+1} - \mu_y)^T(\Sigma_y)^{-1}(\xi_{t+1} - \mu_y)\right) \\
&\quad \times \exp\left(-\frac{1}{2}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - \xi_{t+1})^T(\gamma_t^{-1}\Sigma_t)^{-1}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - \xi_{t+1})\right), \\
p(\xi_{t+1}|\mathcal{D}_n, \Phi_{[-\xi_t]}) &\propto \exp\left(-\frac{1}{2}\left[\xi_{t+1}^T\Sigma_y^{-1}\xi_{t+1} - 2\xi_{t+1}^T\Sigma_y^{-1}\mu_y\right]\right) \\
&\quad \times \exp\left(-\frac{1}{2}\left[\xi_{t+1}^T(\gamma_t^{-1}\Sigma_t)^{-1}\xi_{t+1} - 2\xi_{t+1}^T(\gamma_t^{-1}\Sigma_t)^{-1}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t)\right]\right), \\
p(\xi_{t+1}|\mathcal{D}_n, \Phi_{[-\xi_t]}) &\propto \exp\left(\xi_{t+1}^T(\gamma_t^{-1}\Sigma_t)^{-1}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t) - \frac{1}{2}\xi_{t+1}^T(\gamma_t^{-1}\Sigma_t)^{-1}\xi_{t+1}\right) \\
&\quad \times \exp\left(-\frac{1}{2}\xi_{t+1}^T\Sigma_y^{-1}\xi_{t+1} + \xi_{t+1}^T\Sigma_y^{-1}\mu_y\right), \\
p(\xi_{t+1}|\mathcal{D}_n, \Phi_{[-\xi_t]}) &\propto \exp\left(\xi_{t+1}^T\left[\Sigma_y^{-1}\mu_y + (\gamma_t^{-1}\Sigma_t)^{-1}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t)\right] - \frac{1}{2}\xi_{t+1}^T\left[\Sigma_y^{-1} + (\gamma_t^{-1}\Sigma_t)^{-1}\right]\right), \\
p(\xi_{t+1}|\mathcal{D}_n, \Phi_{[-\xi_t]}) &\sim N\left(V\left(\Sigma_y^{-1}\mu_y + (\gamma^{-1}\Sigma_t)^{-1}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t)\right), V\right).
\end{aligned}
$$

with $V = \left(\Sigma_y^{-1} + (\gamma^{-1}\Sigma_t)^{-1}\right)^{-1}$.

A.5 Full Conditional Posterior Distribution for ρ

Let ρ be $\rho = (\rho_1, \rho_2, \ldots, \rho_p)$. Then, for each ρ_p, $p = 1, \ldots, q$, the full conditional posterior distribution is given by:

As the prior

$$
\begin{aligned}
p(\rho_p) &\sim Beta(\alpha, \beta), \text{ thus} \\
p(\rho_p|\mathcal{D}_n, \Phi_{[-\rho]}) &\propto \rho_p^{\alpha-1}(1-\rho_p)^{\beta-1}\prod_{i=1}^{n}\rho_p^{N_{p,i}}(1-\rho_p)^{1-N_{p,i}} \\
p(\rho_p|\mathcal{D}_n, \Phi_{[-\rho]}) &\propto \rho_p^{\alpha-1}(1-\rho_p)^{\beta-1}\rho_p^{\sum_{i=0}^{t}N_{p,i}}(1-\rho_p)^{n-\sum_{i=0}^{t}N_{p,i}} \\
p(\rho_p|\mathcal{D}_n, \Phi_{[-\rho]}) &\propto \rho_p^{\alpha+\sum_{i=0}^{t}N_{p,i}-1}(1-\rho_p)^{\beta+n-\sum_{i=0}^{t}N_{p,i}-1} \\
p(\rho_p|\mathcal{D}_n, \Phi_{[-\rho]}) &\sim Beta\left(\alpha + \sum_{i=0}^{t}N_{p,i}, \beta + n - \sum_{i=0}^{t}N_{p,i}\right)
\end{aligned}
$$

A.6 Full Conditional Posterior Distribution for N_{t+1}

$$
\begin{aligned}
P(N_{p,t+1} = 1|\Phi_{[-N_t]}) &\propto \rho_p P(\underset{\sim}{y}_{t+1}|N_{p,t+1} = 1, \Phi_{-[N_t]}) \\
P(N_{p,t+1} = 0|\Phi_{[-N_t]}) &\propto (1-\rho_p) P(\underset{\sim}{y}_{t+1}|N_{p,t+1} = 0, \Phi_{-[N_t]})
\end{aligned}
$$

Since N^y can assume only two values, 0 or 1:

$$
P(N_{p,t+1} = 1|\Phi_{-[N_t]}) = \frac{\rho_p P(\underset{\sim}{y}_{t+1}|N_{p,t+1} = 1, \Phi_{-[N_t]})}{\rho_p P(\underset{\sim}{y}_{t+1}|N_{p,t+1} = 1, \Phi_{-[N_t]}) + (1-\rho_p) P(\underset{\sim}{y}_{t+1}|N_{p,t+1} = 0, \Phi_{-[N_t]})}.
$$

Using the idea proposed by Brooks and Prokopczuk (2011), be α a threshold so that:

$$
N_{p,t+1} = \begin{cases} 1 & if \quad P(N_{p,t+1} = 1|\Phi_{-[N_t]}) > \alpha \\ 0 & if \quad P(N_{p,t+1} = 1|\Phi_{-[N_t]}) \leq \alpha \end{cases}.
$$

A.7 Full Conditional Posterior Distribution for Σ_t

The full conditional posterior distribution for $\Sigma_t|\Phi_{[-\Sigma_t]}$ is given by:

As the prior

$$
\begin{aligned}
p(\Sigma_t) &\sim IW(n_t, D_t) \text{ thus} \\
p(\Sigma_t|\mathcal{D}_n, \Phi_{[-\Sigma_t]}) &\propto |\gamma_t^{-1}\Sigma_t|^{-\frac{1}{2}} exp\left(-\frac{1}{2}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)^T(\gamma_t^{-1}\Sigma_t)^{-1}\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)\right) \\
&\quad \times |\Sigma_t|^{-(q+n_t)/2} etr\left(-\frac{1}{2}\Sigma_t^{-1}D_t\right) \\
p(\Sigma_t|\mathcal{D}_n, \Phi_{[-\Sigma_t]}) &\propto |\Sigma_t|^{-(q+n_t+1)/2} etr\left(-\frac{1}{2}\left[\Sigma_t^{-1}D_t + \gamma_t(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)^T\Sigma_t^{-1}(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)\right]\right) \\
p(\Sigma_t|\mathcal{D}_n, \Phi_{[-\Sigma_t]}) &\propto |\Sigma_t|^{-(q+n_t+1)/2} etr\left(-\frac{1}{2}\Sigma_t^{-1}\left[D_t + \gamma_t(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)^T(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)\right]\right) \\
p(\Sigma_t|\mathcal{D}_n, \Phi_{[-\Sigma_t]}) &\sim IW\left(n_t + 1, D_t + \gamma_t(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)^T(\underset{\sim}{y}_t - \underset{\sim}{\theta}_t - J_t)\right)
\end{aligned}
$$

APPENDIX B – Matrix-Beta Evolution Retrospective Analysis

B.1 MSVJM: Sequential Analysis and Smoothing Procedures

In matrix-beta evolution, presented by Prado and West (2010), we have that:

- Posterior at $t-1$: $(\Omega_{t-1}|\mathcal{D}_{t-1}) \sim W(h_{t-1}, D_{t-1}^{-1})$,
- Prior at t: $(\Omega_t|\mathcal{D}_{t-1}) \sim W(\beta h_{t-1}, (\beta D_{t-1})^{-1})$.

where, for simplifying notation, $\mathcal{D}_t$ represents all the available information at time t.

The evolution distribution $p(\Omega_t|\Omega_{t-1}, \mathcal{D}_{t-1})$ is defined as follows.

Given Ω_{t-1}, set $\Omega_t = U_{t-1}^T \Gamma_t U_{t-1} \beta^{-1}$, where U_{t-1} is the upper triangular Cholesky component of Ω_{t-1} and Γ_t is a $q \times q$ matrix random quantity having matrix-beta distribution, denoted by:

$$(\Gamma_t|\mathcal{D}_{t-1}) \sim Beta\left(\frac{\beta h_{t-1}}{2}, \frac{(1-\beta)h_{t-1}}{2}\right) \tag{B.1}$$

for values of βh_{t-1} and $(1-\beta)h_{t-1}$ greater than $q-1$.

For a retrospective analysis and to sample from the joint distribution of $(\Sigma_1, ..., \Sigma_n|\mathcal{D}_n)$, it is necessary to sample from $(\Omega_{t-1}|\Omega_t, \mathcal{D}_{t-1})$, we have that:

$$p(\Omega_{t-1}|\Omega_t, \mathcal{D}_{t-1}) = \frac{p(\Omega_t|\Omega_{t-1}, \mathcal{D}_{t-1}) \times p(\Omega_{t-1}|\mathcal{D}_{t-1})}{p(\Omega_t|\mathcal{D}_{t-1})} \tag{B.2}$$

Note that $\Gamma_t = U_{t-1}^T \Omega_t U_{t-1} \beta$, so we can write:

$$\begin{aligned} p(\Omega_{t-1}|\Omega_t, \mathcal{D}_{t-1}) \propto\ & |{U_{t-1}^T}^{-1} \Omega_t U_{t-1}^{-1} \beta|^{\frac{\beta h_{t-1}-(q+1)}{2}} |I_q - {U_{t-1}^T}^{-1} \Omega_t U_{t-1}^{-1} \beta|^{\frac{(1-\beta)h_{t-1}-(q+1)}{2}} \times \\ & |\Omega_{t-1}|^{\frac{h_{t-1}-(q+1)}{2}} etr\left\{-\frac{D_{t-1}^{-1}\Omega_{t-1}}{2}\right\} \times |{U_{t-1}^T}^{-1} U_{t-1}^{-1} \beta|^{\frac{-(q+1)}{2}} \times \\ & |\Omega_t|^{\frac{-\beta h_{t-1}+(q+1)}{2}} etr\left\{\frac{\beta D_{t-1}^{-1}\Omega_t}{2}\right\} \end{aligned}$$

$$
\begin{aligned}
p(\Omega_{t-1}|\Omega_t, \mathcal{D}_{t-1}) \propto\ & |{U_{t-1}^T}^{-1}\Omega_t U_{t-1}^{-1}\beta|^{\frac{\beta h_{t-1}-(q+1)}{2}} |I_q - {U_{t-1}^T}^{-1}\Omega_t U_{t-1}^{-1}\beta|^{\frac{(1-\beta)h_{t-1}-(q+1)}{2}} \times \\
& |\Omega_{t-1}|^{\frac{h_{t-1}-(q+1)}{2}} \times |\Omega_t|^{-\frac{\beta h_{t-1}-(q+1)}{2}} \times |{U_{t-1}^T}^{-1}U_{t-1}^{-1}|^{\frac{-(q+1)}{2}} \times \\
& etr\left\{-\frac{D_{t-1}^{-1}}{2}\left[\Omega_{t-1} - \beta\Omega_t\right]\right\}
\end{aligned}
$$

Applying the following matrix properties,

- $|AB| = |A||B| = |B||A|$;
- $|A^{-1}| = 1/|A|$;
- $(A^T)^{-1} = (A^{-1})^T$.

we have that:

$$
\begin{aligned}
|{U_{t-1}^T}^{-1}\Omega_t U_{t-1}^{-1}|^{\frac{\beta h_{t-1}-(q+1)}{2}} &= \left(|{U_{t-1}^T}^{-1}| \times |\Omega_t| \times |{U_{t-1}}^{-1}|\right)^{\frac{\beta h_{t-1}-(q+1)}{2}} \\
&= \left(|\Omega_t| \times |{U_{t-1}}^{-1}{U_{t-1}^T}^{-1}|\right)^{\frac{\beta h_{t-1}-(q+1)}{2}} \\
&= \left(|\Omega_t| \times |(U_{t-1}^T U_{t-1})^{-1}|\right)^{\frac{\beta h_{t-1}-(q+1)}{2}} \\
&= \left(|\Omega_t| \times |\Omega_{t-1}|^{-1}\right)^{\frac{\beta h_{t-1}-(q+1)}{2}} \\
&= |\Omega_t|^{\frac{\beta h_{t-1}-(q+1)}{2}} \times |\Omega_{t-1}|^{-\frac{\beta h_{t-1}-(q+1)}{2}}
\end{aligned}
$$

Then,

$$
\begin{aligned}
p(\Omega_{t-1}|\Omega_t, \mathcal{D}_{t-1}) \propto\ & |\Omega_t|^{\frac{\beta h_{t-1}-(q+1)}{2}} \times |\Omega_{t-1}|^{-\frac{\beta h_{t-1}-(q+1)}{2}} |I_q - {U_{t-1}^T}^{-1}\Omega_t U_{t-1}^{-1}\beta|^{\frac{(1-\beta)h_{t-1}-(q+1)}{2}} \\
& \times |\Omega_{t-1}|^{\frac{h_{t-1}-(q+1)}{2}} \times |\Omega_t|^{-\frac{\beta h_{t-1}-(q+1)}{2}} \times |\Omega_{t-1}|^{\frac{-(q+1)}{2}} \times \\
& etr\left\{-\frac{D_{t-1}^{-1}}{2}\left[\Omega_{t-1} - \beta\Omega_t\right]\right\} \\
=\ & |\Omega_{t-1}|^{-\frac{\beta h_{t-1}-(q+1)}{2}} |I_q - {U_{t-1}^T}^{-1}\Omega_t U_{t-1}^{-1}\beta|^{\frac{(1-\beta)h_{t-1}-(q+1)}{2}} \\
& \times |\Omega_{t-1}|^{\frac{h_{t-1}-(q+1)}{2}} \times |\Omega_{t-1}|^{\frac{-(q+1)}{2}} etr\left\{-\frac{D_{t-1}^{-1}}{2}\left[\Omega_{t-1} - \beta\Omega_t\right]\right\}
\end{aligned}
$$

Using the aforementioned matrix properties, it can be seen that:

$$\begin{aligned}
|I_q - {U_{t-1}^T}^{-1}\Omega_t U_{t-1}^{-1}\beta| &= |{U_{t-1}^T}^{-1}U_{t-1}^T||I_q - {U_{t-1}^T}^{-1}\beta\Omega_t U_{t-1}^{-1}||U_{t-1}U_{t-1}^{-1}| \\
&= |{U_{t-1}^T}^{-1}||U_{t-1}^T U_{t-1} - U_{t-1}^T{U_{t-1}^T}^{-1}\beta\Omega_t U_{t-1}^{-1}U_{t-1}||U_{t-1}^{-1}| \\
&= |U_{t-1}^{-1}{U_{t-1}^T}^{-1}| \times |\Omega_{t-1} - \beta\Omega_t| \\
&= |\Omega_{t-1}|^{-1} \times |\Omega_{t-1} - \beta\Omega_t|.
\end{aligned}$$

Then,

$$\begin{aligned}
p(\Omega_{t-1}|\Omega_t, \mathcal{D}_{t-1}) \propto\ & |\Omega_{t-1}|^{-\frac{\beta h_{t-1}-(q+1)}{2}}(|\Omega_{t-1}|^{-1} \times |\Omega_{t-1} - \beta\Omega_t|)^{\frac{(1-\beta)h_{t-1}-(q+1)}{2}} \\
&\times|\Omega_{t-1}|^{\frac{h_{t-1}-(q+1)}{2}} \times |\Omega_{t-1}|^{\frac{-(q+1)}{2}} etr\left\{-\frac{D_{t-1}^{-1}}{2}\left[\Omega_{t-1} - \beta\Omega_t\right]\right\}
\end{aligned}$$

For the terms $|\Omega_{t-1}|$, we have that,

$$\begin{aligned}
&|\Omega_{t-1}|^{-\frac{\beta h_{t-1}-(q+1)}{2}}|\Omega_{t-1}|^{-\frac{(1-\beta)h_{t-1}-(q+1)}{2}} \times |\Omega_{t-1}|^{\frac{h_{t-1}-(q+1)}{2}} \times |\Omega_{t-1}|^{\frac{-(q+1)}{2}} \\
&\qquad |\Omega_{t-1}|^{\frac{-\beta h_{t-1}+(q+1)-(1-\beta)h_{t-1}+(q+1)+h_{t-1}-(q+1)-(q+1)}{2}} \\
&\qquad\qquad |\Omega_{t-1}|^{\frac{(1-\beta)h_{t-1}+2(q+1)-(1-\beta)h_{t-1}-2(q+1)}{2}} = 1
\end{aligned}$$

Therefore, we conclude that,

$$p(\Omega_{t-1}|\Omega_t, \mathcal{D}_{t-1}) \propto |\Omega_{t-1} - \beta\Omega_t|^{\frac{(1-\beta)h_{t-1}-(q+1)}{2}} etr\left\{-\frac{D_{t-1}^{-1}}{2}\left[\Omega_{t-1} - \beta\Omega_t\right]\right\}.$$

Implying that,

$$(\Omega_{t-1} - \beta\Omega_t|\mathcal{D}_{t-1}) \sim W((1-\beta)h_{t-1}, D_{t-1}^{-1}) \tag{B.3}$$

or equivalently,

$$(\Omega_t - \beta\Omega_{t+1}|\mathcal{D}_t) \sim W((1-\beta)h_t, D_t^{-1}) \tag{B.4}$$

B.2 Sampling algorithm

In order to sample from the posterior distribution of Σ_t, define $\Omega_t = \Sigma_t^{-1}$ and $n_t = h_t - q + 1$, where q is the model dimensionality, a retrospective analysis, Prado and West (2010), based on the matrix-beta evolution model, is used. For $h_t > q - 1$ and a discount factor β such that $\beta > \frac{q-2}{q-1}$, at any time t looking back to times $t < n$ a sample from the joint distribution $(\Sigma|\mathcal{D}_n, \Phi_{[-\Sigma_t]})$ can be obtained from the following algorithm:

1. Set $t = n$ and sample $\Omega_n|\mathcal{D}_n, \Phi_{[-\Omega_n]} \sim W(h_n, D_n^{-1})$;
2. Set $t = t - 1$ and sample $\Upsilon_t \sim W\left((1-\beta)h_t, D_t^{-1}\right)$;
3. Make $\Omega_t = \beta\Omega_{t+1} + \Upsilon_t$;
4. If $t > 1$, go back to step 2; otherwise, the sample from $(\Omega_1, \ldots, \Omega_n|\mathcal{D}_n, \Phi_{[-\Sigma_t]})$ is complete;
5. Obtain $\Sigma_t = \Omega_t^{-1}$ for $t = 1, \ldots, n$.

APPENDIX C – Bartlett Decomposition

In this appendix we detail the Bartlett decomposition as in Kshirsagar (1959), addapting its notation to become more coherent with this thesis approach. The Bartlett decomposition is as follows:

Let $\Omega \sim W_q(n, I_q)$, where I_q is the q dimensional identity matrix and n is the degree of freedom parameter of the Wishart distribution, with $n \geq q$ and $|\Omega| > 0$. Ω can be decomposed so that:

$$\Omega \;=\; U^T U,$$

So that the random variables $u_{i,j}$, for $1 \leq i \leq j \leq q$ are independent with $u_{i,i}^2 \sim \chi_{n-i+1}^2$, or equivalently, $u_{i,i}^2 \sim G\left(\frac{n-i+1}{2}, \frac{1}{2}\right)$ for every $i = 1, \ldots q$, and $u_{i,j} \sim N(0,1)$ for every $1 \leq i \leq j \leq q$. That is:

$$U \;=\; \begin{bmatrix} u_{1,1} & u_{1,2} & \ldots & u_{1,q} \\ 0 & u_{2,2} & \ldots & u_{2,q} \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & 0 & u_{q,q} \end{bmatrix}$$

so that

$$\begin{aligned} u_{i,j}^2 &\sim G\left(\frac{n-i+1}{2}, \frac{1}{2}\right), \text{ if } i = j, \\ u_{i,j} &\sim N(0,1), \text{ if } i < j, \\ u_{i,j} &= 0, \text{ if } i > j. \end{aligned}$$

The biggest advantage of the Bartlett decomposition is that allows to sample from a Wishart distribution by sampling from univariate distributions, namely Gamma, or Chi-Squared, and Gaussian distributions.

APPENDIX D – Likelihood Functions for Presented Models

In this appendix the likelihood functions for models developed on this thesis are presented. $L(\cdot)$ represents the likelihood function, Φ represents model parameters, $p(\cdot)$ represents distribution functions. Model static parameters φ are removed from the notation for clarity.

D.1 NGSVJ

The joint distribution of all model quantities is given by

$$p(\underset{\sim}{y}, \underset{\sim}{\lambda}, \underset{\sim}{\gamma}, \underset{\sim}{J^y}|\mathcal{D}_0) = p(\underset{\sim}{y}, \underset{\sim}{\lambda}|\underset{\sim}{\gamma}, \underset{\sim}{J^y}, \mathcal{D}_0) \prod_{t=1}^{n} p_\gamma(\gamma_t) p_J(J_t^y) \tag{D.1}$$

However,

$$p(\underset{\sim}{y}, \underset{\sim}{\lambda}|\underset{\sim}{\gamma}, \underset{\sim}{J^y}, \mathcal{D}_0) = \prod_{t=1}^{n} p(y_t, \lambda_t|\mathcal{D}_{t-1}, \underset{\sim}{\gamma}, \underset{\sim}{J^y}) p(\lambda_0|\mathcal{D}_0) \tag{D.2}$$

and the join densities $p(y_t, \lambda_t|\mathcal{D}_{t-1}, \underset{\sim}{\gamma}, \underset{\sim}{J^y})$ can be decomposed as

$$p(y_t|\mathcal{D}_{t-1}, \lambda_t) p(\lambda_t|\lambda_{t-1}, \mathcal{D}_{t-1}) \tag{D.3}$$

where the dependence on $(\underset{\sim}{\gamma}, \underset{\sim}{J^y})$ was removed from all terms in Eq. D.3 for conciseness. The densities in Eq. D.3 correspond to the model specifications in Eq. 2.1 and Eq. 2.3 respectively, the jump component is composed by a Gaussian magnitude and a Bernoulli jump indicator, with priors distributions provided in Eq. 2.9 and Eq. 2.13, and mixing distribution with prior provided in Eq. 2.6. So, each model component in the distribution $p(\underset{\sim}{y}, \underset{\sim}{\lambda}, \underset{\sim}{\gamma}, \underset{\sim}{J_t^y}|\mathcal{D}_0)$ is uniquely defined.

D.2 NGSVJ-HMM

The joint distribution of all model quantities is given by

$$p(\underset{\sim}{y}, \underset{\sim}{\lambda}, \underset{\sim}{\gamma}, \underset{\sim}{J^y}, \underset{\sim}{s}|\mathcal{D}_0) = p(\underset{\sim}{y}, \underset{\sim}{\lambda}|\underset{\sim}{\gamma}, \underset{\sim}{J^y}, \underset{\sim}{s}, \mathcal{D}_0) \prod_{t=1}^{n} p_\gamma(\gamma_t) p_J(J_t) p_s(s_t) \tag{D.4}$$

However,

$$p(\underset{\sim}{y}, \underset{\sim}{\lambda}|\underset{\sim}{\gamma}, \underset{\sim}{J_t^y}, \underset{\sim}{s}, \mathcal{D}_0) = \prod_{t=1}^{n} p(y_t, \lambda_t|\mathcal{D}_{t-1}, \underset{\sim}{\gamma}, \underset{\sim}{J_t^y}, \underset{\sim}{s}) p(\lambda_0|\mathcal{D}_0) \tag{D.5}$$

and the join densities $p(y_t, \lambda_t|\mathcal{D}_{t-1}, \underset{\sim}{\gamma}, \underset{\sim}{J_t^y}, \underset{\sim}{s})$ can be decomposed as

$$p(y_t|\mathcal{D}_{t-1}, \lambda_t)p(\lambda_t|\lambda_{t-1}, \mathcal{D}_{t-1}) \tag{D.6}$$

where the dependence on $(\underset{\sim}{\gamma}, \underset{\sim}{J_t^y}, \underset{\sim}{s})$ was removed from all terms in Eq. D.6 for conciseness. The densities in Eq. D.6 correspond to the model specifications in Eq. 4.1 and Eq. 4.2 respectively, the jump component is composed by a Gaussian magnitude and a Bernoulli jump indicator, with priors distributions provided in Eq. 2.9 and Eq. 2.13, mixing distribution with prior provided in Eq. 2.6, and states probabilities are defined in **??**. So, each model component in the distribution $p(\underset{\sim}{y}, \underset{\sim}{\lambda}, \underset{\sim}{\gamma}, \underset{\sim}{J_t^y}, \underset{\sim}{s}|\mathcal{D}_0)$ is uniquely defined.

D.3 MSVJM and MSVJB

The joint distribution of all model quantities is given by

$$p(\underset{\sim}{y}, \underset{\sim}{\theta}, \underset{\sim}{\Sigma}, \underset{\sim}{\gamma}, \underset{\sim}{\omega}, \underset{\sim}{J^y}|\mathcal{D}_0) = p(\underset{\sim}{y}, \underset{\sim}{\theta}, \underset{\sim}{\Sigma}|\underset{\sim}{\gamma}, \underset{\sim}{\omega}, \underset{\sim}{J^y}, \mathcal{D}_0) \prod_{t=1}^{n} p_\gamma(\gamma_t)p_\omega(\omega_t)p_J(J_t^y) \tag{D.7}$$

However, $p(\underset{\sim}{y}, \underset{\sim}{\theta}, \underset{\sim}{\Sigma}|\underset{\sim}{\gamma}, \underset{\sim}{\omega}, \underset{\sim}{J^y}, \mathcal{D}_0) \prod_{t=1}^{n} p_\gamma(\gamma_t) =$

$$\prod_{t=1}^{n} p(y_t, \theta_t, \Sigma_t|\mathcal{D}_{t-1}, \underset{\sim}{\gamma}, \underset{\sim}{\omega}, \underset{\sim}{J^y})p(\theta_0|\mathcal{D}_0)p(\Sigma_0|\mathcal{D}_0) \tag{D.8}$$

and the join densities $p(y_t, \theta_t, \Sigma_t|\mathcal{D}_{t-1}, \underset{\sim}{\gamma}, \underset{\sim}{\omega}, \underset{\sim}{J^y})$ can be decomposed as

$$p(y_t|\mathcal{D}_{t-1}, \theta_t, \Sigma_t)p(\Sigma_t|\Sigma_{t-1}, \mathcal{D}_{t-1}, \underset{\sim}{\theta})p(\theta_t|\theta_{t-1}, \mathcal{D}_{t-1}) \tag{D.9}$$

where the dependence on $(\underset{\sim}{\gamma}, \underset{\sim}{\omega}, \underset{\sim}{J^y})$ was removed from all terms in Eq. D.9 for conciseness. The densities in Eq. D.9 correspond to the model specifications in Eq. 5.1 to Eq. 5.3 respectively for MSVJM, and Beta-Bartlett evolution for $p(\Sigma_t|\Sigma_{t-1}, \mathcal{D}_{t-1}, \underset{\sim}{\theta})$ described in Chapter 6 for MSVJB, the jump component is composed by a Multivariate Gaussian magnitude and a Bernoulli jump indicator, with priors distributions provided in Eq. 5.16 and Eq. 5.20, and mixing distribution with prior provided in Eq. 5.12. In this model, $\underset{\sim}{\omega}$ is defined by a single discount factor. So, each model component in the distribution $p(\underset{\sim}{y}, \underset{\sim}{\theta}, \underset{\sim}{\Sigma}, \underset{\sim}{\gamma}, \underset{\sim}{\omega}, \underset{\sim}{J^y}|\mathcal{D}_0)$ is uniquely defined.

APPENDIX E – MCMC Results for Multivariate Model

In this appendix MCMC chains for static parameters and density plots for simulations and applications of MSVJ will be presented.

Figure 59 shows MCMC chains for static parameters for simulation in section 5.4. Convergence was evaluated by graphical methods. Similar results are obtained across alternative simulation scenarios, using different initial values for the parameters. Figure 60 shows density plot for static parameters.

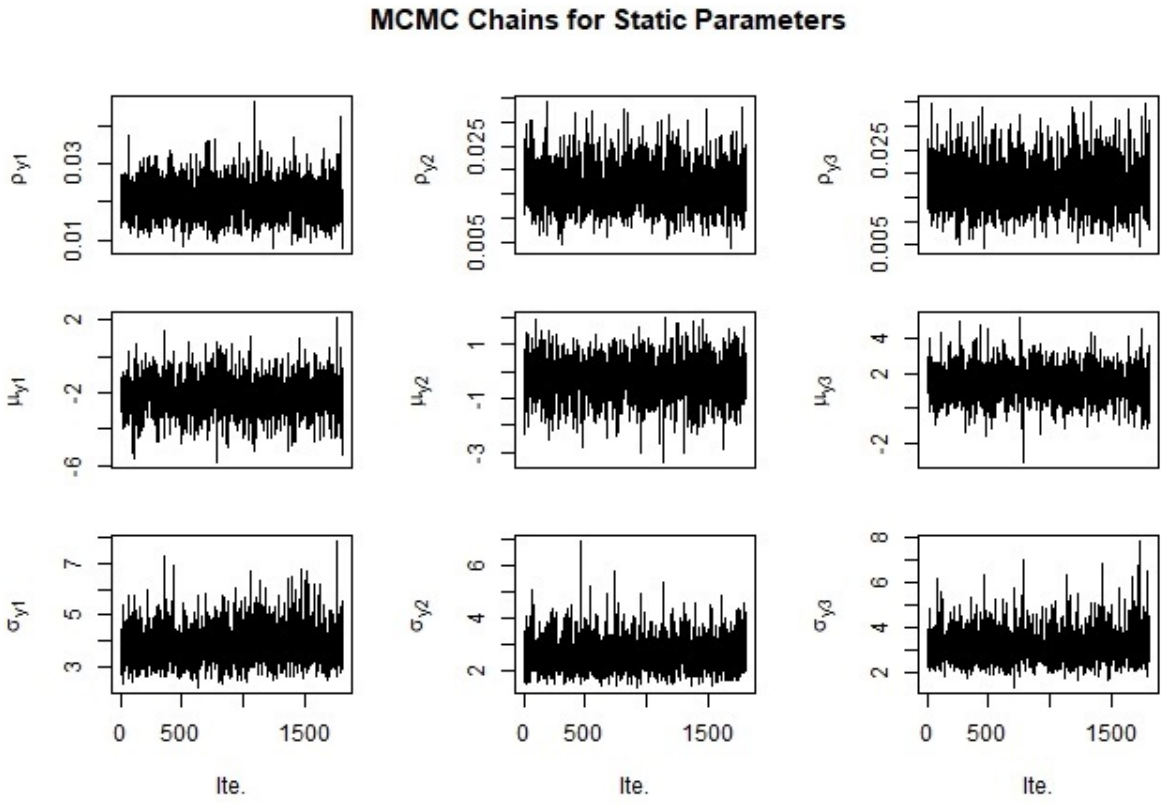


Figure 59 – MCMC chain convergence: trace plot for the static parameters of the model for simulated time series.

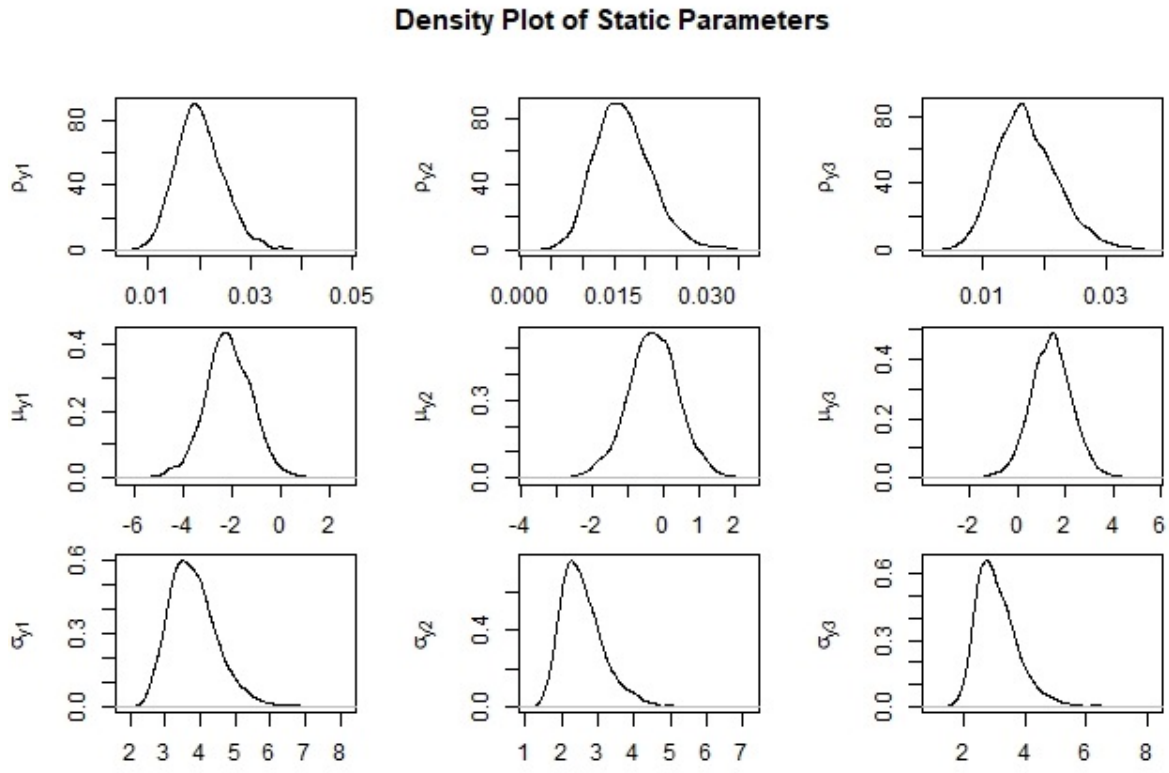


Figure 60 – Density plot for the static parameters of the model for simulated time series.

Figures 61,62 and 63 shows MCMC chains for static parameters in Section 5.5.2. Convergence was achieved for all static parameters. Figures Figures 64, 65 and 66 shows density plot for static parameters.

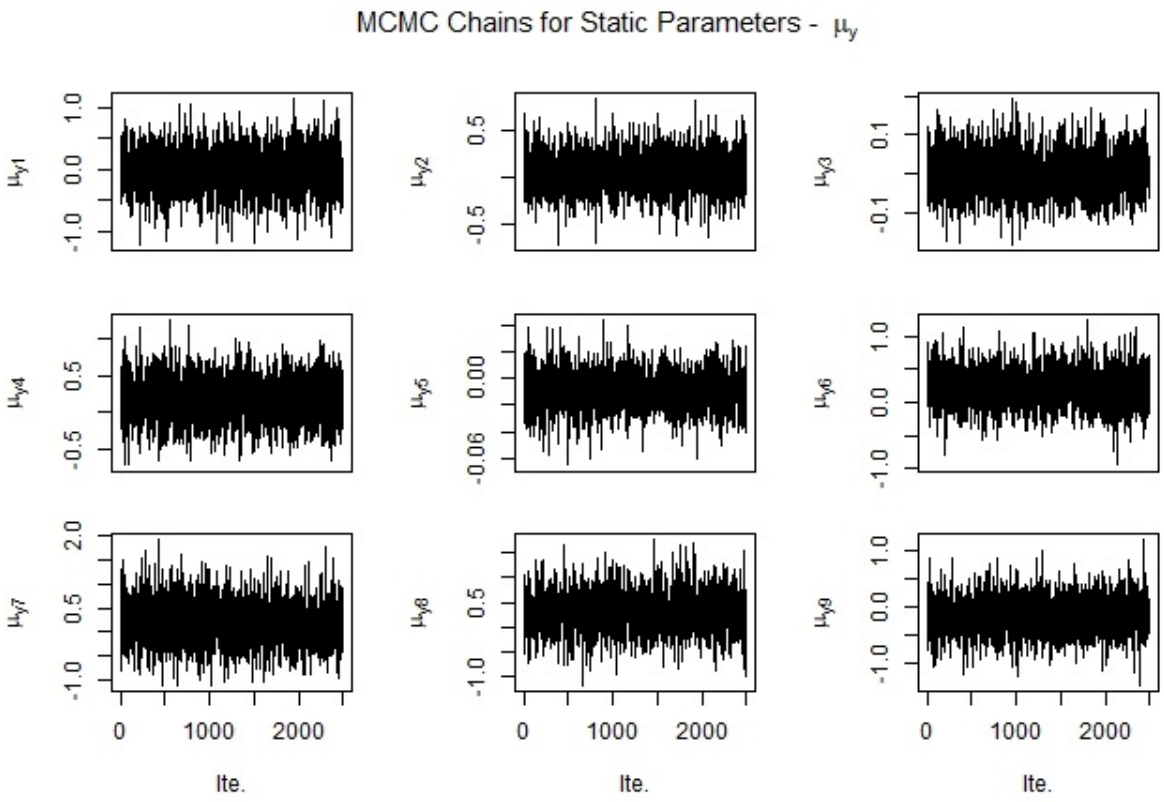


Figure 61 – MCMC chains for effective sample of static parameters.

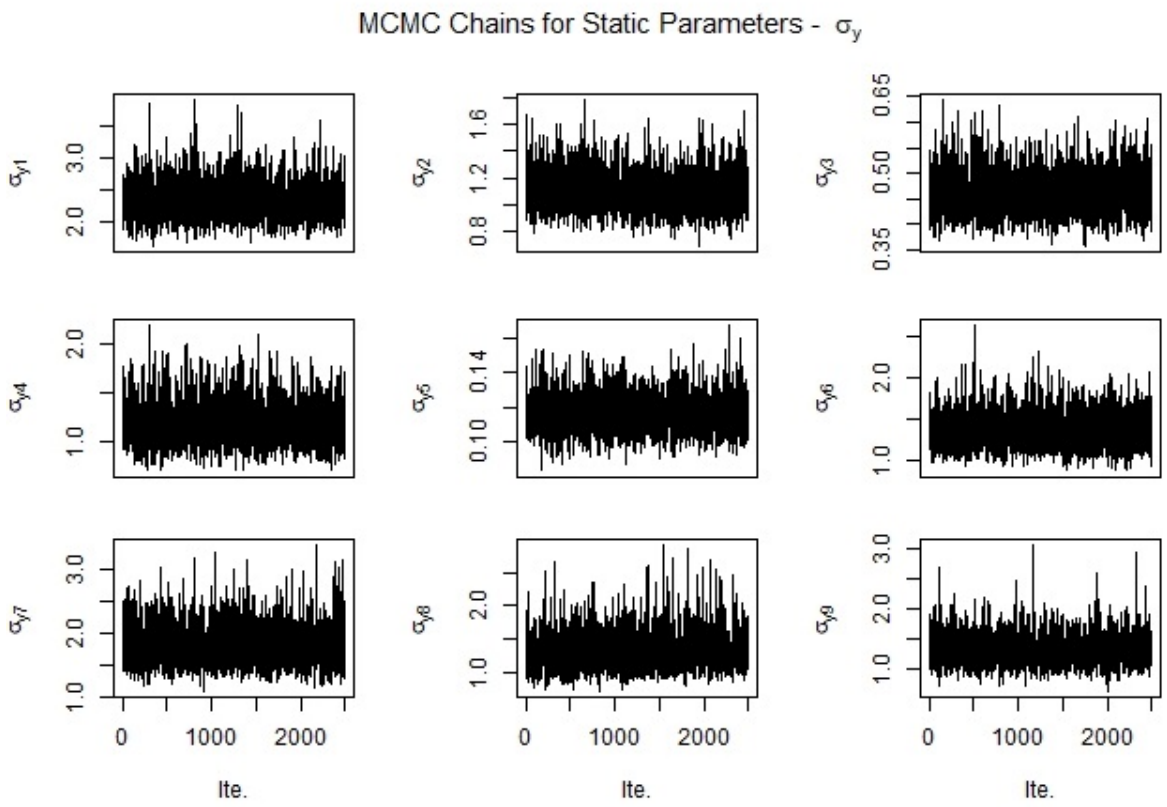


Figure 62 – MCMC chains for effective sample of static parameters.

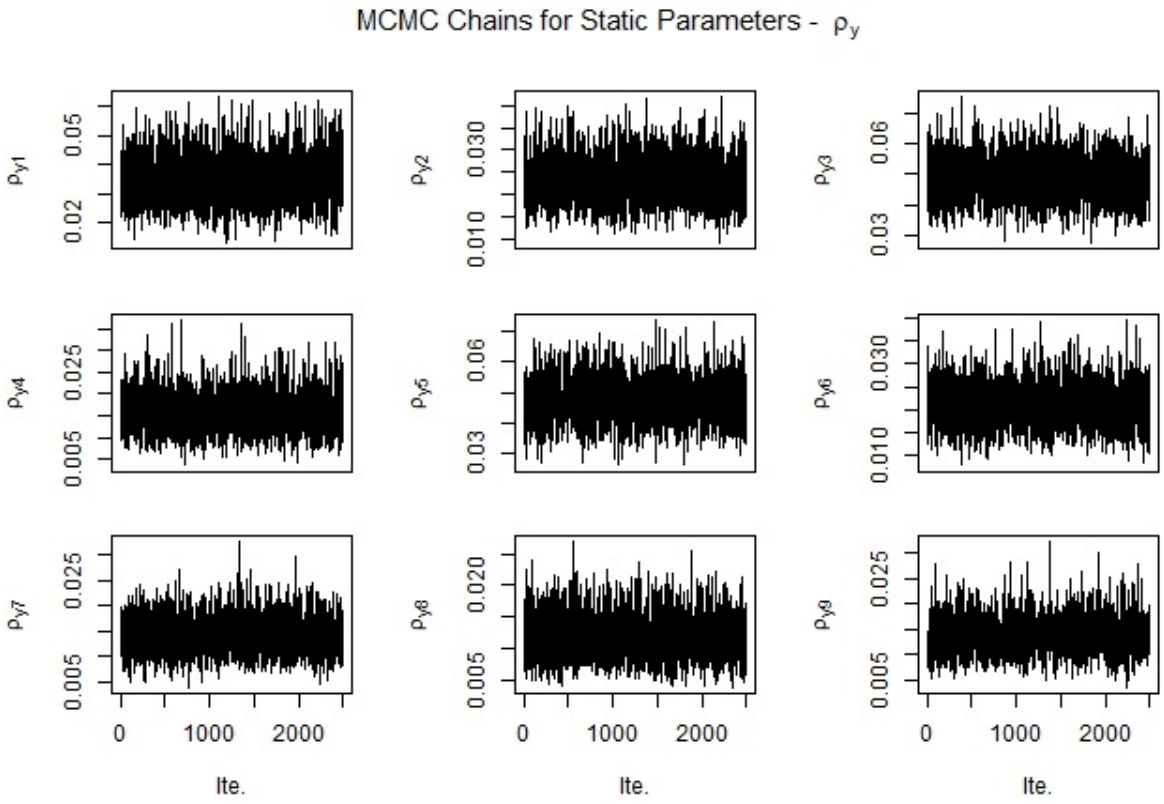


Figure 63 – MCMC chains for effective sample of static parameters.

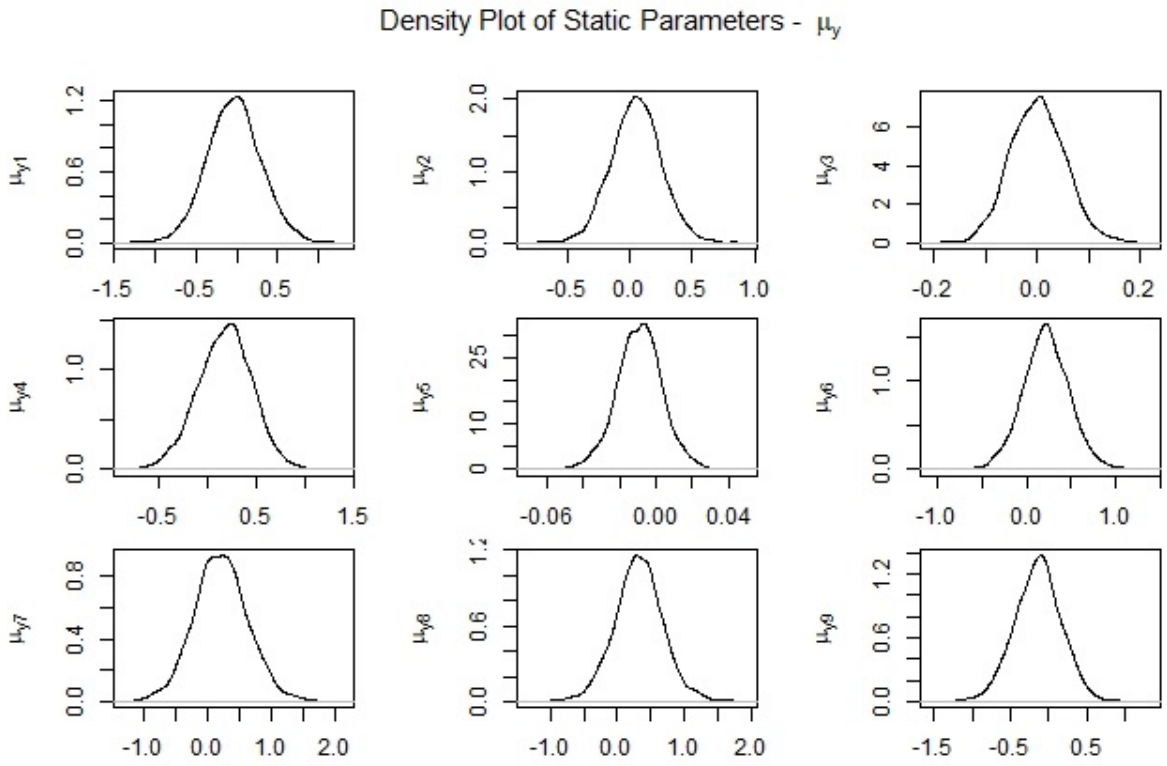


Figure 64 – Density plot for effective sample of static parameters.

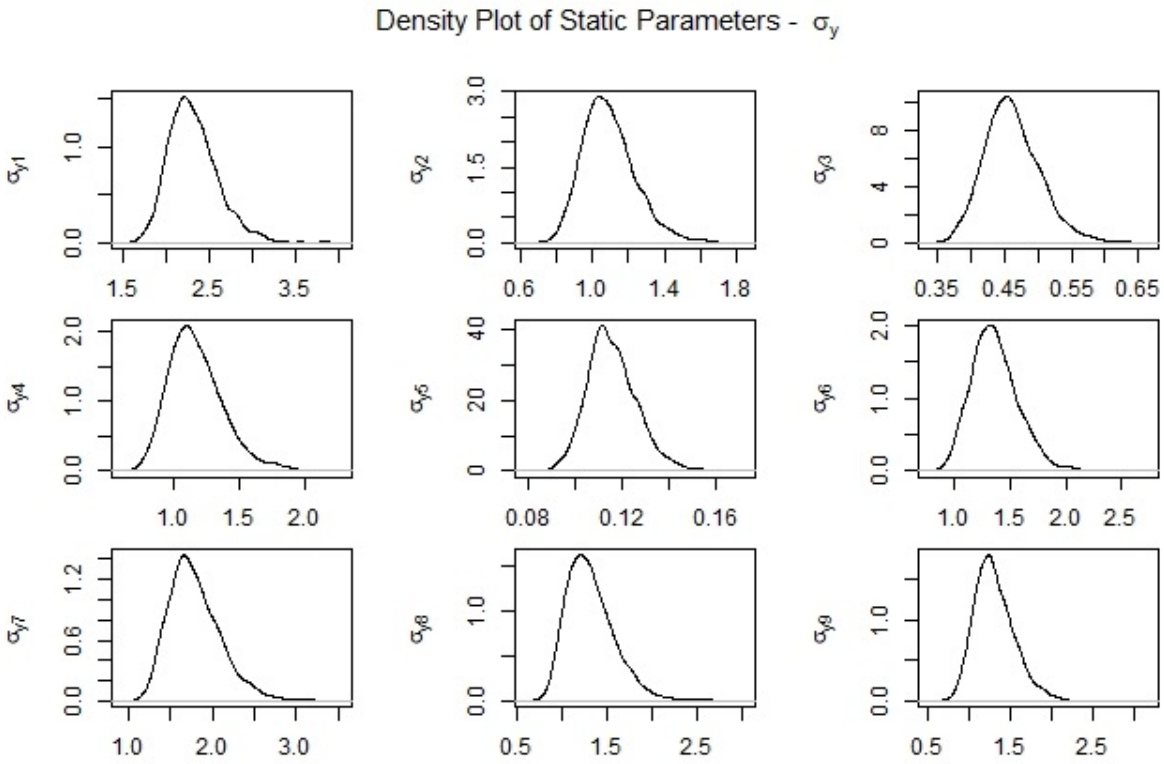


Figure 65 – Density plot for effective sample of static parameters.

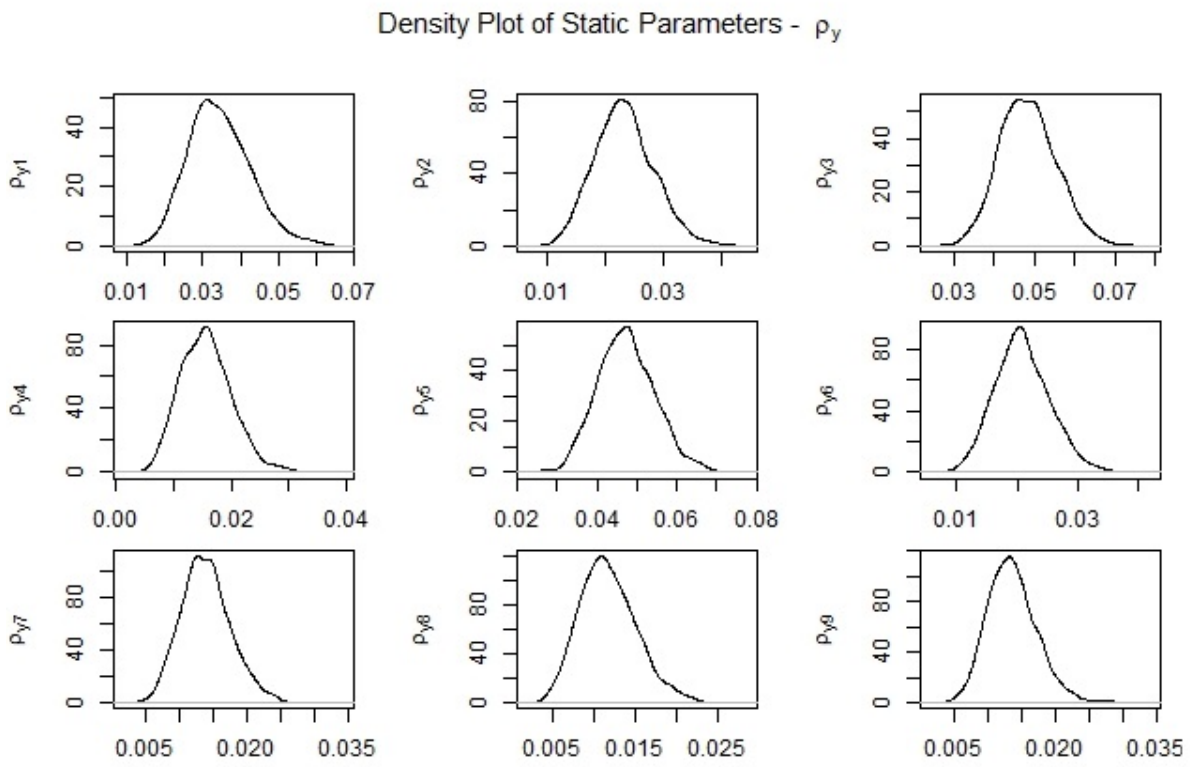


Figure 66 – Density plot for effective sample of static parameters.

Figures 67 to 72 shows MCMC chains and density plots for static parameters in Section 6.3. Convergence was achieved for all static parameters.

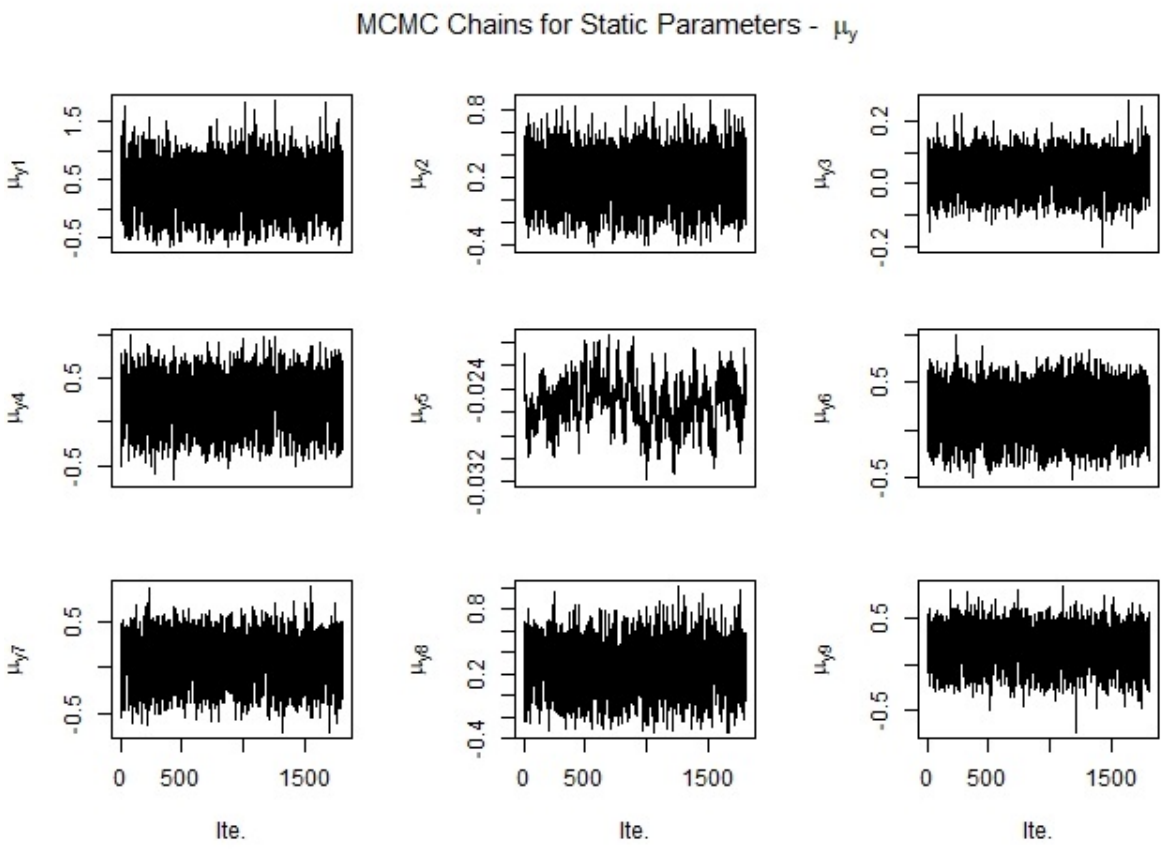


Figure 67 – MCMC chains for effective sample of static parameters.

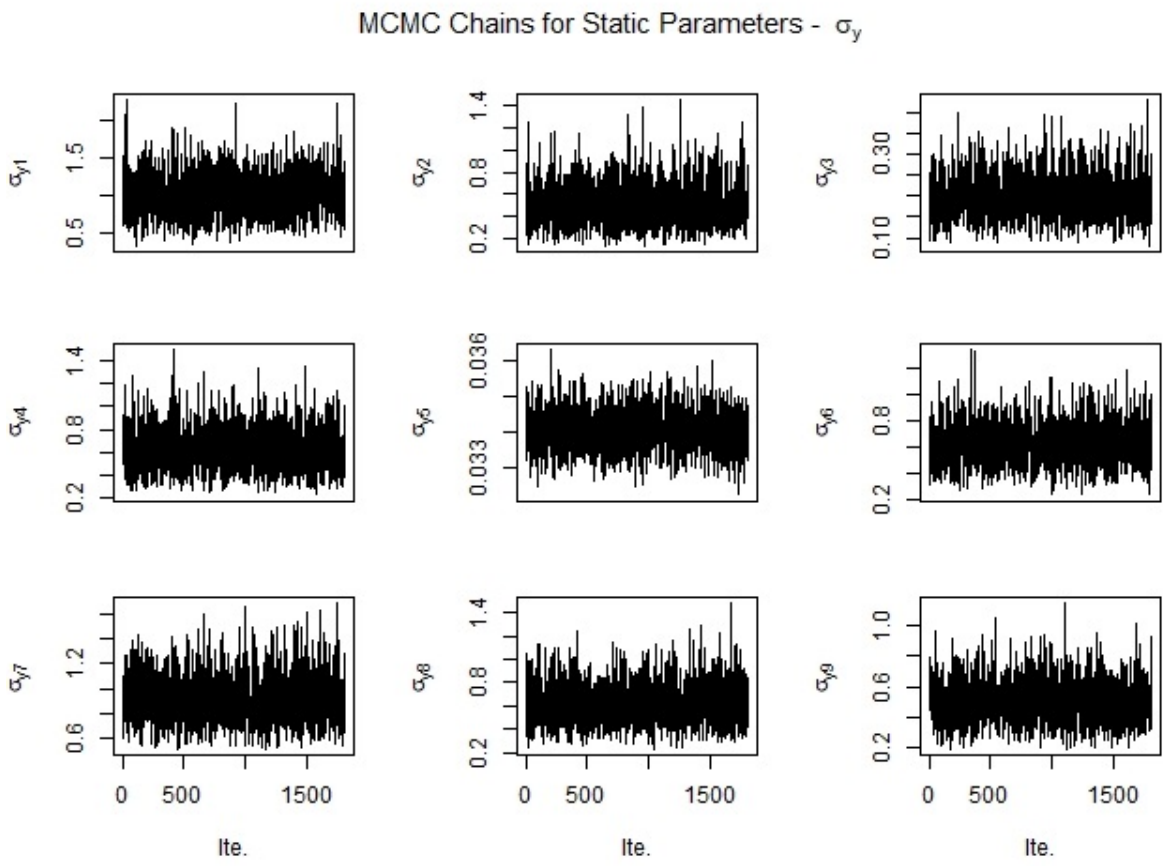


Figure 68 – MCMC chains for effective sample of static parameters.

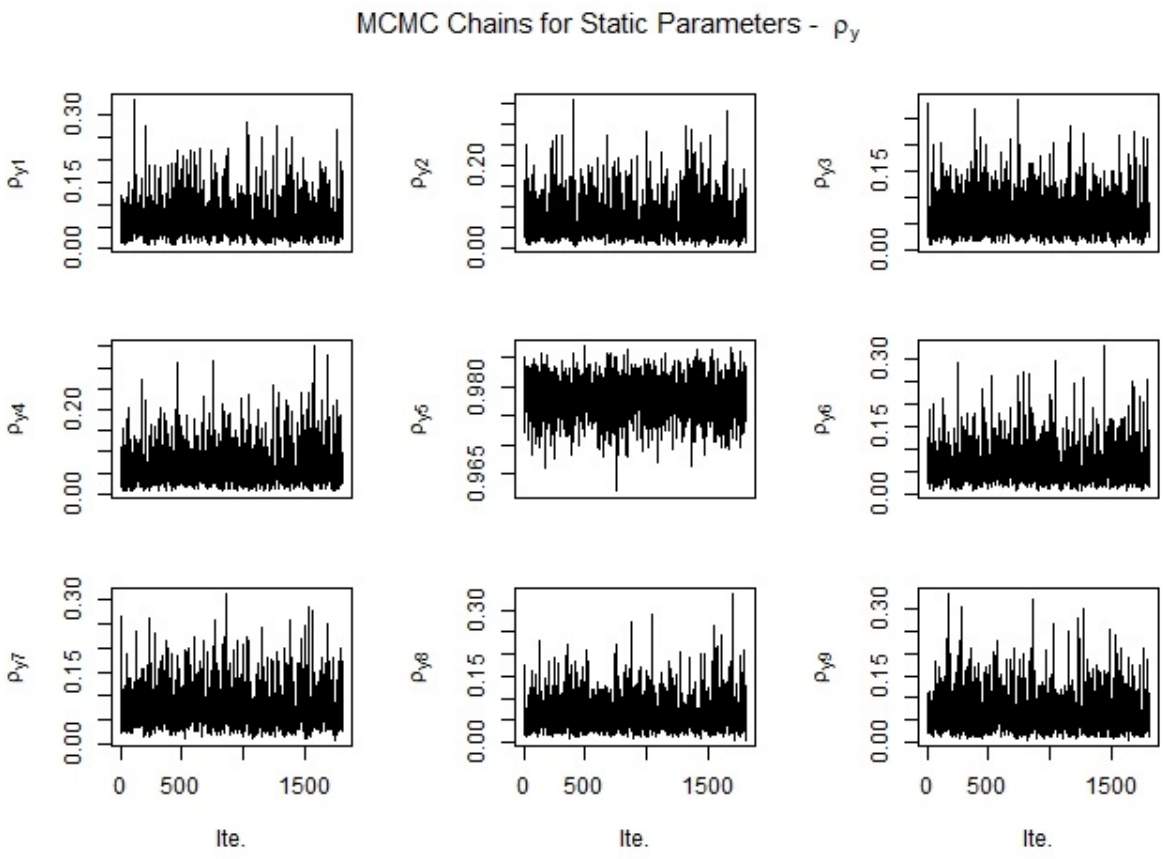


Figure 69 – MCMC chains for effective sample of static parameters.

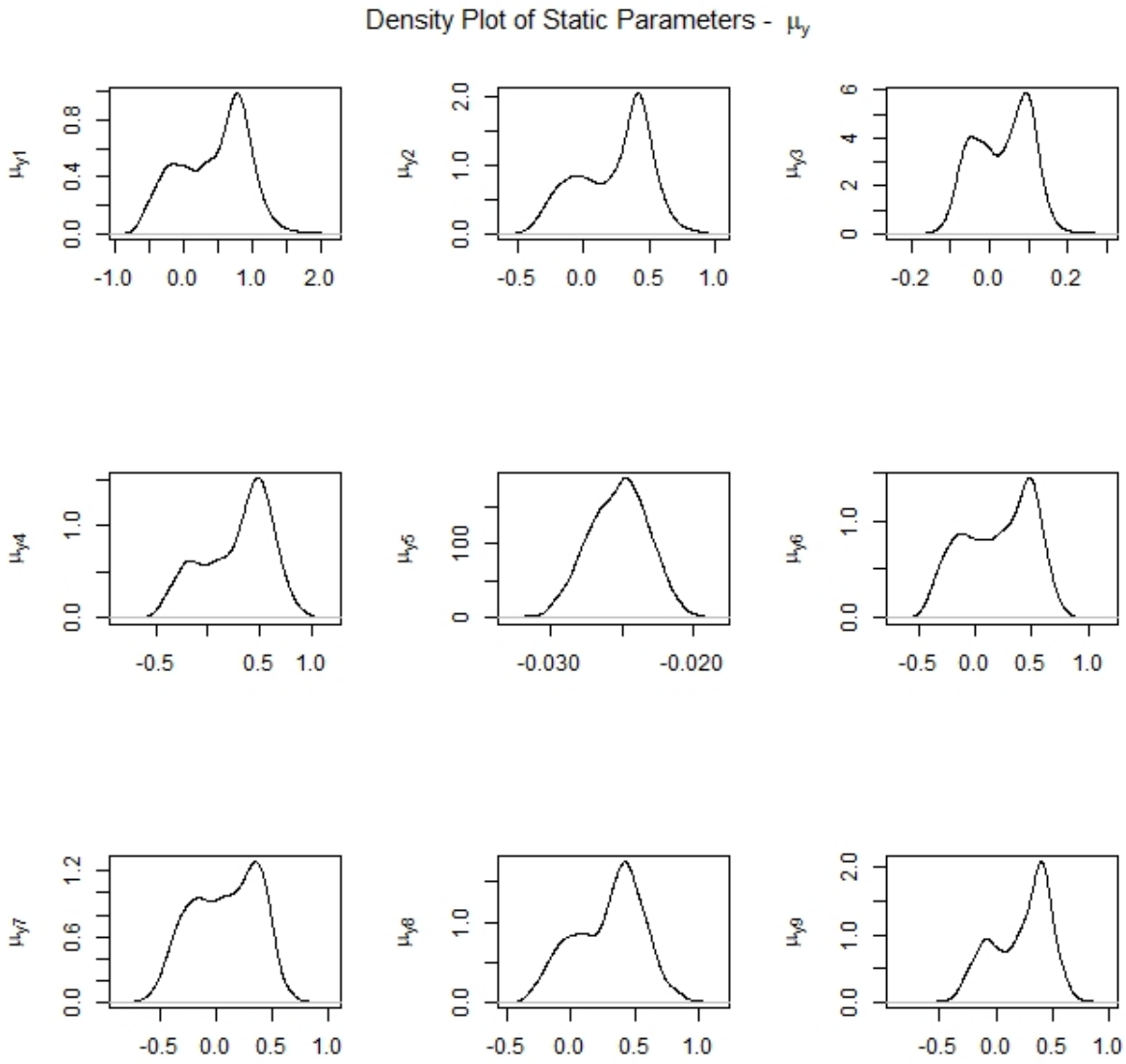


Figure 70 – Density plot for effective sample of static parameters.

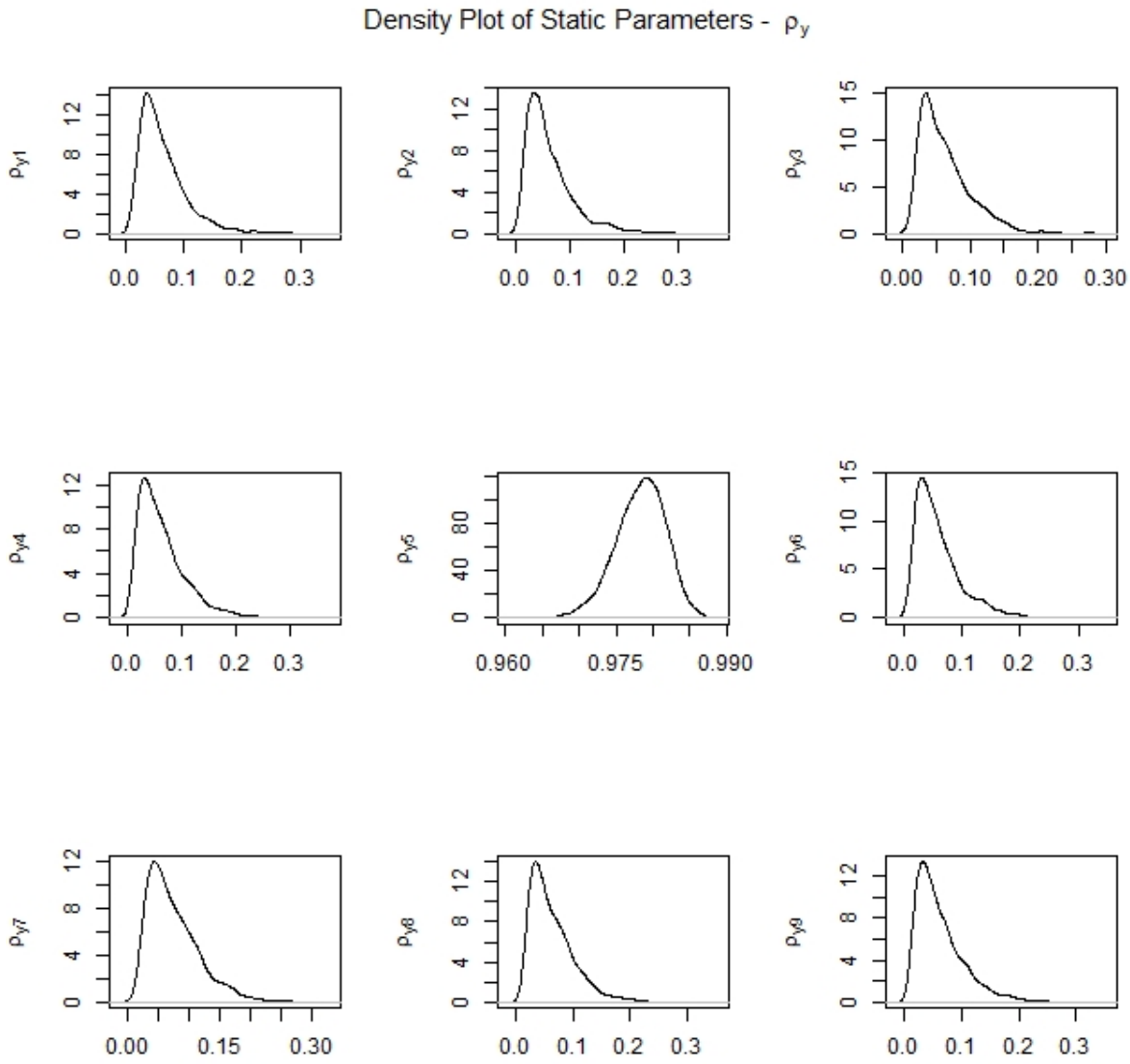


Figure 71 – Density plot for effective sample of static parameters.

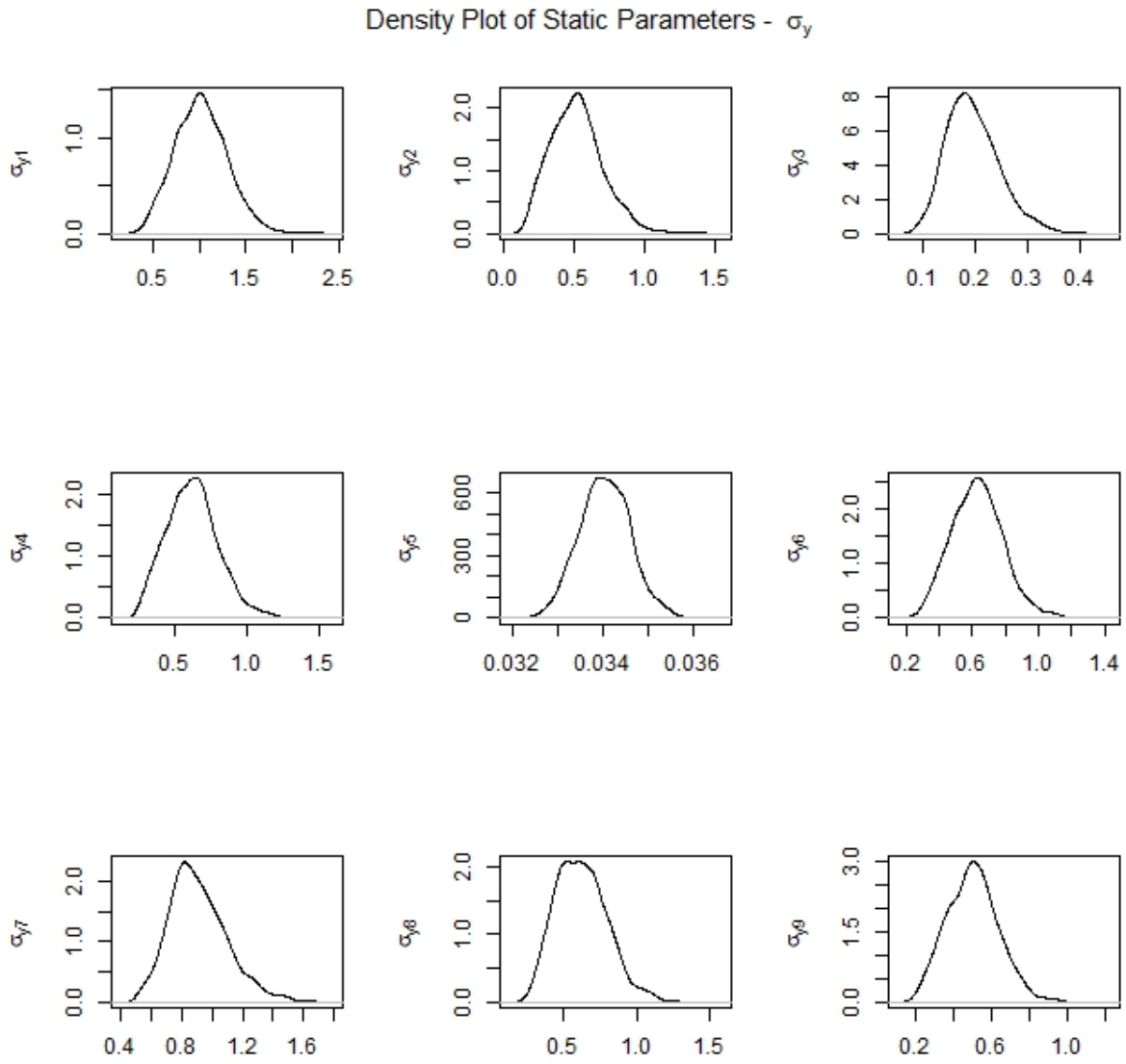


Figure 72 – Density plot for effective sample of static parameters.