THE UNIVERSITY OF SYDNEY

On Gegenbauer long memory stochastic volatility models: A Bayesian Markov chain Monte Carlo approach with applications

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

in the

School of Mathematics and Statistics
Faculty of Science

2018

Abstract

This thesis begins by developing a time series model which has generalised (Gegenbauer) long memory in the mean process with stochastic volatility errors where each process is assumed to have Gaussian errors. We subsequently develop and derive a new Bayesian posterior simulator that couples advanced posterior maximisation techniques, as well as traditional latent stochastic volatility estimation procedures. Details are provided on the estimation process, data simulation, and out of sample performance measures. Several rigorous simulation studies are conducted and verified on the simulator for in and out of sample behaviour. Further, the goodness of fit of the generalised long memory model is compared to the standard long memory model by considering two empirical studies on the US CPI and the US ERP. This model provides a distinct advantage by measuring the long memory attributes in the mean process, whilst also estimating the daily time varying volatility of the error process. The long memory process in particular is generalised so that it encompasses the standard long memory case, as well as the ARMA family.

These findings are then extended to a Gegenbauer long memory stochastic volatility model with leverage and a bivariate Student's t-error distribution to describe the innovations of the observation and latent volatility jointly, with applications to Cryptocurrency time series. The main advantage of pursuing such a model is incorporating the robustness of the Student's t-distribution, and leverage effects to address the deep rooted characteristics found in Cryptocurrencies. The mixing variables in the scale mixture representation of the Student's t-distribution are able to capture outliers which are the occasional jumps found in Cryptocurrency return series'. Understanding the leverage effect and jump behavior helps to evaluate their investability. To date, the literature remains underdeveloped in understanding the properties

of Cryptocurrencies and is a promising area to pursue. In order to do this, a rigorous in-sample simulation study is conducted to assess the performance of the model with nested alternatives and applied to study the behavior of many Cryptocurrencies - in particular Bitcoin. The data analysis is initiated with a broad scope of 114 Cryptocurrencies and then a more detailed understanding of the five most popular Cryptocurrencies and followed up with a specific focus on Bitcoin. The model parameters are estimated using a Bayesian approach and sampled via MCMC. In order to implement model selection, the DIC is used. This is then compared with many popular models including those commonly used in industry. The models are applied to Value-at-Risk (VaR) forecasts and several measures are used to assess the forecast performance.

Finally, it is found that Cryptocurrencies do indeed show highly distinct behaviours that are not present in fiat currencies, and thus require specialized analysis. Cryptocurrencies as of late have commanded global attention on a number of fronts. Most notably, their variance properties are known for being notoriously wild, unlike their fiat counterparts. The third part of this thesis highlights some stylized facts about the variance measures of Cryptocurrencies using the logarithm of daily return range and relates these results to their respective cryptographic designs such as intended transaction speed. This final model in which we arrive to includes even more additional features including buffered autoregressive regime effects as well as jumps. The advantages of including such effects into a more comprehensive model is able to discern between the extreme volatility of Cryptocurrencies as jumps, or from the stochastic volatility itself. The results favor instantaneous oscillatory long run autocorrelations over standard long run autocorrelation filters to model the log daily return range. The overarching implication of this result is the volatility of Cryptocurrencies can be better understood and measured via the use of fast moving autocorrelation functions, as opposed to smoothly decaying functions for fiat currencies.

Acknowledgements

First and foremost, I am thankful to God for giving me the strength to complete my PhD.

Secondly, I would like to thank my supervisor A/Prof. Jennifer S.K. Chan for her constant support, guidance and faith in me throughout my PhD. I have been truly fortunate to have her supervision and mentorship throughout my PhD. I feel very fortunate to have an excellent statistician and a truly nice person as my supervisor. I would also like to thank my associate supervisor A/Prof. Shelton Peiris for his mentorship and for providing me with helpful discussions and directing me to relevant references which assisted me throughout my research.

And last but not least, my wife Cherrie for supporting me throughout all of my PhD years.

Authorship attribution statement

The research in this thesis has resulted in the following publications:

1. Phillip, Andrew, Jennifer S. K. Chan, and Shelton Peiris. 2018. "Bayesian Estimation of Gegenbauer Long Memory Processes with Stochastic Volatility: Methods and Applications." **Studies in Nonlinear Dynamics** & **Econometrics**. In press. (Chapter 3)

2. Phillip, Andrew, Jennifer S. K. Chan, and Shelton Peiris. "On generalised bivariate Student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of Cryptocurrencies with a focus on Bitcoin." **Econometrics and Statistics**. Revision invited on $30/04/2018$. Revision submitted on $30/05/2018$. (Chapter 5)

3. Phillip, Andrew, Jennifer S. K. Chan, and Shelton Peiris. 2018. "A new look at Cryptocurrencies." **Economics Letters** Volume 163, February 2018:6-9 (Chapter 6)

4. Phillip, Andrew, Jennifer S. K. Chan, and Shelton Peiris. 2018. "On long memory effects in the volatility measure of Cryptocurrencies." **Finance Research Letters**. In press. (Chapter 8)

These publications have been included as chapters within this thesis, where the relevant chapter has been stated in parantheses above. The majority of the analysis in every publication listed above was performed by Andrew Phillip. The manuscripts were all written by Andrew Phillip and edited by Andrew Phillip and the respective co-authors.

Statement of originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Signed: ANDREW PHILLIP

Contents

List of Figures

List of Tables

List of Abbreviations

ACF	**A**uto **C**orrelation **F**unction
AI	**A**rtificial **I**ntelligence
APRA	**A**ustralian **P**rudential **R**egulation **A**uthority
AR	**A**uto **R**egressive
ARCH	**A**uto **R**egressive **C**onditional **H**eteroskedasticity
ARMA	**A**uto **R**egressive **M**oving **A**verage
BAR	**B**uffered **A**uto **R**egressive
BNC	**B**rave **N**ew **C**oin
BTC	**B**itcoin
CPI	**C**onsumer **P**rice **I**ndex
DIC	**D**eviance **I**nformation **C**riertion
EP	**E**xponential **P**ower
ERP	**E**quity **R**isk **P**remium
ETH	**E**thereum
(G)ARCH	**(G)**eneralized **A**uto **R**egressive **C**onditional **H**etereoskedasticity
(G)ARFIMA	**(G)**egenbauer **A**uto **R**egressive **F**ractionally **I**ntegrated **M**oving
GFC	**G**lobal **F**inancial **C**risis
GLR	**G**egenbauer **L**og **R**ange
GMA	**G**egenbauer **M**oving **A**verage
HC	**H**eavy **C**ommon tails
HAR	**H**eterogeneous **A**uto **R**egressive
HO	**H**eavy **O**bservational tails
IG	**I**nverse **G**amma
JBAR	**J**ump **B**uffered **A**uto **R**egressive
KF	**K**alman **F**ilter
LM	**L**ong **M**emory
LTC	**L**i**T**e **C**oin
LVG	**L**everage
MA	**M**oving **A**verage
MAP	**M**aximum **A** **P**osteriori
MCL	**M**onte **C**arlo **L**ikelihood
MCMC	**M**arkov **c**hain **M**onte **C**arlo
MH	**M**etropolis **H**astings
(Q)MLE	**(Q)**uasi **M**aximum **L**ikelihood **E**stimation
RV	**R**ealised **V**olatility
SDE	**S**tochastic **D**ifferential **E**quation
STAR	**S**mooth **T**ransition **A**uto**R**egressive
SMN	**S**cale **M**ixture of **N**ormal
SMU	**S**cale **M**ixture of **U**niform
SV	**S**tochastic **V**olatility
TAR	**T**hreshold **A**utogressive **R**egressive
TSV	**T**hreshold **S**tochastic **V**oltatility

VaR	**V**alue **a**t **R**isk
VG	**V**ariance **G**amma
VIX	CBOE **V**olatility **I**nde**X**
VOMR	**V**olatility **O**scillation **M**emory **R**atio
XMR	Monero coin
XRP	**R**ipple

Chapter 1

Introduction

“Begin at the beginning," the King said, very gravely,“and go on ’till you come to the end: then stop."

Lewis Carroll, Alice in Wonderland

1.1 Thesis motivation

This thesis is motivated by the intricacies of measuring risk, which have a long and rich history. The initial developments of risk quantification came about due to portfolio performance being judged purely on portfolio return. This however changed when Harry Markowitz published his groundbreaking seminal paper titled “Portfolio Selection" (Markowitz, 1952), which argued that a fund manager’s performance should not only be judged by return, but also by risk. For example, shares are a riskier investment than bonds and should therefore provide a higher return. The concept of risk was indeed a vague concept, and so Markowitz used a simplified measure, the variance of returns (or volatility). Together with other insights, such as portfolio diversification, this academic movement became known as ‘modern portfolio theory’. It was this pioneering effort that enabled Markowitz to develop a model that investors could measure the trade-offs they faced between risk and return and by doing so, he ensured volatility to be a proxy for risk.

The impact of measuring risk was so profound that in 1973 three academics named Fisher Black, Robert Merton and Myron Scholes published a model to efficiently calculate the value of options based on volatility (Black and Scholes, 1973). This became popularized into the classical Black-Scholes model, and has become the cornerstone

of option pricing until today. The model however assumes the underlying volatility is constant over the life of the derivative, and unaffected by the changes in the price level of the underlying security. This of course is far from reality as there is overwhelming evidence that volatility is not constant across time.

As the financial industry grew, volatility modelling became more important in all aspects of money management. As this was happening, a young academic named Stephen Taylor had suggested a more robust volatility measuring tool that took into account the stochastic nature of the volatility for a financial time series, known as the Stochastic Volatility (SV) model. Research showed that prices of European call options on currencies based on SV models were far more accurate than those based on the simplistic Black-Scholes model.

Further risk management techniques emanated as a result of major losses by financial institutions in the 1980's. The most famous example was after the Black Monday crash of 1987, when JP Morgan Chase Manahattan chairman at the time, Sir Dennis Weatherstone, ordered staff to provide him with a daily report outlining how much value was at risk in any single trading day. This ultimately became known as Value at Risk (VaR), and subsequently became wide spread within the financial industry. The initial proponents of VaR used the most simplistic methodology, known as historical VaR which relies on a pre-specified number of previous observations to serve as the future return distribution. From this point, the race to develop a more sophisticated version of VaR was intensified, with financial services firms hiring statisticians and computer scientists to develop even more sophisticated VaR methodologies.

Additionally, in the 1960s Benoit Mandelbrot was intensely working on mathematical finance, in particular, on his "Random walk hypothesis" (Mandelbrot and Ness, 1968). He proposed that financial assets incrementally innovate from one point to another, and from this work, he came up with the idea of long memory (LM). Long memory refers to the long range dependence between various points in a time series and is characterized by the fact that decay of such dependence is slower than an exponential decay, usually in the form of hyperbolic decay. Long memory time series are rarely introduced with more robust volatility measuring techniques such as the SV model. The intertwining of two great concepts provides opportunities for great model exploration, and therefore a higher level of data extrapolation. Although theoretically and intuitively pleasing, the SV model was unable to be efficiently estimated

with the level of computing power at the time, and the estimation methodology was not clear. It was only until Kim, Shephard, and Chib (1998) provided an efficient Bayesian MCMC method to estimate the SV model that arguably lead to its popularity.

The financial industry is again going through rapid transformation with the advent of digital currencies which early pioneers never anticipated their methods would be applied to. As Cryptocurrencies are emerging, will it gradually replace fiat currency? What is the implication to the world economic order? To answer these questions, there is now a resurgence for more sophisticated models to understand these newly created asset classes. There is indeed overwhelming evidence that Cryptocurrencies display wild volatility and leverage effects, which refers to the negative correlation between current returns and future volatility. Further, there is evidence to suggest the presence of jump diffusion and buffer threshold type effects. These features and their root causes challenge statisticians and econometricians to enhance model structures for measuring Cryptocurrency risk.

The current literature has discussed some features of Cryptocurrencies. Urquhart (2017) stated that the Cryptocurrency market is still in its infancy and is inefficient. Some properties of Cryptocurrencies have also been explained including price clustering (Urquhart, 2017), generalised autoregressive conditional hetereoscedastic (GARCH) effects (Katsiampa, 2017) and standard long run autocorrelation (Jiang, Nie, and Ruan, 2017; Lahmiri, Bekiros, and Salvi, 2018). Lastly, broader risk management issues for Cryptocurrencies were also considered by Hotz-Behofsits, Huber, and Zörner (2018), Catania, Grassi, and Ravazzolo (2018), and Hencic and Gouriéroux (2015).

From the early pioneers of charting financial time series in the 1950s, through to the highly sophisticated artificial intelligence (AI) algorithms currently being deployed by hedge funds in the thousands, there has always been an increasing movement towards greater sophistication to time series estimation. We are deeply motivated by these events to make further contributions to the level of understanding, sophistication and estimation of risk modelling.

1.2 Background

This section provides some basic definitions and fundamental properties of important statistical models considered in this thesis. Literature reviews on SV models and long memory models can be found in Chapters 3, 5 and 8.

1.2.1 Stochastic volatility

Over the last few decades, there has been an increased uptake in the interest of the dynamic nature of volatility. The developmental beginnings of dynamic volatility models initiated from the Black-Scholes rubric in which the stock price S_t is assumed to be geometric Brownian motion and is described by the following stochastic differential equation (SDE)

$$dSt = \mu S_t dt + \sigma S_t dW t. \tag{1.1}$$

The parameter W_t is a standard Brownian motion process at time t, and the expected growth rate μ and the volatility σ are both assumed to be constant. There is however overwhelming evidence from time-series and option price data that indicates the volatility parameter σ in (1.1) should be allowed to vary in time ((G)ARCH model), or vary stochastically in time (SV model).

A widely used class of models for the condition volatility is the ARCH specification of Engle (1982) and the GARCH specification of Bollerslev (1986). These models are able to characterize the stylized features of volatility. The most notable feature of (G)ARCH models is the conditional variance is a deterministic function of previously observed conditional variances and past values of the return itself.

The ARCH(q) model of Engle (1982) is

$$\begin{aligned} y_t &= \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, \sigma_{1,t}^2), \\ \sigma_{1,t}^2 &= \alpha_0^{(1)} + \sum_{i=1}^{q} \alpha_i^{(1)} \varepsilon_{t-i}^2, \end{aligned}$$

and the GARCH(p, q) model of Bollerslev (1986) is

$$\begin{aligned} y_t &= \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, \sigma_{2,t}^2), \\ \sigma_{2,t}^2 &= \mu^{(1)} + \sum_{i=1}^{p} \alpha_i^{(2)} \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{2,t-i}^2. \end{aligned}$$

These two models are in stark contrast to the SV model first proposed by Taylor (1986) which is given by

$$y_t = \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, e^{h_t}), \tag{1.2}$$

$$h_t = \alpha + \beta(h_{t-1} - \alpha) + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma^2), \tag{1.3}$$

where y_t is the return series which is observed at time t and h_t is the latent process which governs the volatility of y_t. The innovations η_t of h_t is assumed to be a Gaussian white noise process with variance σ^2. This 'volatility of volatility' parameter σ^2 intuitively indicates the uncertainty about future volatility and it can be assumed that data with lower estimates of σ^2 have better forecasting power. If the volatility persistence parameter $\beta \to 1$, then the models become explosive and non-stationary. However, it is found that in most financial time series, β is typically close to 1 (Kim, Shephard, and Chib, 1998). Further, we note that as $\beta \to 1$ and $\sigma^2 \to 0$, then h_t becomes constant over time and therefore the model is homoscedastic.

In comparison to the (G)ARCH models which explains the volatility process in a deterministic equation, the SV model is a more robust representation of the real world. SV models have gradually emerged as a successful alternative to the (G)ARCH class of models in accounting for time-varying persistence and volatility of financial returns. The SV model is motivated by the mixture-of-distributions hypothesis which was postulated by Clark (1973). Under this approach, asset returns follow a mixture of normal distributions with a mixing process depending on the unobservable flow of price-relevant information. Further additions include Tauchen and Pitts (1983) and Gallant, Hsieh, and Tauchen (1991) who noted that if the unobserved information flows are positively autocorrelated, then the resulting process with time-varying and autocorrelated conditional variance reveals volatility clustering - a typical feature of financial time series. The mixture of distributions approach naturally leads itself to the idea that asset volatility returns follow their own stochastic process with

unobservable innovations. This is in stark contrast to (G)ARCH models, where the conditional variance given the available information is a deterministic function of past observations and innovations. It is due to this flexibility and realistic view, that SV models demand their popularity (Carnero, Pena, and Ruiz, 2003; Ghysels and Perron, 1996).

The main properties of the SV model have also been recorded in Taylor (1994), Shephard (1996), Ghysels and Jasiak (1994), Capobianco (1996), and Barndorff-Nielsen and others (2001). The conditional variance of y_t in (1.2) is given by

$$\text{Var}[y_t|\boldsymbol{\theta}] = e^{\alpha+\frac{1}{2}\sigma_h^2},$$

where $\boldsymbol{\theta} = (\alpha, \beta, \sigma^2)$ and $\sigma_h^2 = \sigma^2/(1-\beta^2)$. The kurtosis of y_t is denoted as

$$\kappa_y = \kappa_\epsilon e^{\sigma_h^2},$$

where κ_ϵ is the kurtosis of ε_t.

In recent years, there has been further developments on the time-dependent return and volatility processes for the SV model. One prominent extension is the incorporation of long memory process in the return equation of (1.2).

1.2.2 Long memory

Long range dependence modelling, also known as long memory (LM), has become a fundamental aspect of time series modelling in a host of applications and plays a significant role in many fields such as hydrology, econometrics, DNA sequencing and traffic engineering amongst others. In a general sense, a stationary time series displays long memory if there is a divergence of the absolute sum of the autocorrelation function (ACF). Essentially, a stationary long memory time series displays a slowly decaying autocorrelation function towards zero. The most common class of such time series is the ARFIMA models which were popularized by Granger and Joyeux (1980). The general expression for ARFIMA processes may be defined by the equation

$$\phi(B)y_t = \psi(B)(1-B)^{-d}\varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \tag{1.4}$$

where the autoregressive (AR) and moving average (MA) polynomials are $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$, $\psi(B) = 1 + \psi_1 B + \ldots + \psi_q B^q$ respectively, B is the backshift operator and d is the long memory parameter. The term $(1 - B)^{-d}$ therefore can be considered a fractional differencing operator and is given by

$$(1 - B)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j + d)}{\Gamma(j + 1)\Gamma(d)} B^j.$$

If the polynomials $\phi(\cdot)$ and $\psi(\cdot)$ in (1.4) have no common zeros, and $d \in (-1, 1/2)$, then:

1. If the zeros of $\phi(\cdot)$ lie outside the unit circle $z : |z| = 1$ then there is a unique stationary solution of (1.4) given by $y_t = \sum_{j=-\infty}^{\infty} \varphi_j \varepsilon_{t-j}$ where φ_j are the coefficients of the following polynomial $\varphi(z) = (1 - z)^{-d}\psi(z)/\phi(z)$;

2. If the zeros of $\phi(\cdot)$ lie outside the closed unit disk $z : |z| \leq 1$, then the solution $\{y_t\}$ is causal; and

3. If the zeros of $\psi(\cdot)$ lie outside the closed unit disk $z : |z| \leq 1$, then the solution $\{y_t\}$ is invertible.

One typical treatment to deal with a non-stationary time series is to keep differencing until stationarity has been achieved. If the original time series is not differenced, then it has an infinite variance (strictly speaking) and is cumbersome to work with. Some statisticians argue that taking the difference may lead to data loss at lower spectral densities (power spectrums), which express the strength of variations (energy) as functions of frequencies instead of time. A mathematical definition is given by

$$f_s(\omega) = \sum_{t=-\infty}^{\infty} \gamma(t) e^{-2\pi i \omega t} \quad \text{where} \quad \gamma(t) = \int_{-1/2}^{1/2} e^{2\pi i \omega t} f_s(\omega) d\omega,$$

and $\gamma(t)$ is the autocovariance function. A solution to data loss at lower spectral densities is to take fractional differences.

A typical stationary time series has a spectral density function bounded at the frequency 0, and the ACF displays exponential decay. This may be the case for the standard autoregressive time series. Yet, this may not be the case for other time series and fitting these time series to models which assume that the spectral density is peaked at 0 may result in failure. The most common example is the time series with the standard long memory filter of Granger and Joyeux (1980), which assumes the

spectral density does not necessarily peak at 0, and is indeed unbounded. It is due to these unique properties that long memory as a concept is an exciting and practical topic that is of great interest to econometricians.

A further generalization of (1.4) is the Gegenbauer autoregressive moving average (GARMA) class which is defined as

$$\phi(B)(1-2uB+B^2)^d y_t = \psi(B)\varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0,\sigma^2), \tag{1.5}$$

where $|u| \leq 1, |d| \leq 1$ are real parameters. A few important properties about the GARMA model are that:

1. The power spectrum is given by

$$f_s(\omega) = C(\omega) \times [4(\cos\omega - u)^2]^{-d}, \quad -\pi < \omega < \pi, \tag{1.6}$$

   where $C(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}\left(\frac{\psi(e^{-i\omega})}{\phi(e^{-i\omega})}\right)^2$ and $i = \sqrt{-1}$; and

2. The process in (1.5) is deemed to be long memory when $(\{|u| < 1, 0 < d < 0.5\} \cup \{|u| = 1, 0 < d < 0.25\})$. These features are characterized by the hyperbolic decay of the autocorrelation function (ACF) and the unbounded spectrum at the Gegenbauer frequency, $\omega = \omega_g = \cos^{-1}(u)$.

1.2.3 Other stylized facts

The SV and LM models are powerful, yet do not fully explain all the dynamics of modern financial time series. There has been an explosion into the research of the unique stylized facts of financial time series over the last 10 years for two main reasons. Firstly, volatility spikes are more prevalent in financial markets post the GFC, and secondly, technology growth has had a stronger impact on the financial world (chiefly, AI and digital currencies) than before. It is due to these two reasons that interest has been growing in volatility measurement. This section explores a whole host of these and other stylized facts which are discussed at length throughout this thesis.

The unique and unconventional financial time series that exist today are different to what were considered the norm ten years ago. This is due to the increase in computational efficiency, and the interconnectedness of the modern world. By stark

contrast, financial news in the 1800's would travel between market participants via pigeons and as such, the price of financial assets took weeks or months to be fully reflective of that news. In the 1930s, the first teleprinter was used to distribute news to London newspapers, and in turn, the market took days to reflect incoming news. Nowadays, news is reflected almost instantly, and can experience multiple shocks in very short time intervals. Given this extreme change in how market news is received and disseminated, it is a natural assertion that the way these returns are measured must now become more vigilant. Hence, the traditional SV and LM effects must be combined with newer, and more resilient effects to fully measure these new stylized facts.

One stylized fact commonly observed in financial returns is a negative correlation between returns and volatilities (Asai, 2008). This phenomena is known as the leverage effect and was first discussed by Black (1976), who observed the volatility of stocks tend to increase when the price drops. The typical assertion made is that in response to bad financial news, the price of a stock decreases thereby increasing the debt-to-equity ratio of a firm. This in turn makes it a riskier investment and therefore increases future expected volatility. Hence, in empirical applications where volatility responds negatively to returns, SV models with leverage (SV-LVG) are utilized. The method of estimating this leverage effect has been dealt with in several ways.

One well established method to model the leverage effect in the SV model is via the negative correlation in a bivariate distribution between $(\varepsilon_t, \eta_{t+1})$ - the error terms for returns and latent future volatility respectively. This is a marginal approach and was investigated by Meyer and Yu (2000), Omori et al. (2007), and Choy and Chan (2000). Essentially they propose to factorize the bivariate distribution into a marginal distribution for volatility and a conditional distribution for returns. With t-innovations, the model is firstly expressed as a scale mixture of bivariate normals and then each bivariate normal distribution is factorized into a marginal and a conditional normal distribution. Alternatively, Wang, Chan, and Choy (2011) reversed the order to factorize the bivariate t into a marginal and a conditional t-distribution first and then expressed each t component as a scale mixture of normals. The advantage of this particular approach is a separate scale mixture of normals representation for each component and so it enables the distinction between outliers generated by the return or volatility processes. The disadvantage, however, is that it is algebraically

tedious to derive.

The second approach (a conditional approach) involves incorporating exogenous variables conditionally into the latent volatility equation. There is no universally agreed approach on how to perform this, however a few popular examples are discussed in Asai and McAleer (2005) and include:

1. $h_t = \alpha + \beta(h_{t-1} - \alpha) + \gamma|y_t| + \eta_t$; and
2. $h_t = \alpha + \beta(h_{t-1} - \alpha) + \gamma\{I(\varepsilon_t) - E[I(\varepsilon_t)]\} + \eta_t$,

where γ is a proxy for leverage and $I(\cdot)$ is an indicator function such that $I(x) = 1$ if $x < 0$ and $I(x) = 0$ otherwise.

The current SV and LM models mostly assume their error processes to be normally distributed. This is in line with the findings of Andersen et al. (2001), who demonstrated the assumption of Gaussian returns is indeed fair and adequate. Notwithstanding this, critics have argued the Gaussian assumption for error processes seems rather presumptuous and is highly non-reflective of the real world. Presumably, this assumption has been adopted mainly to simplify parameter inference (Taylor, 1986; Mahieu and Schotman, 1994; Kim, Shephard, and Chib, 1998). This conditional normality assumption is arguably too restrictive and suffers from non-robustness in the presence of outliers (Jacquier, Polson, and Rossi, 1994). A much better assumption is the use of heavy tailed distributions (Ruiz, 1994). The initial proponents in favor of including such an assumption into the SV model incorporated a scaled Student's t-distribution (Harvey, Ruiz, and Shephard, 1994). Their findings were later exonerated to confirm that heavy tails are more deeply connected and are a stylized fact of financial time series due to their leptokurtic distributions, and their slowly decaying autoregressive volatility behavior (Liesenfeld and Jung, 2000). Further notable extensions of SV models include:

1. The use of the generalised Student's t-distribution in a scale mixture of uniforms (SMU) (Wang, Choy, and Chan, 2013b);
2. The normal inverse Gaussian distribution as a scale mixture of normals (SMN) with an IG mixing distribution (Barndorff-Nielsen, 1997); and
3. The generalised hyperbolic skew t-distribution as a SMN with a generalised Inverse-Gamma (IG) mixing distribution (Nakajima and Omori, 2012). Note

that scale mixture distributions are discussed in further details Chapter 4.

It should also be noted these models focus mainly on distributional choices and leverage effects but do not consider other flexible modelling alternatives.

Another important modelling alternative is the inclusion of non-linear effects to uniquely measure the mean structure. Non-linear time series have gained traction within academic circles since the 1970s when non-linearity in many time series were observed and further investigated. One popular type of model to explain non-linear time series are threshold models. The first proposition of such a type of model was the Threshold Autoregregressive (TAR) model of Tong (1990). In comparison to their linear counterparts, a threshold model provides a much wider set of possible dynamics for financial and economic time series. It measures a set of regimes and switches between these regimes based on the levels of threshold variables. Such effects are important to measure the asymmetry of stock returns, which was studied in Li and Lam (1995). These non-linear stock returns occur during bull and bear markets and hence can be measured by threshold models. The results in Li and Lam (1995) reveal this conditional mean structure could depend significantly on the rise and fall of the market from previous time periods. Some alternative threshold extensions which combine the SV model include So, Lam, and Li (2002) who proposed a threshold SV in response to good or bad news and Chen, Liu, and So (2008b) who imposed a threshold model for both returns and the SV component.

The final effect to discuss are jump diffusions which capture discontinuous behavior in returns. In the classical literature, this jump behavior was introduced in order to measure some rare outlying event (Merton, 1976). However, with the growing interest of digital assets, these jump effects are now more relevant than before. The inclusion of jump effects are popular extensions to financial asset pricing models (Bates, 1996; Ball and Torous, 1985) and have also been applied in conjunction with the SV model (Barndorff-Nielsen and others, 2001; Chernov et al., 2003).

1.2.4 Bayesian MCMC method

Although intuitively appealing, the SV model was not often used due to its intractable likelihood which involves T dimensional integral with respect to the unknown volatility parameter $h_t,\ t = 1, \ldots, T$

$$p(y_t|\boldsymbol{\theta}) = \int p(y_t|h_t, \boldsymbol{\theta}) p(h_t|\boldsymbol{\theta}) dh_t, \tag{1.7}$$

where $\boldsymbol{\theta} = (\alpha, \beta, \sigma^2)$. Clearly, this observed data likelihood in its analytical form is near impossible to analytically evaluate. This has resulted in a number of techniques to estimate it. The earliest of such techniques used simulated maximum likelihood and was introduced by Geyer (1991). This method is less efficient, but is relatively easier to compute. Other approaches include quasi maximum likelihood (QMLE) (Bollerslev and Wooldridge, 1992; Harvey, Ruiz, and Shephard, 1994) and the efficient method of moments (Andersen, Chung, and Sørensen, 1999).

In a Bayesian setting, the evaluation of such a likelihood is a routine process. MCMC techniques were popularized into the SV literature by Kim, Shephard, and Chib (1998) and Jacquier, Polson, and Rossi (2004). In fact, Andersen, Chung, and Sørensen (1999) showed that MCMC is the most efficient method for estimating the SV model. Undoubtedly, MCMC techniques have operationalized the common use of the SV model. Other advantages of using the Bayesian method include the incorporation of prior information to supplement the data and the straightforward approach of using the posterior predictive distribution for inference. Moreover, Strasser (1975) showed the asymptotic equivalence of Bayes and maximum likelihood estimation.

The Bayes rule is the cornerstone of Bayesian analysis, and factorizes the posterior distribution into its constituents

$$p(\boldsymbol{\theta}, h_t|y_t) = \frac{p(y_t|\boldsymbol{\theta}, h_t) p(\boldsymbol{\theta}, h_t)}{p(y_t)} \propto p(y_t|\boldsymbol{\theta}, h_t) p(h_t|\boldsymbol{\theta}) p(\boldsymbol{\theta}),$$

where $p(y_t|\boldsymbol{\theta}, h_t)$ is the observed data likelihood, $p(h_t|\boldsymbol{\theta})$ is the conditional density and $p(\boldsymbol{\theta})$ is the joint prior density of the model parameters, also known as *the prior*. The predictive density $p(y_t)$ equals to

$$\int p(y_t|\boldsymbol{\theta}, h_t) p(\boldsymbol{\theta}, h_t) d\boldsymbol{\theta} dh_t,$$

and is independent of $\boldsymbol{\theta}$ or h_t. Hence it can be considered as an integrating constant. In essence, the Bayesian method of estimation is instigated with a joint conditional density, $p(\boldsymbol{\theta}, h_t|y_t)$, which is referred to as the *posterior distribution*. The posterior distribution is a summary of all the information that is known about the model and this information includes the information from the observed data as well as priors. The posterior distribution forms the basis for Bayesian inference, whereas the prior distribution allows the inclusion of prior beliefs on the parameters and provides any economic or practical interpretation of the parameters to be implemented. Throughout this thesis, we consider mostly non-informative or diffuse priors. For instance, the persistence parameter β of the latent volatility equation in (1.3) needs to be truncated between $[-1, 1]$ to ensure stationarity and is therefore assigned a uniform prior on $[-1, 1]$. Other parameters with real support include the mean parameters μ and α which are assigned a normal prior, whereas parameters with positive support such as the volatility of volatility σ^2 are assigned an IG prior in order to achieve mostly standard (conjugate) distributions.

Gibbs sampling

With the advent of cheap computing power in recent years, there has been an uptake in the use of MCMC methods to estimate high dimensional integrals. The Gibbs sampler is one of the most popular sampling methods to estimate a high dimensional parameter vector in these integrals, and is commonly used in Bayesian statistics (Tanner and Wong, 1987; Gelfand and Smith, 1990; Smith and Roberts, 1993).

The Gibbs sampler was first used in the image analysis literature by Geman and Geman (1984) and was later uprooted into the statistical literature by Besag and York (1989). It is an extremely powerful tool to sample from multidimensional posterior distributions. To see how it works, we consider the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{h}|\boldsymbol{y})$ of the SV model as a $T+3$-dimensional distribution where sampling directly from it is close to impossible. Specifically, the Gibbs sampler deals with the curse of dimensionality problem by approximating the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{h}|\boldsymbol{y})$ by blocks of conditional distribution, in which samples are more readily available.

The Gibbs sampler iterates through the complete set of conditional distributions as below

1. $p(\boldsymbol{h}|\boldsymbol{y}, \alpha, \beta, \sigma^2)$;
2. $p(\alpha|\boldsymbol{y}, \boldsymbol{h}, \beta, \sigma^2)$;
3. $p(\beta|\boldsymbol{y}, \boldsymbol{h}, \alpha, \sigma^2)$; and
4. $p(\sigma^2|\boldsymbol{y}, \boldsymbol{h}, \alpha, \beta)$.

These individual posterior distributions uniquely determine the complete joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{h}|\boldsymbol{y})$. Hence, sampling from these distributions is equivalent to sampling from the joint posterior distribution, up to some proportionality constant. In essence, the Gibbs sampler constructs the Markov Chain by iterating through the parameter space of the posterior distributions. Classical examples of estimating the SV model using the Gibbs sampler include Jacquier, Polson, and Rossi (1994), Kim, Shephard, and Chib (1998), and Chib and Greenberg (1994).

Metropolis-Hastings algorithm

Even though the dimensions for parameter sampling can be greatly reduced using the Gibbs sampler, the complexity of sampling still remains if the conditional distributions are not available. To deal with nonstandard distributions, a well suited and more generalised alternative to construct MCMC samplers is to use rejection sampling. Under this method, proposed values are drawn from a proposal/candidate generating distribution where it approximates the target (posterior conditional) density. These samples are corrected via an acceptance and rejection mechanism, so that asymptotically their behaviour resembles random observations from the target density. This is the mechanism for methods such as the Metropolis Hastings (MH) algorithm. A considerable amount of attention within the Bayesian sphere is devoted to the MH algorithm of Metropolis et al. (1953), and was later generalised by Hastings (1970). The Gibbs sampler is in fact a special case of the MH algorithm. The motivation for such a process is to draw candidate observations from a distribution, conditional upon the last observation and therefore invoking a Markov chain. The most definitive aspect of the MH algorithm is the approximating candidate density is improved at each step in the chain. This is in contrast to rejection sampling where

the candidate density remains the same. Assuming that the parameter of interest is θ, the MH algorithm can be summarized as follows:

Step 1: Set $m = 1$ while $m \leq M$.

Step 2: Generate θ from $q(\theta|\theta^{(m)})$ and u from a Uniform$(0, 1)$.

Step 3: Assign $\theta^{(m+1)} = \theta^*$ if $u \leq \mathbb{P}_a(\theta^*, \theta^{(m)})$, otherwise, let $\theta^{(m+1)} = \theta^{(m)}$, $m = m+1$,

where $q(\cdot|\cdot)$ is the candidate generating function, $p(\theta) = p(\theta|\boldsymbol{y}, \boldsymbol{\theta}^-)$ where $\boldsymbol{\theta}^- = \boldsymbol{\theta}\ \{\theta\}$ is the target density and $\mathbb{P}_a(\cdot, \cdot)$ is the acceptance probability. If the candidate θ^* is accepted, then the process will move to θ^*, otherwise it will stay at $\theta^{(m)}$. When the MCMC is currently at stage $\theta^{(m)}$, the value θ^* is generated from $q(\theta|\theta^{(m)})$, and is accepted for $\theta^{(m+1)}$ with the acceptance probability

$$\varrho(\theta^*, \theta^{(m)}) = \min\left(1, \frac{p(\theta^*)q(\theta^{(m)}|\theta^*)}{p(\theta^{(m)})q(\theta^*|\theta^{(m)})}\right),$$

where $p(\theta)$ is the target conditional posterior distribution. In Section 4.2.5, we utilize the method of Chib and Greenberg (1994) to construct the standard proposal density from a conjugate distribution. There are two special cases of the MH which are used throughout this thesis and deserve some attention.

Case 1: Independent Metropolis-Hastings algorithm

This algorithm allows a proposal distribution $q(\theta)$ to be independent of the current state $\theta^{(m)}$ of the chain (Tierney, 1994). This sampler is useful when the proposal value is independent of previous states and its efficiency depends on the proposal distribution $q(\theta)$ being close to $p(\theta)$. If this is not practically possible, Case two may be better. This case implies an acceptance probability of

$$\varrho(\theta^*, \theta^{(m)}) = \min\left(1, \frac{p(\theta^*)q(\theta^{(m)})}{p(\theta^{(m)})q(\theta^*)}\right).$$

Case 2: Random-walk Metropolis-Hastings algorithm

The candidate state is obtained by adding noise to the current state $\theta^* = \theta^{(m)} + e,\ e \sim N(0, \sigma^2)$. Specifically, the candidate density $q(\theta^*|\theta^{(m)}) = f(\theta^{(m)} - \theta^*)$, for some density $f(\cdot)$ which is symmetric about zero and is generated from a symmetric distribution centred at the current state. A common choice of $f(\cdot)$ is Gaussian with mean zero and variance σ^2. The symmetry property of the proposal transition,

$q(\theta^*|\theta^{(m)}) = q(\theta^{(m)}|\theta^*)$, leads to a simple form for the acceptance probability given by

$$\varrho(\theta^*, \theta^{(m)}) = \min\left(1, \frac{p(\theta^*)}{p(\theta^{(m)})}\right). \tag{1.8}$$

By using such an algorithm, the posterior sampler accepts new proposals according to the ratio of posterior distributions. This is commonly used when the data likelihood is difficult to work with. A relevant example is the sampling of the volatility persistence parameter β as given in Section 4.2.5.

Adaptive MCMC algorithms

The choice of an effective proposal distribution for the random walk Metropolis algorithm, for example, is essential in order to obtain reasonable results by simulation in a limited amount of time. A possible remedy is provided by adaptive algorithms, which use the history of the process in order to 'tune' the proposal distribution suitably. This tuning parameter is then modified by monitoring the acceptance rate. A high acceptance rate means that most proposed draws are being sampled around the current point, whereas a low acceptance rate means the chain is moving too slowly and not exploring the parameter space enough. Adaptive MCMC algorithms are designed with the intention of achieving an optimal acceptance rate, which is typically in the vicinity of 20 percent through to 30 percent (Roberts and Rosenthal, 2001); see Rue, Steinsland, and Erland (2004) for an overview of adaptive algorithms. By using an adaptive MCMC algorithm, the transition kernel is sequentially modified at a pre-specified number of steps throughout the simulation to obtain optimal efficiency (Roberts, Gelman, and Gilks, 1997; Haario, Saksman, and Tamminen, 2001).

In general, adaptive MCMC procedures can be summarized as follows:

1. Define a measurable function $q_m \times \theta^m \mapsto \theta$ such that $q_m(\theta|\theta^{(m)}, \gamma_m)$, $m = 1, 2, \ldots, M$ is a transition kernel with $\gamma_m = g_m(\theta^{(1)}, \ldots, \theta^{(m)}|\gamma_0, \theta^0)$ and function $g_m : \Theta^m \mapsto \mathbb{R}$.

2. Initialize the adaption chain with some arbitrary but fixed values $(\theta^0, \gamma_0) \in \Theta \times \Gamma$.

3. Sample θ^* from $q_m(\theta|\theta^{(m)}, \gamma_m)$ at iteration $m \geq 1$ given $(\gamma_0, \theta^0, \ldots, \theta^{(m)})$ and $\gamma_m = g_m(\theta^{(1)} \ldots, \theta^{(m)}|\theta^0, \gamma^0)$.

4. Return the value of $\theta^{(m+1)}$ according to the transition probability $\mathbb{P}_a(\theta^*, \theta^{(m)}, \gamma_m)$.

In practice, tuning is carried out with a multiplicative constant to the scale of the proposal distribution. If the acceptance rate deviates out of some predefined bounds, the constant is scaled accordingly to ensure an adequate acceptance rate will be achieved. For example, the transition kernel can be a Gaussian proposal centred on the current state with variance calculated using all of the previous states and recursive updating. In this case, $q_m(\theta|\theta^{(m)}, \gamma_m) = N(\theta^{(m)}, \gamma_m)$ where $\gamma_m = s_m^2(\theta^{(1)}, \ldots, \theta^{(m)}|\gamma_0, \theta^0)$ is the sample variance of $(\theta^{(m)}, \ldots, \theta^{(1)})$, $q_0 = N(\theta^{(0)}, \gamma_0)$ and $(\theta^{(0)}, \gamma_0)$ are some initial values. This transition kernel also determines the probability of moving to the next step. Alternatively, we may set $\gamma_m = c_m^2 V$ where V is a certain variance level and c_m^2 is a certain scale parameter of V to be tuned. A tuning example can be found in Appendix A.

Modal and distributional approximations

The previous sections describe Gibbs and MH algorithms which work well in low-dimensional problems. However, with more complicated models, it is sometimes difficult to even sample the posterior distribution directly. One approach to overcome this is the maximum a posteriori (MAP) estimation technique, which utilizes the mode of the posterior as a point estimate for the parameter of interest. The posterior mode is often under the guise of a penalized likelihood estimate, where the logarithm of the prior density is considered a penalty function.

The method first initiates by finding the mode of the posterior distribution. The mode is sought as a way to begin mapping the posterior density. If the posterior is multimodal, the global maximum should be found. In the event where multiple modes are found, then a mode-finding algorithm should be run from multiple starting points to ensure that a global maximum has been found. A wide variety of techniques exist for solving optimisation problems, and any of these, in principle, can be applied to find the mode of a posterior density. Examples of these techniques include stepwise ascent, Newton's method, quasi-Newton-gradient methods and the numerical computation of derivatives.

Once the mode is found, it can be used as a point estimate of the parameter of interest, and then used to sample the parameters accordingly. The maximum likelihood estimate for θ based on the data likelihood function is

$$\arg\max_{\theta} f(\boldsymbol{y}|\theta).$$

Under this framework, the posterior distribution is

$$f(\theta|y) = \frac{f(y|\theta)p(\theta)}{\int f(y|\vartheta)p(\vartheta)d\vartheta},$$

where $p(\theta)$ is the prior density of θ. The method of MAP estimation then estimates the mode of the posterior distribution as

$$\hat{\theta}_{\text{MAP}} = \arg\max_{\theta}\ p(\theta|\boldsymbol{y}) = \arg\max_{\theta} \frac{f(\boldsymbol{y}|\theta)\, p(\theta)}{\displaystyle\int_{\Theta} f(\boldsymbol{y}|\vartheta)\, p(\vartheta)\, d\vartheta} = \arg\max_{\theta}\ f(\boldsymbol{y}|\theta)\, p(\theta).$$

1.3 Thesis outline

This introductory section motivates our research by starting with some real financial problems that demand technological and statistical advancements. This is carried out by defining the fundamental concepts related to these problems, reviewing model development in the literature and providing technical details for the Bayesian MCMC sampling scheme which are applied in later chapters. These serve as building blocks towards further exploration. The remaining chapters are structured as below:

Chapter 2 explores the basic SV model, derives the normal conjugates and supplements the sampling techniques for the parameters in the volatility model that are applied in Chapter 3.

Chapter 3 is the first publication on Gegenbauer long memory SV models which extends the basic SV model. This chapter provides concepts of long memory and details of sampling schemes for long memory parameters. It further discusses some computational issues of importance and tests the accuracy of parameter estimates through extensive simulation studies. The tuning of the long memory parameters and the sampling of latent volatilities through mixtures of normals are given in Appendices A and B. Thereafter, several empirical applications, including forecasting the US equity

risk premium are performed and details of these posterior simulation and performance evaluations are also provided. The estimation of the marginal likelihood for performance evaluation is given in Appendix C.

Although this model has pleasing results, Chapter 4 further expands upon it in an iterative manner by incorporating popular financial stylized features such as leverage and heavy tailed distributions. Similar to Chapter 2, this chapter aims to provide background information for Chapters 5 and 6 which respectively report our second paper currently under review and our third published paper. Chapter 4 provides fundamental concepts for the factorization of a bivariate leverage effect model and for the scale mixture representation of the Student's t-distribution - both with an aim to facilitate sampling from posterior conditional distribution. Furthermore, this chapter also provides sampling of the volatility parameters, the leverage effect parameter, mixing variables and degrees of freedom, and simulation studies are conducted to test the performance of these parameters. Lastly, we provide a brief overview on Cryptocurrencies and detail how they are relevant to the future of statisticians. The findings from this chapter are then culminated in the form of a complete model and analyzed through the lens of Cryptocurrency data in Chapter 5.

Chapter 5 introduces this bivariate Student's t-long memory SV model with leverage effect and compares this extended model to two similar models. A list of the submodels are contained in Appendix E. The sampling for the long memory parameters and latent volatilities which adopts a new scheme is detailed in Appendix F. Simulation studies are again performed to assess the efficiency of estimators and some results are given in Appendix D. The effectiveness of this model is demonstrated via analyzing 114 Cryptocurrencies. Thereafter, special focus is drawn to the top five Cryptocurrencies by market capitalization and finally, a forecasting exercise is conducted using Bitcoin. The superiority of these forecasts is shown using several forecasting measures.

Chapter 6 considers this bivariate Student's t-leverage effect long memory SV model again, however the focus is on the practical issues of the application to Cryptocurrency. This Chapter also discusses in greater details the special properties of Cryptocurrencies such as their wild volatilities and their technological set-up in greater detail and relates these properties with the model-fitting findings.

Chapter 7 discusses alternative model structures that prove to be useful when modelling Cryptocurrencies, namely: realised volatility, buffered threshold and jumps. Again, this chapter supplements Chapter 8 by reviewing model developments of these three model structures and providing sampling schemes for their parameters.

Chapter 8 reports our fourth published paper which further explores the application of Cryptocurrencies in conjunction with the extended model developed in Chapter 7. The unique variance properties of Cryptocurrencies prompt a generalised model which is able to encapsulate their nature.

Finally, this thesis concludes with Chapter 9 which reviews all contributions and provides future research avenues.

Chapter 2

The stochastic volatility model

"All models are wrong, but some are useful."

George E.P. Box

This Chapter introduces the basic SV model and presents some Bayesian MCMC sampling techniques for estimating its parameters. These techniques are similarly applied and further extended to the Gegenbaur long memory SV model in Chapter 3.

2.1 Background

The SV model defined in (1.2) and (1.3) is central to this thesis. This model and various extensions in later chapters are listed in Appendix E. In the most typical Bayesian applications of the SV model, the routine software OpenBUGS/WinBUGS (Yu, 2005) or Rstan are used. As briefly mentioned in Section 1.2.4, the conditional posterior distributions discussed throughout this thesis are highly non-standard. As such, these off-the-shelf Bayesian programming tools may be inadequate or inefficient for the extensions we consider. Therefore, the relevant posterior of each parameter should be derived from first principles, and subsequently manually programmed. We choose the programing language MATLAB for all of our model implementations in this thesis.

2.1.1 Review of normal conjugates

Recall the standard result that if we are given the observations $\boldsymbol{X} = [x_1, \ldots, x_T]$ where $x_i \sim \mathrm{N}(\mu, \sigma^2)$ and $p(\mu) \sim \mathrm{N}(m, \tau^2)$, then the likelihood of $\boldsymbol{X}$ is

$$\frac{1}{(\sqrt{2\pi}\sigma)^T} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\right\},$$

and the posterior distribution of μ is

$$\begin{aligned}
p(\mu|\boldsymbol{X},\sigma^2) &\propto \exp\left\{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2} - \frac{(\mu-m)^2}{2\tau^2}\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\frac{(\sum_{i=1}^n x_i^2 - 2n\bar{x}\mu + n\mu^2)}{\sigma^2} + \frac{(\mu^2 - 2\mu m + m^2)}{\tau^2}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{-2n\bar{x}\mu + n\mu^2}{\sigma^2} + \frac{\mu^2 - 2\mu m}{\tau^2}\right)\right\} \quad \text{(dropping constant independent to } \mu\text{)} \\
&= \exp\left\{-\frac{1}{2}\left(\frac{\mu^2(n\tau^2+\sigma^2) - 2\mu(m\sigma^2 + n\bar{x}\tau^2)}{\sigma^2\tau^2}\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\frac{\mu^2 - 2\mu\frac{(m\sigma^2+n\bar{x}\tau^2)}{(n\tau^2+\sigma^2)}}{\frac{\sigma^2\tau^2}{(n\tau^2+\sigma^2)}}\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\frac{\left[\mu - 2\mu\frac{(m\sigma^2+n\bar{x}\tau^2)}{(n\tau^2+\sigma^2)}\right]^2 - \left(\frac{(m\sigma^2+n\bar{x}\tau^2)}{(n\tau^2+\sigma^2)}\right)^2}{\frac{\sigma^2\tau^2}{(n\tau^2+\sigma^2)}}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{\left[\mu - 2\mu\frac{(m\sigma^2+n\bar{x}\tau^2)}{(n\tau^2+\sigma^2)}\right]^2}{\frac{\sigma^2\tau^2}{(n\tau^2+\sigma^2)}}\right)\right\}.
\end{aligned}$$

Therefore

$$p(\mu|\boldsymbol{X},\sigma^2) \quad \sim \quad \mathrm{N}\left(\frac{m\sigma^2 + n\bar{x}\tau^2}{n\tau^2+\sigma^2}, \frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}\right).$$

Now, the mean can be written as

$$\begin{aligned}
\frac{m\sigma^2 + n\bar{x}\tau^2}{n\tau^2+\sigma^2}\frac{\div\sigma^2\tau^2}{\div\sigma^2\tau^2} &= \frac{\frac{m}{\tau^2} + \frac{n\bar{x}}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \\
&= \frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}m + \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\bar{x},
\end{aligned}$$

and the variance as

$$\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}\frac{\div\sigma^2\tau^2}{\div\sigma^2\tau^2} = \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}},$$

so that

$$\begin{aligned} p(\mu|\boldsymbol{X},\sigma^2) &\sim \mathrm{N}\left(\frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}m+\frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\bar{x},\frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\right) \\ &\sim \mathrm{N}\left(V_\mu\left(\frac{m}{\tau^2}+\frac{n\bar{x}}{\sigma^2}\right),V_\mu\right), \quad \text{where } V_\mu=\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}. \end{aligned} \tag{2.1}$$

2.2 Bayesian inference for the SV model

2.2.1 Sampling the volatility level parameter α in the SV model

We denote the vector of volatilities $h_t, t=1,\ldots,T$ as $\boldsymbol{h}$, the vector of $h_t, t=2,\ldots,T$ as $\boldsymbol{h}_{-1}$ and the vector of $h_t, t=1,\ldots,T-1$ as $\boldsymbol{h}_{-T}$. Let the prior distribution of α be $\mathrm{N}(\mu_\alpha,\sigma_\alpha^2)$. The posterior distribution of α is expressed as

$$
\begin{aligned}
& p(\alpha|\boldsymbol{h},\beta,\sigma^2) \propto f(h_1|\alpha,\beta,\sigma^2) f(\boldsymbol{h}_{-1}|\boldsymbol{h}_{-T},\alpha,\beta,\sigma^2) \times p(\alpha) \\
= \;& \frac{(1-\beta^2)^{-\frac{1}{2}}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2}\right\} \times \\
& \frac{1}{(\sqrt{2\pi}\sigma)^{T-1}} \exp\left\{-\frac{\sum_{t=2}^{T}[h_t-\alpha-\beta(h_{t-1}-\alpha)]^2}{2\sigma^2}\right\} \times \frac{1}{\sqrt{2\pi}\sigma_\alpha} \exp\left\{-\frac{(\alpha-\mu_\alpha)^2}{2\sigma_\alpha^2}\right\} \\
\propto \;& \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}\left((h_1-\alpha)^2(1-\beta)^2 + \sum_{t=2}^{T}[h_t-\alpha-\beta(h_{t-1}-\alpha)]^2\right) + \frac{(\alpha-\mu_\alpha)^2}{\sigma_\alpha^2}\right]\right\} \\
= \;& \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}\left((h_1^2-2h_1\alpha+\alpha^2)(1-\beta^2) + \sum_{t=2}^{T}\{h_t^2 - 2h_t[\alpha(1-\beta)+\beta h_{t-1}]\right.\right.\right. \\
& \left.\left.\left. + \; [\alpha(1-\beta)+\beta h_{t-1}]^2\}\right) + \frac{\alpha^2-2\alpha\mu_\alpha+\mu_\alpha^2}{\sigma_\alpha^2}\right]\right\} \\
\propto \;& \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}\left((-2h_1\alpha+\alpha^2)(1-\beta^2) + \sum_{t=2}^{T}[-2h_t\alpha(1-\beta)+\alpha^2(1-\beta)^2+2\alpha(1-\beta)\beta h_{t-1}]\right)\right.\right. \\
& \left.\left. + \frac{\alpha^2-2\alpha\mu_\alpha}{\sigma_\alpha^2}\right]\right\} \\
= \;& \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}\left(-2h_1\alpha(1-\beta^2)+\alpha^2(1-\beta^2) - 2\alpha(1-\beta)\sum_{t=2}^{T}h_t + \alpha^2(1-\beta)^2(T-1)\right.\right.\right. \\
& \left.\left.\left. +2\alpha(1-\beta)\beta\sum_{t=2}^{T}h_{t-1}\right) + \frac{\alpha^2}{\sigma_\alpha^2} - \frac{2\alpha\mu_\alpha}{\sigma_\alpha^2}\right]\right\} \\
= \;& \exp\left\{-\frac{1}{2}\left[\alpha^2\left(\frac{(1-\beta^2)+(1-\beta)^2(T-1)}{\sigma^2} + \frac{1}{\sigma_\alpha^2}\right)\right.\right. \\
& \left.\left. -2\alpha\left(\frac{h_1(1-\beta^2)+(1-\beta)\sum_{t=2}^{T}h_t - (1-\beta)\beta\sum_{t=2}^{T}h_{t-1}}{\sigma^2} + \frac{\mu_\alpha}{\sigma_\alpha^2}\right)\right]\right\} \\
= \;& \exp\left\{\frac{\alpha^2 - 2\alpha\left(\frac{\frac{h_1(1-\beta^2)+(1-\beta)\sum_{t=2}^{T}h_t-(1-\beta)\beta\sum_{t=2}^{T}h_{t-1}}{\sigma^2}+\frac{\mu_\alpha}{\sigma_\alpha^2}}{\frac{(1-\beta^2)+(1-\beta)^2(T-1)}{\sigma^2}+\frac{1}{\sigma_\alpha^2}}\right)}{-2\left(\frac{(1-\beta^2)+(1-\beta)^2(T-1)}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)^{-1}}\right\} \\
\propto \;& \exp\left\{\frac{\left[\alpha - \left(\frac{\frac{h_1(1-\beta^2)+(1-\beta)\sum_{t=2}^{T}h_t-(1-\beta)\beta\sum_{t=2}^{T}h_{t-1}}{\sigma^2}+\frac{\mu_\alpha}{\sigma_\alpha^2}}{\frac{(1-\beta^2)+(1-\beta)^2(T-1)}{\sigma^2}+\frac{1}{\sigma_\alpha^2}}\right)\right]^2}{-2\left(\frac{(1-\beta^2)+(1-\beta)^2(T-1)}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)^{-1}}\right\} \\
= \;& \exp\left\{\frac{\left[\alpha - \left(\frac{\frac{h_1(1-\beta^2)+(1-\beta)\sum_{t=2}^{T}[h_t-\beta h_{t-1}]}{\sigma^2}+\frac{\mu_\alpha}{\sigma_\alpha^2}}{\frac{(1-\beta^2)+(1-\beta)^2(T-1)}{\sigma^2}+\frac{1}{\sigma_\alpha^2}}\right)\right]^2}{-2\left(\frac{(1-\beta^2)+(1-\beta)^2(T-1)}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)^{-1}}\right\}
\end{aligned}
$$

Therefore,

$$p(\alpha|\boldsymbol{h}, \beta, \sigma^2) \sim \mathrm{N}\left(V_\alpha M_\alpha, V_\alpha\right) \tag{2.2}$$

where
$M_\alpha = \dfrac{(1-\beta^2)h_1 + (1-\beta)\sum_{t=2}^{T}[h_t - \beta h_{t-1}]}{\sigma^2} + \dfrac{\mu_\alpha}{\sigma_\alpha^2}$, and
$V_\alpha = \left(\dfrac{(1-\beta^2) + (1-\beta)^2(T-1)}{\sigma^2} + \dfrac{1}{\sigma_\alpha^2}\right)^{-1}$.

2.2.2 Sampling the volatility persistence parameter β in the SV model

A similar principle can be applied to estimate β, whereby the posterior distribution of β is

$$\begin{aligned} & p(\beta|\boldsymbol{h}, \alpha, \sigma^2) \propto f(h_1|\alpha, \beta, \sigma^2) f(\boldsymbol{h}_{-1}|\boldsymbol{h}_{-T}, \alpha, \beta, \sigma^2) \times p(\beta) \\ = & \frac{\sqrt{1-\beta^2}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2}\right\} \times \frac{1}{(\sqrt{2\pi}\sigma)^{T-1}} \exp\left\{-\sum_{t=1}^{T-1}\frac{(h_{t+1}-\alpha-\beta(h_t-\alpha))^2}{2\sigma^2}\right\} \\ & \times \frac{1}{\sqrt{2\pi}\sigma_\beta} \exp\left\{-\frac{(\beta-\mu_\beta)^2}{2\sigma_\beta^2}\right\}. \end{aligned}$$

Unlike parameter α, the posterior distribution of β cannot be represented in a Gaussian form due to the existence of the prior $p(\beta)$ and the marginal distribution $f(h_1|\alpha, \beta, \sigma^2)$ and is therefore non-standard. Hence, another approach is needed to estimate β with a non-standard posterior distribution.

The work of Chib and Greenberg (1994) was directed to estimate variants of the ARMA(p, q) model. They also face the problem of non-standard posteriors. The essence of their idea is to implement the MH algorithm to sample β from the proposal (or candidate) density which equals to the conditional likelihood, $\prod_{t=2}^{T} f(h_t|h_{t-1}, \alpha, \beta, \sigma^2)$, and set the target density to be the marginal likelihood $f(h_1|\alpha, \beta, \sigma^2)$.

In order to implement this scheme, we first consider the conditional likelihood of $\boldsymbol{h}_{-1}$ as

$$
\begin{aligned}
f(\boldsymbol{h}_{-1}|\boldsymbol{h}_{-T},\alpha,\beta,\sigma^2) &= \frac{1}{(\sqrt{2\pi}\sigma)^{T-1}}\exp\left\{-\frac{\sum_{t=2}^{T}[h_t-\alpha-\beta(h_{t-1}-\alpha)]^2}{2\sigma^2}\right\} \\
\propto &\quad \exp\left\{-\frac{\sum_{t=2}^{T}[h_t-[\alpha(1-\beta)+\beta h_{t-1}]]^2}{2\sigma^2}\right\} \\
\propto &\quad \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{t=2}^{T}\{h_t^2-2h_t[\alpha(1-\beta)+\beta h_{t-1}]+[\alpha(1-\beta)+\beta h_{t-1}]^2\}\right]\right\} \\
\propto &\quad \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{t=2}^{T}\{-2h_t(-\alpha\beta+\beta h_{t-1})+[\alpha^2(1-\beta)^2+2\alpha\beta h_{t-1}(1-\beta)+\beta^2h_{t-1}^2]\}\right]\right\} \\
= &\quad \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{t=2}^{T}\{2h_t(\alpha-h_{t-1})\beta+[\alpha^2(1-2\beta+\beta^2)+2\alpha h_{t-1}\beta-2\alpha h_{t-1}\beta^2+h_{t-1}^2\beta^2]\}\right]\right\} \\
\propto &\quad \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{t=2}^{T}\{(\alpha^2-2\alpha h_{t-1}+h_{t-1}^2)\beta^2+[2h_t(\alpha-h_{t-1})-2\alpha^2+2\alpha h_{t-1}]\beta\}\right]\right\}.
\end{aligned}
$$

We note that $(\alpha^2-2\alpha h_{t-1}+h_{t-1}^2)$ is equal to $(h_{t-1}-\alpha)^2$ and

$$
\begin{aligned}
[2h_t(\alpha-h_{t-1})-2\alpha^2+2\alpha h_{t-1}] &= 2[h_t\alpha-h_{t-1}h_t-\alpha^2+\alpha h_{t-1}] \\
&= -2[(h_t-\alpha)(h_{t-1}-\alpha)].
\end{aligned}
$$

Therefore we have,

$$
\begin{aligned}
f(\boldsymbol{h}_{-1}|\boldsymbol{h}_{-T},\alpha,\beta,\sigma^2) &\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{t=2}^{T}(h_{t-1}-\alpha)^2\beta^2-2\sum_{t=2}^{T}[(h_t-\alpha)(h_{t-1}-\alpha)]\beta\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\frac{\beta^2-2\beta\frac{\sum_{t=2}^{T}[(h_t-\alpha)(h_{t-1}-\alpha)]}{\sum_{t=2}^{T}(h_{t-1}-\alpha)^2}}{\sigma^2(\sum_{t=2}^{T}(h_{t-1}-\alpha)^2)^{-1}}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\frac{\left(\beta-\frac{\sum_{t=2}^{T}[(h_t-\alpha)(h_{t-1}-\alpha)]}{\sum_{t=2}^{T}(h_{t-1}-\alpha)^2}\right)^2}{\sigma^2(\sum_{t=2}^{T}(h_{t-1}-\alpha)^2)^{-1}}\right\}.
\end{aligned}
$$

Hence neglecting the prior $p(\beta)$ and $f(h_1|\alpha,\beta,\sigma^2)$, we have,

$$
\beta|\boldsymbol{h},\alpha,\sigma^2 \sim N(V_\beta M_\beta, V_\beta), \tag{2.3}
$$

where

$$M_\beta = \frac{1}{\sigma^2}\sum_{t=1}^{T-1}[(h_{t+1}-\alpha)(h_t-\alpha)] \quad \text{and} \quad V_\beta = \left(\frac{1}{\sigma^2}\sum_{t=1}^{T-1}(h_t-\alpha)^2\right)^{-1}.$$

In order to implement the MH step, we first sample β^* from (2.3). Given the current value $\beta^{(m-1)}$ at the $(m-1)^{th}$ MCMC loop, we accept β^* with probability $\min\left\{\frac{q(\beta^*)}{q(\beta^{(m-1)})}, 1\right\}$, where $q(\beta)$ is

$$q(\beta) = p(\beta)\frac{\sqrt{1-\beta^2}}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2}\right\}.$$

We work with $\log q(x)$ so that

$$\log q(\beta) = \log p(\beta) + \frac{1}{2}\log(1-\beta^2) - \frac{1}{2}\log(2\pi\sigma^2) - \frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2},$$

and β^* is accepted with probability $\min\left\{\exp[q(\beta^*)-q(\beta^{(m-1)})], 1\right\}$.

2.2.3 Sampling the volatility of volatility parameter σ^2 in the SV model

Assuming a prior of $p(\sigma^2) \sim \text{IG}(\frac{a}{2}, \frac{b}{2})$, we have

$$\begin{aligned}
& p(\sigma^2|\boldsymbol{h},\alpha,\beta) = f(h_1|\alpha,\beta,\sigma^2)f(\boldsymbol{h}_{-1}|\boldsymbol{h}_{-T},\alpha,\beta,\sigma^2)\times p(\sigma^2) \\
= \;& \frac{\sqrt{1-\beta^2}}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2}\right\}\times \\
& \frac{1}{(\sqrt{2\pi}\sigma)^{T-1}}\exp\left\{-\frac{\sum_{t=2}^{T}[h_t-\alpha-\beta(h_{t-1}-\alpha)]^2}{2\sigma^2}\right\}\times\frac{(\frac{b}{2})^{\frac{a}{2}}}{\Gamma(\frac{a}{2})}\sigma^{-2(\frac{a}{2}+1)}\exp\left(-\frac{(\frac{b}{2})}{\sigma^2}\right) \\
\propto \;& \sigma^{-2(\frac{T}{2}+\frac{a}{2}+1)}\exp\left\{-\frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2}-\frac{\sum_{t=2}^{T}[h_t-\alpha-\beta(h_{t-1}-\alpha)]^2}{2\sigma^2}-\frac{b}{2\sigma^2}\right\}.
\end{aligned}$$

Hence, we have

$$\sigma^2|\boldsymbol{h},\alpha,\beta \sim \text{IG}\left(\frac{T+a}{2}, \frac{b+(h_1-\alpha)^2(1-\beta^2)+\sum_{t=2}^{T}[h_t-\alpha-\beta(h_{t-1}-\alpha)]^2}{2}\right). \quad (2.4)$$

2.2.4 Sampling the latent volatility vector h of the SV model

Given the normal conjugate result, we now apply it to the SV model in (1.2) and (1.3). It is assumed that h_t is stationary, so the mean and variance of the unconditional marginal distribution of h_1 is found by

$$\begin{aligned}\mathbb{E}[h_1] &= \alpha + \beta(\mathbb{E}[h_1] - \alpha) \\ &= \alpha,\end{aligned}$$

and

$$\begin{aligned}\mathbb{V}\text{ar}[h_1] &= \beta^2 \mathbb{V}\text{ar}[h_1] + \sigma^2 \\ &= \frac{\sigma^2}{1-\beta^2}.\end{aligned}$$

Hence,

$$h_1|\alpha, \beta, \sigma^2 \sim N\left(\alpha, \frac{\sigma^2}{1-\beta^2}\right).$$

Now, the full conditional distribution of h_t is given as

$$h_t|h_{t-1}, \alpha, \beta, \sigma^2 \sim \begin{cases} \text{N}\left(\alpha, \frac{\sigma^2}{1-\beta^2}\right), & t = 1, \\ \text{N}(\alpha + \beta(h_{t-1} - \alpha), \sigma^2), & t > 1. \end{cases}$$

The complete estimation method of h_t is described in detail in Appendix B.

2.3 Conclusion

After considering the basic SV model, we discuss in the upcoming Chapter the Gegenbauer long memory SV model. This model motivates the model progression of this thesis which aims to combine the Gegenbauer long memory time series model within the SV modelling framework to overcome the shortcomings of each one alone. This model and its applications are reported in the next Chapter as our first publication in *Studies in Nonlinear Dynamics and Econometrics*.

Chapter 3

Gegenbauer long memory processes with stochastic volatility

"If it looks like a duck, swims like a duck, and quacks like a duck, then it probably is a duck..."

Unknown

This paper discusses a time series model which has generalised long memory in the mean process with stochastic volatility errors and develops a new Bayesian posterior simulator that couples advanced posterior maximisation techniques, as well as traditional latent stochastic volatility estimation procedures. Details are provided on the estimation process, data simulation, and out of sample performance measures. We conduct several rigorous simulation studies and verify our results for in and out of sample behaviour. We further compare the goodness of fit of the generalised process to the standard long memory model by considering two empirical studies on the US Consumer Price Index (CPI) and the US Equity Risk Premium (ERP).

3.1 Introduction

Applications in econometrics, hydrology and other scientific disciplines have motivated time-series developments in fractionally differenced, or long range models over the past two decades. The seminal work of Granger and Joyeux (1980) and Hosking (1981) introduced the autoregressive fractionally integrated moving average (ARFIMA) model. A stationary time series y_t is said to be long memory if $\sum_{k=0}^{\infty}|\delta(k)|$

diverges, where $\delta(k)$ is the k^{th}-lag autocovariance. This class of time series generalizes the usual Box-Jenkins ARIMA model by modelling long term correlation structures as suggested by Mandelbrot and Ness (1968).

The prominence of the ARFIMA model can be seen through various extensions such as the long memory in stochastic volatility process (Baillie, 1996), ARFIMA model with ARCH errors (Hauser and Kunst, 1998) and the fractionally integrated GARCH model with leverage (Baillie, Bollerslev, and Mikkelsen, 1996). Although theoretically pleasing, the implementation of the ARFIMA model was a major deterrent. Chan and Palma (1998) operationalized the ARFIMA model by considering a state space representation and an approximate maximum likelihood estimator by means of the Kalman filter. Since then, the ARFIMA model has been applied extensively to a myriad of contexts with various extensions. Goldman et al. (2013) estimate a threshold fractionally integrated model with efficient jumps to better model intraday Exchange Traded Funds data. Iglesias, Jorquera, and Palma (2006) proposed a new methodology to better handle residuals which exhibit long-memory. Reisen, Rodrigues, and Palma (2006) discuss the estimation of seasonal fractional long memory models. More recent advances include the development of a heterogeneous infinite order autoregressive long memory estimate by ordinary least squares (Hwang and Shin, 2014). Conrad and Karanasos (2005) develop a dual long memory process by first estimating the conditional variance from the ARFIMA-FIGARCH model and then use Grander methods to test for bidirectional effects. Carlos and Gil-Alana (2016) use Chebyshev polynomials to examine the interaction between non-linear deterministic trends and long memory with one or more non-zero power spectrum frequencies.

Most notably, one recent suggestion is the ARFIMA stochastic volatility (ARFIMA-SV) model of Bos, Koopman, and Ooms (2014a), which models long memory in the time series itself, and measures the variance as a latent stochastic volatility (SV) process of Taylor (1986). The authors find good results measuring the Consumer Price Index of the United States (US CPI). We see this as an exciting path as the amalgamation of long memory and SV are representative of two important stylized facts of financial time series. The process is able to capture long memory effects, and the variance is able to develop more freely over time compared to the usual white noise case. The seminal work of Jacquier, Polson, and Rossi (1994) sheds light on a Bayesian approach on the estimation of SV models. Most notably, the estimation

of the latent variable h_t was later refined by Kim, Shephard, and Chib (1998) as a multi-move sampler which is briefly discussed later. Notable extensions of SV models include the Threshold Stochastic Volatility Model (So, Lam, and Li, 2002), SV models with fat-tails and correlated errors (Jacquier, Polson, and Rossi, 2004), SV models with Markov Switching (So, Lam, and Li, 1998) and the generalization of the return distribution using the generalised-t distribution (Wang, Choy, and Chan, 2013a).

An appealing generalization of traditional long memory models are generalised autoregressive fractional integrated moving average models (GARFIMA); whereby Gegenbauer polynomials replace the plain long memory fractional differencing operator. Gegenbauer polynomials were first introduced to the time-series community by Gray, Zhang, and Woodward (1989a). The novelty in such polynomials lie in their orthogonality and recursion properties. Bordignon, Caporin, and Lisi (2007) considered Gegenbauer fractionally integrated GARCH (FIGARCH) processes to measure intraday volatility. Lopes and Prass (2013) further extended this by including seasonality.

Evidently, we see a worthwhile pursuit in the coupling of the GARFIMA model as well as the SV model: the GARFIMA-SV model. We note the GARFIMA-SV (and therefore the ARFIMA-SV) is a special case of the so called doubly fractional model of Artiach and Arteche (2012). The authors use a sequential estimation strategy based on the Whittle appoximation to maximum likelihood in order to estimate the model in sample only, by first estimating a GARFIMA mean model, then using the residuals to estimate a GARFIMA-SV model. We purposely note here that our first contribution is to detail a new Bayesian estimation procedure, and discuss in detail forecasting techniques. The Bayesian approach includes many added benefits. Instinctively, Bayesian schemes are advantageous over frequentist approaches as complex hierarchical model structures can be specified and estimated with MCMC. Further, the inferences made are conditional on the observed data without relying on asymptotic approximations and the output provides credibility intervals which are easy to interpret. Our estimation method is also straight forward such that alternative mean structures (e.g., AR(p), exogenous variables, trends, jump points, outliers to name a few) can be easily implemented via the design matrix . We rely on the exploitation of matrix structures and efficient Kalman filtering techniques as employed by Chan (2013) to greatly reduce the computational burden.

Pursuing the GARFIMA-SV model has several motivations stemming from applied and

theoretical reasons. Bhardwaj and Swanson (2006) showed that long memory models can provide better out-of-sample results than ARMA and GARCH models for the prediction of macroeconomic and financial time series. Gray, Zhang, and Woodward (1989a) also notes that some time series processes do not necessarily display slowly dampening autocorrelations, yet are still valid candidates for generalised fractional differencing. In practice, the time series analyst can be conflicted with persistent residual autocorrelations at high lags which are not accounted for with parsimonious model choices. A potential candidate model would therefore be a long-memory filter, or more generally, a Gegenbauer filter with time dependent stochastic residuals. Long memory models in the past have been criticized for mistaking trend as long memory. Sowell (1992) outlined a hypothesis for testing trend stationarity versus difference stationarity. Crato and Rothman (1994) further supported these claims and found strong evidence of difference stationarity in popular macroeconomic time series. Supplementary evidence has found the existence of long memory in exchange rates (Cheung, 1993; Gil-Alana and Toro, 2002; Fei-xue, Yan, and Tie-shan, 2009); unemployment (Mikhail, Eberwein, and Handa, 2006; Lahiani and Scaillet, 2009; Gil-Alana, 2002); and equity returns
(Lillo and Farmer, 2004; Aye et al., 2014; Turkyilmaz and Balibey, 2014).

Our second contribution is the detailing and implementation of a Bayesian forecasting scheme and directly applying our findings to the US equity risk premium (US ERP), which is found to be non-stationary under the ordinary ARFIMA-SV specification. A prominent point of interest for practitioners globally has been extremely high bouts of volatility, looming deflationary talks and sub-par equity returns. The ERP intuitively delivers an extremely strong case to exhibit long memory as well as time dependent residuals as it moves with economic cycles, and exhibits long-term autocorrelations. Indeed, there are deeply rooted practical reasons to better understanding the ERP also, as macroeconomic based asset managers view the ERP as a gauge of investor sentiment. We are currently not aware of any papers which consider the long memory properties of the ERP. Considering the ERP is one way to better understand the relationship between stocks and bonds, and therefore there is merit in understanding its structural process.

The remainder of the article is organised as follows. The GARFIMA-SV model is introduced in Section 3.2, and we describe its relationship to the ARFIMA-SV model.

In Section 3.3, we outline the complete MCMC sampling scheme and some important computational issues. Section 3.4.2 describes our in and out of sample simulation studies and some further computational issues. Our empirical applications to the US CPI and US ERP are detailed in Section 3.5, and we finally conclude with Section 3.6.

3.2 The Gegenbauer long memory in mean with stochastic volatility model

Let $y_t, t = 1, 2, ..., T$ be a stochastic process satisfying the equations

$$\phi(B)(1 - 2uB + B^2)^d(y_t - \mu) = \psi(B)\varepsilon_t, \quad \varepsilon_t|F_{t-1} \sim \text{N}(0, e^{h_t}) \tag{3.1}$$

$$h_t = \alpha + \beta(h_{t-1} - \alpha) + \eta_t, \qquad \eta_t \sim \text{N}(0, \sigma^2) \tag{3.2}$$

where the autoregressive (AR) and moving average (MA) polynomials are $\phi(B) = 1-\phi_1 B-\ldots-\phi_p B^p$, $\psi(B) = 1+\psi_1 B+\ldots+\psi_q B^q$ respectively and B is the backshift operator.

We assume that y_t is stationary and invertible such that the zeros of $\phi(z)$ and $\psi(z)$ lie outside the unit circle with no common zeros, μ is a constant and F_{t-1} is the natural filtration of $\{y_t\}_{t\geq 0}$. It is known that y_t is causal when $(\{|u|< 1, d < 0.5\} \cup \{|u|= 1, d < 0.25\})$, and invertible when $(\{|u|< 1, d > -0.5\} \cup \{|u|= 1, d > -0.25\})$. The class of time series generated by (3.1) and (3.2) is called a GARFIMA(p, q)-SV time series process and long memory when $(\{|u|< 1, 0 < d < 0.5\}\cup\{|u|= 1, 0 < d < 0.25\})$.

Clearly, h_t is the log-volatility, which evolves according to the state equation (3.2) for $t = 1, \ldots, T$, α is the constant level of the volatility, β is the persistence of the volatility process and σ^2 is the volatility of volatility such that $E[\varepsilon_t\eta_t] = 0 \ \forall \ t$ and $E[\varepsilon_t\eta_s] = 0 \ \forall \ t, s$. We assume $|\beta|< 1$ so h_t is stationary and initialized with $h_0 \ \sim N(\alpha, \sigma^2/(1 - \beta^2))$.

For simplicity, we discuss the generalised fractional stochastic volatility noise process when $\phi(B) = \psi(B) = 1$ such that $(1-2uB+B^2)^d(y_t-\mu_t) = \varepsilon_t$. Under the assumption that y_t is causal, we have the following MA(∞) representation

$$y_t - \mu = (1 - 2uB + B^2)^{-d}\varepsilon_t = \sum_{j=0}^{\infty} \lambda_j \varepsilon_{t-j}, \tag{3.3}$$

where λ_j are the Gegenbauer coefficients, initialized with $\lambda_0 = 1, \lambda_1 = 2ud$ and follow the recursion

$$\lambda_j = 2u\left(\frac{d-1}{j}+1\right)\lambda_{j-1} - \left(\frac{2(d-1)}{j}+1\right)\lambda_{j-2}, \quad j \geq 2. \tag{3.4}$$

Further details on the Gegenbauer polynomial and its properties can be found in Rainville (1960). A truncated moving average representation of the Wold representation in (3.3) arises from truncating at lag J so that

$$y_t - \mu = (1 - 2uB + B^2)^{-d}\varepsilon_t \approx \sum_{j=0}^{J} \lambda_j \varepsilon_{t-j}. \tag{3.5}$$

The choice of J is discussed further in Section 3.2. The power spectrum of (3.3), conditional of h_t, is given by

$$f_{y_t|h_t}(\omega) = C[4(\cos\ \omega - u)^2]^{-d} \quad -\pi < \omega < \pi,$$

where C is a suitable constant, and the Gegenbauer frequency is $\omega = \cos^{-1}(u)$.
It is duly important to note the special case when $u = 1$, (3.3) collapses to the ARFIMA-SV model of Bos, Koopman, and Ooms (2014a). In this special case, $(1 - B)^{2d}(y_t - \mu) = \varepsilon_t$, where $(1 - B)^{2d}$ is said to be an integrating filter of order $2d$ and typically defined in terms of its Taylor series expansion.

3.3 Estimation

3.3.1 Sampling scheme of the (G)ARFIMA-SV

This section explains an efficient sampling scheme of the long-memory GARFIMA(1,0)-SV model. It should be noted that other mean structures can also be easily implemented. The observation equation in (3.1) now becomes

$$y_t = \mu + \phi y_{t-1} + \sum_{j=0}^{J_t^*} \lambda_j \varepsilon_{t-j}, \tag{3.6}$$

where $E[y_t] = \mu/(1-\phi)$, $|\phi| < 1$ and $J_t^* = \min(t, J)$. For a given set of observations $(y_1, \ldots, y_T)$, consider the following matrix representation

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\Xi} + \boldsymbol{G}_J\boldsymbol{\varepsilon}, \tag{3.7}$$

where

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & y_0 \\ 1 & y_1 \\ \vdots & \\ 1 & y_{T-1} \end{bmatrix};$$

$\boldsymbol{\Xi}' = (\mu, \phi)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_T)'$ is a $T \times 1$ vector of stochastic innovations which have the joint multivariate Gaussian distribution $\mathbf{N}(\mathbf{0}, \boldsymbol{V})$ with $\boldsymbol{V} = \text{diag}(e^{h_1}, \ldots, e^{h_T})$. $\boldsymbol{G}_J$ is a $T \times T$ lower triangular banded matrix with J Gegenbauer truncated moving average parameters in each column, and ones on the diagonal as given below

$$\boldsymbol{G}_J = \begin{bmatrix} 1 & 0 & \ldots & \ldots & \ldots & \ldots & \ddots & \ldots & 0 \\ \lambda_1 & 1 & \ldots & \ldots & \ldots & \ldots & \ddots & \ldots & 0 \\ \lambda_2 & \lambda_1 & \ldots & \ldots & \ldots & \ldots & \ddots & \ldots & 0 \\ \vdots & \lambda_2 & \ldots & \ldots & \ldots & \ldots & \ddots & \ldots & \vdots \\ \lambda_J & \vdots & \ldots & \ldots & \ldots & \ldots & \ddots & \ldots & \vdots \\ 0 & \lambda_J & \ldots & \ldots & \ldots & \ldots & 0 & 0 & 0 \\ \vdots & 0 & \ldots & \ldots & \ldots & \ldots & 1 & 0 & 0 \\ \vdots & \vdots & \ldots & \ldots & \ldots & \ldots & \lambda_1 & 1 & 0 \\ 0 & 0 & \ldots & \ldots & \ldots & \ldots & \lambda_2 & \lambda_1 & 1 \end{bmatrix}.$$

It is elementary to see that the conditional distribution $\boldsymbol{Y}|\boldsymbol{\Xi}, \boldsymbol{h}, \boldsymbol{G}_J \sim \mathbf{N}(\boldsymbol{X}\boldsymbol{\Xi}', \boldsymbol{\Gamma})$ where $\boldsymbol{\Gamma} = \boldsymbol{G}_J\boldsymbol{V}\boldsymbol{G}_J'$, and $|\boldsymbol{G}_J| = 1$ such that $|\boldsymbol{\Gamma}| = \exp(\sum_{t=1}^{T} h_t)$. Therefore, the log-likelihood of the Gaussian GARFIMA(1,0)-SV model is

$$\log f(\boldsymbol{Y}|\boldsymbol{\Xi}, \boldsymbol{h}, \boldsymbol{G}_J) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} h_t - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Xi})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Xi}) \tag{3.8}$$

where $\boldsymbol{h} = (h_1, \ldots, h_T)'$.

The posterior sampler of the GARFIMA-SV model is globally a Gibbs sampler with 6

blocks. Essentially, we use a combination of Gibbs sampling and Metropolis-within-Gibbs to sample the full conditional posterior distributions.

In order to estimate u and d, we consider two independent truncated normal priors with support in the region where generalised long-memory holds

$$\begin{aligned} u &\sim \mathrm{N}(\mu_u, \sigma_u^2)1_{ud} \\ d &\sim \mathrm{N}(\mu_d, \sigma_d^2)1_{ud} \end{aligned}$$

where $1_{ud} = 1(\{-1 < u < 1, 0 < d < 0.5\} \cup \{|u| = 1, 0 < d < 0.25\})$ and 1 is an indicator function. Note that we impose Gegenbauer long-memory stationarity through the prior distributions of u and d.

As for other parameters, we assume the following independent priors

$$\boldsymbol{\Xi} \sim \mathrm{N}_2(\boldsymbol{\mu}_\Xi, \boldsymbol{\Sigma}_\Xi)1(|\boldsymbol{\Xi}| < 1), \quad \alpha \sim \mathrm{N}(\mu_\alpha, \sigma_\alpha^2), \quad \beta \sim \mathrm{N}(\mu_\beta, \sigma_\beta^2)1(|\beta| < 1), \quad \sigma^2 \sim \mathrm{IG}(\frac{a}{2}, \frac{b}{2})$$

where $\mathrm{N}_2(\cdot,\cdot)$ is the bivariate normal distribution, $\boldsymbol{\Sigma}_\Xi$ is a diagonal variance-covariance matrix and $\mathrm{IG}(\cdot,\cdot)$ is the Inverse-Gamma distribution. We assume that all the priors are independent such that

$$p(\boldsymbol{\Xi}, u, d, \alpha, \beta, \sigma^2) = p(\boldsymbol{\Xi})p(u)p(d)p(\alpha)p(\beta)p(\sigma^2).$$

Let $\boldsymbol{Y}^* = \boldsymbol{X}^*\boldsymbol{\Xi} + \boldsymbol{\varepsilon}$, where $\boldsymbol{Y}^*$ and $\boldsymbol{X}^*$ are the vector and matrix $\boldsymbol{G}_J^{-1}\boldsymbol{Y}$ and $\boldsymbol{G}_J^{-1}\boldsymbol{X}$ respectively, such that $\boldsymbol{Y}^* \sim \mathbf{N}(\boldsymbol{X}^*\boldsymbol{\Xi}, \boldsymbol{V})$. Hence, it is standard to see the posterior distribution of $\boldsymbol{\Xi}$ is

$$\boldsymbol{\Xi}|u, d, \boldsymbol{h} \sim \mathrm{N}_2((\boldsymbol{X}^{*'}\boldsymbol{V}^{-1}\boldsymbol{X}^* + \boldsymbol{\Sigma}_{\boldsymbol{\Xi}}^{-1})^{-1}\boldsymbol{X}^{*'}\boldsymbol{V}^{-1}\boldsymbol{Y}^*, (\boldsymbol{X}^{*'}\boldsymbol{V}^{-1}\boldsymbol{X}^* + \boldsymbol{\Sigma}_{\boldsymbol{\Xi}}^{-1})^{-1})1(|\boldsymbol{\Xi}| < 1). \tag{3.9}$$

Note that sampling from (3.9) is a draw from a truncated bivariate normal distribution; see Robert (1995) for details.

Once the mean structure has been sampled, we then concentrate out the volatility process given by $\boldsymbol{\varepsilon} = \boldsymbol{G}_J^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Xi})$. One of the earliest samplers to estimate the SV model is that of Kim, Shephard, and Chib (1998) which measures the stochastic volatility model using a mixture of linear Gaussian models. We do not describe the details due to space constraints, but direct readers to the original article. We first

sample from

$$\alpha|\boldsymbol{h},\beta,\sigma^2 \sim \mathrm{N}\left(V_\alpha\left(\frac{(1-\beta^2)h_1+(1-\beta)\sum_{t=1}^{T-1}(h_{t+1}-\beta h_t)}{\sigma^2}+\frac{\mu_\alpha}{\sigma_\alpha^2}\right),V_\alpha\right)$$

where

$$V_\alpha=\left(\frac{1-\beta^2+(T-1)(1-\beta)^2}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)^{-1}.$$

In order to sample β, we use a specialized version of a Metropolis-Hastings algorithm first suggested by Chib and Greenberg (1994). Given the current value $\beta^{(m-1)}$ at the $(m-1)^{th}$ iteration, sample a proposal value β^* from $\mathrm{N}(\hat{\beta},V_\beta)$ where $\hat{\beta}=\sum_{t=1}^{T-1}(h_{t+1}-\alpha)(h_t-\alpha)/\sum_{t=1}^{T-1}(h_t-\alpha)^2$ and $V_\beta=\sigma^2/\sum_{t=1}^{T-1}(h_t-\alpha)^2$. If $|\beta^*|<1$, then accept with probability $\min\{1,\varrho\}$ where $\varrho=\exp\left\{g(\beta^*)-g(\beta^{m-1})\right\}$ and

$$g(x)=\log p(x)+\frac{1}{2}\log(1-x^2)-\frac{(h_1-\alpha)^2(1-x^2)}{2\sigma^2}$$

and $p(\cdot)$ is the prior distribution of β. Under the assumption of a conjugate prior $\sigma^2 \sim \mathrm{IG}(\frac{a}{2},\frac{b}{2})$, the posterior distribution of σ^2 is a standard conjugate result given by

$$\sigma^2|\boldsymbol{h},\alpha,\beta \sim \mathrm{IG}\left(\frac{T+a}{2},\frac{b+(h_1-\alpha)^2(1-\beta^2)+\sum_{t=1}^{T-1}[h_{t+1}-\alpha-\beta(h_t-\alpha)]^2)}{2}\right).$$

The posterior of both u and d are complicated and do not have a tractable conjugate form, and subsequently samples from these distributions cannot be obtained directly. In order to sample from u and d, we use an approximation based on posterior modes from Gelman et al. (2013), coupled with a proposal distribution precision tuning algorithm which we conduct only within the burn-in period. Details of this are provided in Appendix A. We briefly note that attempts to estimate $[u,d]$ using the Metropolis algorithm proved futile due to extremely slow convergence, and "boundary trap" issues. Consider the following independence chain Metropolis-Hastings algorithm

1. Maximimize the log posterior of u and d to find the modes $\tilde{u}$ and $\tilde{d}$ respectively. The log posterior modes are found by maximising

$$\log p_u(u|d,\boldsymbol{\Xi},\boldsymbol{h})=\log f(\boldsymbol{Y}|d,\boldsymbol{\Xi},\boldsymbol{h})+\log \mathrm{N}(\mu_u,\sigma_u^2)1_{ud}$$
$$\log p_d(d|u,\boldsymbol{\Xi},\boldsymbol{h})=\log f(\boldsymbol{Y}|u,\boldsymbol{\Xi},\boldsymbol{h})+\log \mathrm{N}(\mu_d,\sigma_d^2)1_{ud}$$

maximising (minimizing) a univariate function on a fixed interval is easily and quickly performed on most routine statistical packages (`fminbnd` in `MATLAB`, `optimx` in `R`, etc).

2. Sample u^* from the proposal distribution $\text{N}(\tilde{u}, c_u^2 V_u)$ denoted by q_u, where c_u is the scaling parameter. See Appendix A for details.

3. Reject u^* unless $(\{-1 < u^* < 1, 0 < d < 0.5\} \cup \{|u^*| = 1, 0 < d < 0.25\})$.[1] Otherwise, accept u^* with probability ζ, where

$$\zeta = \min\left\{1, \frac{p_u(u^*|d, \boldsymbol{\Xi}, \boldsymbol{h}) q_u(u^{(m)})}{p_u(u^{(m)}|d, \boldsymbol{\Xi}, \boldsymbol{h}) q_u(u^*)}\right\}.$$

4. Repeat steps 2-3 by replacing d with d^* and u^* with u.

If we accept u^* and d^*, then we update $u^{(m+1)} = u^*$ and $d^{(m+1)} = d^*$ respectively, and generate the updated $\boldsymbol{G}_J$ using the new Gegenbauer parameters. We use the multi-move algorithm to sample $\boldsymbol{h}$ due to its superiority over the single-move sampler as documented in Kim, Shephard, and Chib, 1998, and full details are provided in Appendix B. The global sampling procedure is then repeated many times until we are sampling from the true parameter posterior distributions.

Further, under the advice of a referee, we include the derivation of the marginal likelihood in Appendix C so the GARFIMA-SV model can be easily compared to other processes.

3.3.2 Computational issues

The computational burden of evaluating the log-likelihood and the choice of J are contemporaneous to one another and deserve some commentary. Evaluating the log likelihood function can be a time consuming process during each MCMC sweep. In order to speed this procedure up, we make a few changes to exploit the nature of the problem.

Firstly, computing $\boldsymbol{\Gamma}^{-1}$ can happen hundreds if not thousands of times during the sampling of u^* and d^* due to the optimisation process. We take advantage of the

[1] Practically, when $u^* \geq 0.99$, then we set $u^* = 1$ in order to give the event $|u| = 1$ non-zero probabilities. This adjustment is also applied to the Metropolis-Hastings ratio of Step 3.

banded structure of $\mathbf{\Gamma}$ in order to do this. A much quicker alternative is to evaluate $(\gamma^{-1})^T(\gamma^{-1}) = \mathbf{\Gamma}^{-1}$ where γ is the Cholesky decomposition of $\mathbf{\Gamma}$. This reduces the computational time of evaluating $\mathbf{\Gamma}^{-1}$ by a factor of $\frac{O(T^3)}{2\times O(T)}$. Second, since $\boldsymbol{G}_J$ is sparse, we work with sparse matrix packages which are readily available on most statistical software. This speeds up the computation of the quadratic term in (3.8) by storing only non-zero elements of $\boldsymbol{G}_J$ together with their indices. Ultimately, this means the computational time is reduced by eliminating operations on zero elements. Finally, we adopt an improved MCMC algorithm for estimating the latent variable $\boldsymbol{h}$ using the so-called precision sampler of Chan, 2013. The novelty in the approach is using recent advances in state space simulation techniques to exploit the banded structure of $\mathbf{\Gamma}$. We avoid details not to detract from our argument and direct enthusiastic readers to the original article.

In actual fact, after these changes are made we still find the estimation of u and d can consume upwards of 50% of the overall computational time depending on T and J. We find that in practice, increasing the value of J is more computationally expensive than T. Evidently, the less sparse $\boldsymbol{G}_J$ becomes, the slower the evaluation of the quadratic term in (3.8) becomes.

Dissanayake, Peiris, and Proietti, 2016 discusses an optimal truncation point of the moving average Gegenbauer white noise innovation process from an expected mean square argument via simulation studies. It is found that an optimal lag order using the Kalman filter is between $[29, 35]$. Although our main focus is not to determine what the optimal lag order is, sensitivity analysis reveals that $J > 35$ does not greatly increase accuracy, however $J < 29$ does have some material impact on the results. We therefore find that using $J = 35$ is a good trade-off between accuracy and computational time.

3.4 Simulation studies

Our proposed model leaves some open-ended questions such as what sample size is required in order to achieve reliable results, and if the values of (u, d) have an impact on the estimation of the global model. Given these challenges, we see a comprehensive simulation study as a sensible choice to answer these questions. Clearly, the

limitation in what follows is the limited scope of parameter choices and priors. A more in depth analysis is mandatory before this model is applied in practice.

3.4.1 Parallelization issues of MCMC

We now discuss some computational issues with our simulation study, which are also applicable to our empirical applications in latter sections. MCMC is a notoriously computationally intensive exercise since chains are dependent on previous values and it proves difficult to invoke multiple computer cores to evaluate one single chain. However, the nature of our simulation study is embarrassingly parallel, and can be scaled across multiple computer cores relatively easily. Each MCMC chain is independent so our work is scaled up by creating multiple execution threads across multiple cores, across multiple machines. This can be concurrently executed on a multicore machine. Without parallel computing, each chain must be run sequentially which is an extremely time consuming task.

Randomness is what drives the parameter estimation procedure, and it is critical to generate uncorrelated randomness across multiprocessor cores. There are some issues pertaining to the simulation of random quantities which need to be addressed. By default, we assume that all simulated realizations of $\boldsymbol{\theta}$ are independent. This requires that not only randomness be achieved within cores, but also between cores. Standard pseudo-random number generators (PRNGs) are unsatisfactory in a parallel context, as executing the same command in parallel will result in the same stream. A widely used solution to this is to use a different seed on each processor. The most commonly used method in practice is to use a seed value equal to the current system time, or the system time multiplied by an number unique to the parallel loop (e.g., loop number). This is still however unsatisfactory.

By default, each stream is generated with the same deterministic function $\omega(\cdot)$ and has a finite periodicity. Clearly, there is a chance of overlap between streams as they are generated using the same function $\omega(\cdot)$ thereby inducing dependance. Correlation in pseudo-random number sequences can lead to serious and undetectable errors.

This can be resolved with the use of initial value parametrized PRNGs. In essence, each deterministic function $\omega(.)$ is parameterized according to the seed which is used. In our case, each function which generates the stream will be dependent on the seed

value of computer system time multiplied by loop number. This avoids any potential overlap between random streams, and therefore ensures each MCMC chain is independent and does not overlap. A popular choice is the so called multiplicative lagged-Fibonacci generator. The details of this algorithm are out of the scope of this paper, however interested readers are encouraged to see Mascagni and Srinivasan, 2004 for details.

All of our experiments are coded in MATLAB and run on a Dell PowerEdge R630 server with a Intel Xenon E5-2680V3 CPU and DDR3 128GB of resident RAM. Each node consists of 24 cores, and we are able to execute up to 256 threads at once. The SV sampling engine is implemented using a MATLAB script provided by Joshua Chan, which can be downloaded from his website[2].

3.4.2 In sample

For completeness, we first discuss how to simulate a GARFIMA(1,0)-SV process. We are compelled by the work of Bardet et al. (2003) to use a MA expansion over other means to simulate the process, as the authors show this to be more robust than the Durbin-Levinson algorithm, and the AR aggregation process of Granger (1980). Our method is general enough such that other mean processes can alternatively be considered (unobserved components, ARMA etc). It is important to note that simulating a long memory stochastic volatility model is the same, except we replace Step 3 of the following algorithm with ARFIMA moving average truncated coefficients; see Hosking (1981).

Algorithm 1

1. Initialize h_0 as $h_0 \sim \text{N}\left(\alpha, \frac{\sigma^2}{1-\beta^2}\right)$, and iterate h_t for $t = 1, \ldots, J, \ldots, T+R$ forward in time with transition equation (3.2) as $h_t = \alpha + \beta(h_{t-1} - \alpha) + \eta_t$ where η_t is a draw from $\text{N}(0, \sigma^2)$, and R is the burn-in period.[3]
2. Simulate the SV errors ε_t as $\exp(h_t/2)z_t \; \forall \, t$, where $z_t \sim \text{N}(0,1)$.
3. Generate the vector of Gegenbauer coefficients using (3.4).
4. Initialize $y_0 = \frac{\mu}{1-\phi}$, and iterate y_t forward in time as $y_t = \mu + \phi y_{t-1} + \sum_{j=0}^{J_t^*} \lambda_j z_{t-j}$ for $t = 1, \ldots, J, \ldots, T+R$, where $J_t{}^* = \min(t, J)$.

[2] http://people.anu.edu.au/joshua.chan/

[3] We use $R = 1,000,000$ always.

5. Discard the first R values.

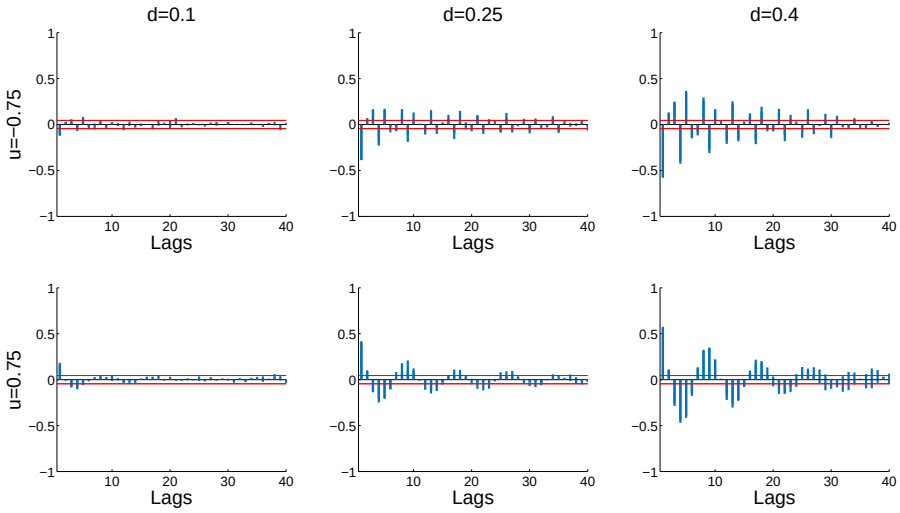

FIGURE 3.1: Sample autocorrelation functions (SACFs) for the 40 lag truncation point GARFIMA-SV process with $\mu = 0, \alpha = 0, \beta = 0.9, \sigma^2 = 0.2$. Clearly, the larger the value of d, the further away the process is from being purely randomness and invokes cyclicality into the autocorrelation structure. Interestingly, positive values of u introduce smoother autocorrelation cycles, whilst negative values of u cause jaggered autocorrelation patterns. These two properties are useful when identifying evidence of generalised long memory during the initial exploratory data analysis process undertaken by the time series analyst.

We now outline a comprehensive simulation study in order to assess the performance of our proposed sampling scheme. First, data is generated from a GARFIMA-SV model according to Equations (3.5) and (3.2) (using Algorithm 1) and the parameters are estimated subsequently. We consider $u = [-0.5, 0.5, 0.9, 1]$ and $d = [0.05, 0.25, 0.45]$ on the Gegenbauer parameter plane. The AR structure is set as $\mu = 0$ and $\phi = 0.90$, and the stochastic residuals are simulated according to the parameters $\alpha = 0, \beta = 0.95$ and $\sigma^2 = 0.2$. Our simulated process has the expression

$$(1 - 0.9B)(1 - 2uB + B^2)^d y_t = \varepsilon_t, \quad \varepsilon_t | F_{t-1} \sim \mathrm{N}(0, e^{h_t})$$
$$h_t = 0.95 h_{t-1} + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, 0.2).$$

Three different time lengths T are considered: $500, 1000, 1500$. The hyperparameter $\boldsymbol{\mu}_{\boldsymbol{\Xi}}$ is set to $\begin{bmatrix} 0 & 0.8 \end{bmatrix}'$, and we find that $\boldsymbol{\Sigma}_{\boldsymbol{\Xi}} = \begin{bmatrix} \sqrt{10} & 0 \\ 0 & 5 \end{bmatrix}$ is a sensible choice. μ_u is set to half the search region to 0, and similarly, μ_d is taken as the mid-point of the support region to be 0.25 for the Gegenbauer filter, and 0.125 for the long

memory filter. The variance is set as half the range so that $\sigma_u^2 = 1$ and $\sigma_d^2 = 0.25$ for the Gegenbauer case, and $\sigma_d^2 = 0.125$ for the plain long memory case. We find that setting V_u and V_d to be relatively larger works best with tuning. As such, V_u = 2 and $V_d = 0.50$ for the Gegenbauer case, and $V_d = 0.25$ for the plain long memory case.

For the SV parameters, the choice of prior is typically not very influential as the likelihood carries most of the information. We however briefly discuss the motivation behind our prior choices for completeness. A vague prior is typically used for α, but we however favour a slightly more informative prior depending on the nature of the time series. For most financial series, daily log returns data have a variance of less than 0.0001 which implies an α of $\log(0.0001) \approx -9$. On the other hand, percentage log returns exhibit a variance of 1, hence a log volatility of 0 is suitable. We will assume that our synthetic time series' are percentage log returns, so that $\mu_\alpha = 0$. As for σ_α^2, some popular choices in the literature are 5 (Jacquier, Polson, and Rossi, 1994), 1 (Omori et al., 2007) and $\sqrt{10}$ (Kim, Shephard, and Chib, 1998). We find good results with $\sigma_\alpha^2 = \sqrt{10}$. Kim, Shephard, and Chib (1998) notes that $\hat{\beta}$ is typically estimated ≈ 1 for most financial time series, and accepted with a high acceptance rate $(> 99\%)$. Therefore, we set $\mu_\beta = 0.95$ and $\sigma_\beta^2 = 5$ so that it is non-informative. The choice of hyperparameters for σ^2 is not very influential in most settings as long as it is kept away from 0. Hence, we use the prior choice of Kim of $a = 5$ and $b = 0.05$.

The process in equations (3.5) and (3.2) are simulated $\Omega = 1,000$ times, and estimated each time using the GARFIMA-SV model. We report the estimated mean of each parameter, the root mean squared error (RMSE) and the mean of the standard errors in parentheses . We use $M = 10,000$ iterates after a burn-in period of $10,000$. We purposely choose a burn-in period of half the total number of iterations under the advice of Gelman et al. (2013).

Table 1 reports our findings when the length of the time series is $T = 500$. We first note that when d is low, and u is close to 0, then $\hat{u}$ has a positive bias and $\hat{d}$ has a negative bias. Moreover, μ and ϕ are estimated quite accurately as the likelihood carries enough information about the observation equation. However, α has a slight upward bias, and σ^2 has a negative bias. Clearly, volatility is a latent process and a sample size of 500 is inadequate to provide enough information to estimate the parameters of the latent process. We see slightly more improved parameter estimates when T is set to $1,000$, in Table 2 but are still not adequate for inferential purposes.

A sample size giving more reliable results seems to be $T = 1,500$ as evidenced by Table 3. We see lower standard errors, as well as lower RMSEs, but however σ^2 is still underestimated. We particularly note the same effect is seen with lower values of σ^2, and is therefore not an artifact of the choices of hyperparameters. Clearly, σ^2 requires even larger sample sizes in order to attain greater accuracy. Indeed, the same phenomena is found using a plain SV model. Gelman (2006) notes that the inverse-gamma is a poor choice as when σ^2 is near 0, the resulting inferences are sensitive to a and b, and advises to use a non-informative prior instead. We however favour this specification due to the clean conjugacy properties and most importantly because as our purpose is focus on the relative merits of the generalised long-memory process and its long memory counterpart.

We include some diagnostics in Appendix D which deserve some commentary. Table $D.1$ are the Gelman-Rubin statistics for each parameter in the case of $T = 1,500$. We see that all parameters have converged as they are lower than 1.2 and close to 1. The only notable remark is we see that u tends to have a slightly higher statistic in the case when $d = 0.05$. This is expected as the process becomes "less long memory" as $d \to 0$. This is further reaffirmed with the SACF charts in Figures $D.1$ and $D.2$.

3.4.3 Out of sample

Time series forecasting in a Bayesian setting is an intuitive process. We provide details on how this is performed in our setting, and once again, the method is general enough to be applied to more complicated models. In essence, parameter vector draws from the posterior distribution are used to generate a new data set under the model. This new data set is used to make inferences after averaging out. This contrasts sharply to the frequentist who bases forecasting on one particular set of estimated parameter values.

The predictive density is particularly important to the Bayesian as it is not only used to forecast, but also to measure out-of-sample fitness via the so called Bayes Factor (BF). A distinct advantage here is that we are able to measure the uncertainty of our forecasted value via the posterior predictive variance, which is not readily available to the frequentist. In order to assess the goodness of fit between two competing models,

u	d	$\hat{u}$	RMSE($\hat{u}$)	$\hat{d}$	RMSE($\hat{d}$)	$\hat{\mu}$	RMSE($\hat{\mu}$)	$\hat{\phi}$	RMSE($\hat{\phi}$)	$\hat{\alpha}$	RMSE($\hat{\alpha}$)	$\hat{\beta}$	RMSE($\hat{\beta}$)	$\hat{\sigma}^2$	RMSE($\hat{\sigma}^2$)
0.100	0.100	0.220	0.120	0.077	0.023	-0.008	0.008	0.893	0.007	0.003	0.003	0.956	0.057	0.152	0.048
		(0.158)		(0.054)		(0.373)		(0.019)		(0.363)		(0.019)		(0.049)	
0.100	0.250	0.101	0.001	0.250	0.000	-0.025	0.025	0.895	0.005	0.031	0.031	0.955	0.055	0.153	0.047
		(0.015)		(0.030)		(0.351)		(0.019)		(0.350)		(0.019)		(0.049)	
0.100	0.400	0.100	0.000	0.399	0.001	-0.024	0.024	0.896	0.004	0.059	0.059	0.957	0.054	0.153	0.047
		(0.007)		(0.026)		(0.329)		(0.018)		(0.352)		(0.018)		(0.047)	
0.250	0.100	0.358	0.108	0.076	0.024	0.002	0.002	0.895	0.005	0.052	0.052	0.957	0.055	0.152	0.048
		(0.177)		(0.055)		(0.389)		(0.019)		(0.362)		(0.019)		(0.047)	
0.250	0.250	0.250	0.000	0.249	0.001	-0.001	0.001	0.895	0.005	0.029	0.029	0.956	0.055	0.154	0.046
		(0.015)		(0.028)		(0.354)		(0.019)		(0.368)		(0.019)		(0.051)	
0.250	0.400	0.250	0.000	0.400	0.000	0.008	0.008	0.896	0.004	0.034	0.034	0.956	0.054	0.154	0.046
		(0.007)		(0.026)		(0.356)		(0.017)		(0.352)		(0.017)		(0.049)	
0.500	0.100	0.567	0.067	0.075	0.025	0.005	0.005	0.895	0.005	0.007	0.007	0.957	0.055	0.153	0.047
		(0.186)		(0.055)		(0.384)		(0.022)		(0.356)		(0.022)		(0.047)	
0.500	0.250	0.501	0.001	0.251	0.001	-0.018	0.018	0.896	0.004	0.016	0.016	0.956	0.054	0.155	0.045
		(0.013)		(0.026)		(0.386)		(0.018)		(0.366)		(0.018)		(0.047)	
0.500	0.400	0.500	0.000	0.400	0.000	-0.009	0.009	0.895	0.005	0.037	0.037	0.957	0.055	0.151	0.049
		(0.006)		(0.024)		(0.374)		(0.019)		(0.358)		(0.019)		(0.049)	
0.750	0.100	0.743	0.007	0.100	0.000	0.004	0.004	0.890	0.010	0.037	0.037	0.956	0.060	0.154	0.046
		(0.097)		(0.035)		(0.452)		(0.034)		(0.371)		(0.034)		(0.048)	
0.750	0.250	0.751	0.001	0.251	0.001	-0.011	0.011	0.895	0.005	0.014	0.014	0.955	0.055	0.155	0.045
		(0.012)		(0.024)		(0.482)		(0.021)		(0.369)		(0.021)		(0.050)	
0.750	0.400	0.750	0.000	0.402	0.002	0.004	0.004	0.895	0.005	0.014	0.014	0.955	0.055	0.157	0.043
		(0.005)		(0.021)		(0.531)		(0.019)		(0.357)		(0.019)		(0.049)	

TABLE 3.1: In-sample simulation results when $T = 500$, $u = [0.1, 0.25, 0.5, 0.75]$, $d = [0.1, 0.25, 0.4]$, $\mu = 0$, $\phi = 0.90$, $\alpha = 0$, $\beta = 0.95$ and $\sigma^2 = 0.2$. $\hat{\theta}$ is reported as the in-sample mean of $10,000$ MCMC sweeps, and standard errors are reported in parentheses. u has a larger RMSE for lower values of d. $\hat{\phi}$ has a negative bias throughout all variations of $[u, d]$ but is correct to 2dp, whilst $\hat{\alpha}$ has a slight positive bias. $\hat{\beta}$ is also slightly positively biased while $\hat{\sigma}^2$ is negatively biased.

u	d	$\hat{u}$	RMSE($\hat{u}$)	$\hat{d}$	RMSE($\hat{d}$)	$\hat{\mu}$	RMSE($\hat{\mu}$)	$\hat{\phi}$	RMSE($\hat{\phi}$)	$\hat{\alpha}$	RMSE($\hat{\alpha}$)	$\hat{\beta}$	RMSE($\hat{\beta}$)	$\hat{\sigma}^2$	RMSE($\hat{\sigma}^2$)
0.100	0.100	0.127	0.027	0.098	0.002	-0.005	0.005	0.898	0.002	-0.002	0.002	0.953	0.052	0.178	0.022
		(0.082)		(0.030)		(0.257)		(0.013)		(0.261)		(0.013)		(0.039)	
0.100	0.250	0.099	0.001	0.249	0.001	0.002	0.002	0.898	0.002	0.005	0.005	0.952	0.052	0.177	0.023
		(0.009)		(0.022)		(0.242)		(0.013)		(0.267)		(0.013)		(0.039)	
0.100	0.400	0.100	0.000	0.399	0.001	0.007	0.007	0.897	0.003	0.019	0.019	0.952	0.053	0.177	0.023
		(0.005)		(0.020)		(0.231)		(0.013)		(0.267)		(0.013)		(0.038)	
0.250	0.100	0.270	0.020	0.098	0.002	0.013	0.013	0.898	0.002	0.004	0.004	0.952	0.052	0.177	0.023
		(0.085)		(0.030)		(0.280)		(0.013)		(0.264)		(0.013)		(0.038)	
0.250	0.250	0.250	0.000	0.250	0.000	0.002	0.002	0.898	0.002	-0.006	0.006	0.953	0.052	0.177	0.023
		(0.009)		(0.021)		(0.248)		(0.013)		(0.269)		(0.013)		(0.040)	
0.250	0.400	0.250	0.000	0.400	0.000	-0.002	0.002	0.898	0.002	0.005	0.005	0.953	0.052	0.177	0.023
		(0.004)		(0.019)		(0.241)		(0.013)		(0.276)		(0.013)		(0.038)	
0.500	0.100	0.511	0.011	0.098	0.002	0.015	0.015	0.898	0.002	0.012	0.012	0.953	0.052	0.177	0.023
		(0.075)		(0.027)		(0.265)		(0.014)		(0.253)		(0.014)		(0.039)	
0.500	0.250	0.500	0.000	0.250	0.000	0.002	0.002	0.898	0.002	-0.007	0.007	0.952	0.052	0.177	0.023
		(0.008)		(0.020)		(0.268)		(0.013)		(0.269)		(0.013)		(0.040)	
0.500	0.400	0.500	0.000	0.401	0.001	-0.007	0.007	0.897	0.003	0.001	0.001	0.953	0.053	0.177	0.023
		(0.004)		(0.018)		(0.252)		(0.013)		(0.273)		(0.013)		(0.040)	
0.750	0.100	0.757	0.007	0.102	0.002	-0.006	0.006	0.894	0.006	0.016	0.016	0.953	0.056	0.176	0.024
		(0.049)		(0.021)		(0.304)		(0.024)		(0.276)		(0.024)		(0.038)	
0.750	0.250	0.750	0.000	0.251	0.001	0.009	0.009	0.898	0.002	0.011	0.011	0.952	0.052	0.178	0.022
		(0.006)		(0.017)		(0.319)		(0.013)		(0.262)		(0.013)		(0.038)	
0.750	0.400	0.750	0.000	0.400	0.000	0.000	0.000	0.897	0.003	0.018	0.018	0.951	0.053	0.179	0.021
		(0.003)		(0.015)		(0.370)		(0.013)		(0.254)		(0.013)		(0.039)	

TABLE 3.2: In-sample simulation results when $T = 1,000$, $u = [0.1, 0.25, 0.5, 0.75]$, $d = [0.1, 0.25, 0.4]$, $\mu = 0$, $\phi = 0.90$, $\alpha = 0$, $\beta = 0.95$ and $\sigma^2 = 0.2$. $\hat{\theta}$ is reported as the in-sample mean of $10,000$ MCMC sweeps, and standard errors are reported in parentheses. The RMSE and standard errors have fallen for all parameters relative to $T = 500$. Notably, the estimation of $\hat{u}$ has improved when d is close to 0, but u is still least accurate for low values of d. All of the latent parameters are being estimated well except for σ^2. $\hat{\sigma}^2$ is still negatively biased.

u	d	$\hat{u}$	RMSE($\hat{u}$)	$\hat{d}$	RMSE($\hat{d}$)	$\hat{\mu}$	RMSE($\hat{\mu}$)	$\hat{\phi}$	RMSE($\hat{\phi}$)	$\hat{\alpha}$	RMSE($\hat{\alpha}$)	$\hat{\beta}$	RMSE($\hat{\beta}$)	$\hat{\sigma}^2$	RMSE($\hat{\sigma}^2$)
0.100	0.100	0.109	0.009	0.099	0.001	-0.005	0.005	0.898	0.002	-0.005	0.005	0.952	0.052	0.186	0.014
		(0.053)		(0.023)		(0.215)		(0.011)		(0.229)		(0.011)		(0.031)	
0.100	0.250	0.100	0.000	0.249	0.001	0.005	0.005	0.898	0.002	-0.015	0.015	0.951	0.052	0.186	0.014
		(0.007)		(0.018)		(0.193)		(0.010)		(0.228)		(0.010)		(0.033)	
0.100	0.400	0.100	0.000	0.400	0.000	-0.005	0.005	0.899	0.001	0.004	0.004	0.951	0.051	0.185	0.015
		(0.004)		(0.016)		(0.189)		(0.010)		(0.226)		(0.010)		(0.032)	
0.250	0.100	0.255	0.005	0.099	0.001	0.006	0.006	0.899	0.001	-0.013	0.013	0.952	0.051	0.186	0.014
		(0.045)		(0.020)		(0.221)		(0.011)		(0.219)		(0.011)		(0.033)	
0.250	0.250	0.250	0.000	0.250	0.000	-0.004	0.004	0.899	0.001	-0.009	0.009	0.951	0.051	0.186	0.014
		(0.007)		(0.018)		(0.211)		(0.010)		(0.218)		(0.010)		(0.032)	
0.250	0.400	0.250	0.000	0.399	0.001	0.006	0.006	0.899	0.001	0.009	0.009	0.951	0.051	0.185	0.015
		(0.003)		(0.016)		(0.201)		(0.010)		(0.229)		(0.010)		(0.033)	
0.500	0.100	0.504	0.004	0.099	0.001	-0.001	0.001	0.898	0.002	-0.004	0.004	0.951	0.052	0.185	0.015
		(0.044)		(0.019)		(0.223)		(0.011)		(0.228)		(0.011)		(0.032)	
0.500	0.250	0.500	0.000	0.251	0.001	-0.001	0.001	0.898	0.002	0.006	0.006	0.951	0.052	0.186	0.014
		(0.006)		(0.016)		(0.218)		(0.011)		(0.221)		(0.011)		(0.031)	
0.500	0.400	0.500	0.000	0.400	0.000	0.003	0.003	0.899	0.001	-0.001	0.001	0.952	0.051	0.184	0.016
		(0.003)		(0.015)		(0.214)		(0.010)		(0.222)		(0.010)		(0.033)	
0.750	0.100	0.751	0.001	0.101	0.001	-0.003	0.003	0.897	0.003	0.001	0.001	0.951	0.053	0.185	0.015
		(0.024)		(0.016)		(0.238)		(0.015)		(0.228)		(0.015)		(0.034)	
0.750	0.250	0.750	0.000	0.251	0.001	0.003	0.003	0.898	0.002	-0.007	0.007	0.951	0.052	0.186	0.014
		(0.005)		(0.014)		(0.272)		(0.010)		(0.218)		(0.010)		(0.032)	
0.750	0.400	0.750	0.000	0.400	0.000	0.020	0.020	0.899	0.001	0.008	0.008	0.952	0.051	0.186	0.014
		(0.003)		(0.013)		(0.293)		(0.010)		(0.226)		(0.010)		(0.032)	

TABLE 3.3: In-sample simulation results when $T = 1,500$, $u = [0.1, 0.25, 0.5, 0.75]$, $d = [0.1, 0.25, 0.4]$, $\mu = 0$, $\phi = 0.90$, $\alpha = 0$, $\beta = 0.95$ and $\sigma^2 = 0.2$. $\hat{\theta}$ is reported as the in-sample mean of $10,000$ MCMC sweeps, and standard errors are reported in parentheses. We once again see further improvements to the measurement of u for low values of d by increasing T. All of the Gegenbauer long memory parameters are being estimated well and are statistically significant with lower RMSE relative to $T = 500$ and $T = 1,000$. The latent parameters are once again estimated well, and $\hat{\sigma}^2$ has slightly improved.

we rely on the log Bayes factor which can be thought of as the Bayesian equivalent of the frequentists likelihood ratio test; see Geweke and Amisano (2011) for details.

First recall the one-step ahead predictive likelihood at time $T+1$ is defined as

$$\text{PL}_{T+1} = p(y_{T+1}|y_T^o) = \int_{\boldsymbol{\theta}} \text{p}(y_{T+1}|y_T^o, \boldsymbol{\theta}) \times p(\boldsymbol{\theta}|y_T^o) d\boldsymbol{\theta}.$$

where $y_T^o = [y_T, \ldots, y_1]$. In practice, we can compute this by "averaging out" the parameter vector $\boldsymbol{\theta}$ over iterates in the posterior sample as

$$\widehat{\text{PL}}_{T+1} \approx \sum_{m=1}^{M} p(y_{T+1}|y_T^o, \boldsymbol{\theta}^{(m)})/M. \tag{3.10}$$

If we are considering two competing models, A and B, for some given data $\boldsymbol{Y}$ the one-step ahead log Bayes factor K_{T+1} is given by

$$\begin{aligned} K_{T+1} &= \log\left(\frac{p_A(y_{T+1}|y_T^o)}{p_B(y_{T+1}|y_T^o)}\right) \\ &= \log\left(\frac{\int_{\boldsymbol{\theta}} p_A(y_{T+1}|y_T^o, \boldsymbol{\theta}_A) \times p_A(\boldsymbol{\theta}_A|y_T^o) d\boldsymbol{\theta}_A}{\int_{\boldsymbol{\theta}} p_B(y_{T+1}|y_T^o, \boldsymbol{\theta}_B) \times p_B(\boldsymbol{\theta}_B|y_T^o) d\boldsymbol{\theta}_B}\right) \\ &\approx \log\left[\frac{\widehat{\text{PL}}_{A,T+1}}{\widehat{\text{PL}}_{B,T+1}}\right] = \log\widehat{\text{PL}}_{A,T+1} - \log\widehat{\text{PL}}_{B,T+1}. \end{aligned} \tag{3.11}$$

The higher the log Bayes factor (3.11), the stronger the evidence is for model A over B. An advantage of the log Bayes factor is that it includes a model structure penalty and therefore protects against over fitting (Kass and Raftery, 1995). The log Bayes factor is cumulative over forecast horizons such that if we wish to evaluate the out-of-sample performance across several forecast periods $(T+1, \ldots, T+s)$, with each based on a data window of T observations, then the one-step ahead cumulative log Bayes factor (CBF_{T+s}) is

$$\text{CBF}_{T+s} = \sum_{t=T+1}^{T+s} K_t = \sum_{t=T+1}^{T+s} \left(\log\widehat{\text{PL}}_{A,t} - \log\widehat{\text{PL}}_{B,t}\right). \tag{3.12}$$

The information available to time T is the observed data, while out of sample prediction will begin at time $T+1$ and end at $T+s$. We now provide practical details on how one can generate the predictive density, and evaluate the predictive likelihood at time $T+1$ given the observer is currently at time T.

Algorithm 2

1. Iterate through the posterior sampling scheme M times so that we obtain $\boldsymbol{\theta}^{(m)}$ for $m = 1, \ldots, M$.

2. Denote the m^{th} Gegenbauer error at time t as $e_t^{(m)} = e^{h_t/2}\varepsilon_t$. For each m

   (a) Obtain a draw from the conditional density $h_{T+1}^{(m)}|\boldsymbol{\theta}^{(m)} \sim \text{N}\left(\alpha^{(m)} + \beta^{(m)}(h_T^{(m)} - \alpha^{(m)}), \sigma^{2(m)}\right)$.

   (b) Obtain the Gegenbauer errors recursively as $\hat{e}_t^{(m)} = y_t - \mu^{(m)} - \sum_{j=1}^{J_t^*} \lambda_j^{(m)} \hat{e}_{t-j}^{(m)}$ for $t = 1, \ldots, J, \ldots, T$, where $J_t{}^* = \min(t, J)$.

   (c) Evaluate the predictive likelihood at time $T+1$ as per (3.10) by evaluating $\text{N}\left(\mu^{(m)} + \sum_{j=1}^{J} \lambda_j^{(m)} \hat{e}_{T+1-j}^{(m)}, \exp\left\{h_{T+1}^{(m)}\right\}\right)$ at y_{T+1} denoted as $\text{PL}_{T+1}^{(m)}$.

   (d) Draw one sample from the predictive density at time $T+1$ from $\text{N}\left(\mu^{(m)} + \sum_{j=1}^{J} \lambda_j^{(m)} \hat{e}_{T+1-j}^{(m)}, \exp\left\{h_{T+1}^{(m)}\right\}\right)$ denoted as $\hat{y}_{T+1}^{(m)}$.

3. Evaluate $\widehat{\text{PL}}_{T+1} = \sum_{m=1}^{M} \text{PL}_{T+1}^{(m)}/M$.

4. Evaluate $\hat{y}_{T+1} = \sum_{m=1}^{M} \hat{y}_{T+1}^{(m)}/M$.

In order to further reaffirm the reliability of our model, we perform a out-of-sample simulation study. A GARFIMA(1,0)-SV model is simulated with parameters $u = 0.5, d = 0.2, \mu = 0, \phi = 0.90, \alpha = 0, \beta = 0.95$ and $\sigma^2 = 0.2$ of length $1,501$. The observation window is set to $1,500$, and we forecast one-step ahead using both the GARFIMA-SV and ARFIMA-SV models. This procedure is repeated $\Omega = 1,000$ times, with different simulated data sets. Our study can be summarized as follows

1. For $i = 1, \ldots, \Omega$

   (a) Simulate a GARFIMA(1,0)-SV model with parameters $u = 0.5, d = 0.2, \mu = 0, \phi = 0.90, \alpha = 0, \beta = 0.95$ and $\sigma^2 = 0.2$ of length $1,501$.

   (b) Forecast $\hat{y}_{1,501}|y_{1,500}$ and calculate the one-step ahead log Bayes factor denoted as $K_{T+1}^{(i)}$ for the i^{th} simulated data set, according to (3.11).

2. Evaluate the one-step ahead cumulative log Bayes factor as $\text{CBF}^{(i)} = \sum_{\tau=1}^{i} K_{T+1}^{(\tau)}$.

where $K_{T+1}^{(\tau)}$ represents the log Bayes factor of the τ^{th} simulated data set. Figure 3.2 (a) is the evolution of $\text{CBF}^{(i)}$ for $i = 100, \ldots, \Omega$ of the GARFIMA-SV model over the

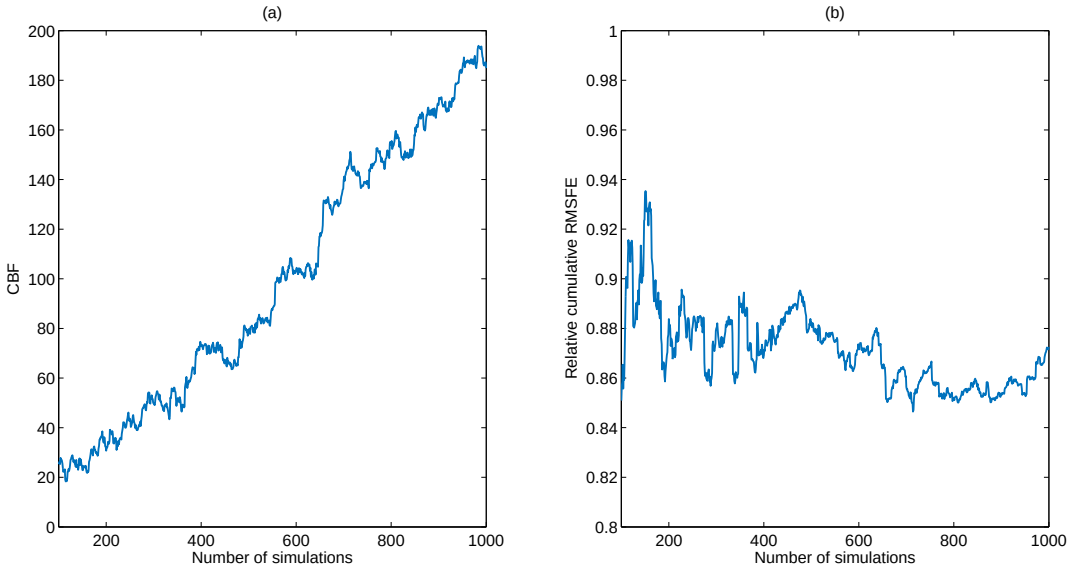

FIGURE 3.2: (a): Simulated GARFIMA-SV data one-step ahead cumulative log Bayes factor: GARFIMA-SV Vs. ARFIMA-SV. (b): Simulated GARFIMA-SV data cumulative relative RMSFE: GARFIMA-SV Vs. ARFIMA-SV. The GARFIMA-SV model is clearly the superior model choice under both sets of criteria. Note that both graphs start from 100 simulations to avoid distorting the vertical axis of Figure (b).

ARFIMA-SV model across the $\Omega = 1,000$ replicates (we ignore the first 99 values for aesthetic reasons). Clearly, the more positive this value is, the more we favour the GARFIMA filter over the ARFIMA filter. There is clear evidence the GARFIMA model is by far the favored model.

In Figure 3.2 (b), we again show the evolution of the relative cumulative root mean squared forecast error of both models, which is calculated as

$$\text{Relative cumulative RMSFE}^{(i)} = \left(\frac{\sum_{\tau=1}^{i} (\hat{y}_{A,T+1}^{(\tau)} - y_{T+1}^{(\tau)})^2}{\sum_{\tau=1}^{i} (\hat{y}_{B,T+1}^{(\tau)} - y_{T+1}^{(\tau)})^2} \right)^{0.5} \tag{3.13}$$

where, for example, $\hat{y}_{A,T+1}^{(\tau)}$ represents the one-step ahead forecast of the τ^{th} simulated data set under model A. The most interesting feature is that once again the GARFIMA model is superior. Clearly, the limitations of this study are the parameter choices.

3.5 Empirical Evidence

We now focus our attention to empirical data applications to further investigate the performance of the GARFIMA-SV model.

3.5.1 Sample fit: U.S. Consumer Price Index (CPI)

We now compare the ARFIMA-SV model and the GARFIMA-SV model using empirical data. A popular time series that is commonly used throughout the long-memory literature is CPI (such as (Mandelbrot, 1969; Baillie, 1996; Geweke and Porter-Hudak, 1983) to name a few). Sustained periods of deflation are ultimately bad for growth assets. Interestingly, there has been a growing social interest in deflation as there has been a rising trend in Google searches for deflation relative to inflation since the start of 2014. CPI exhibits long memory properties presumably from the argument proposed by Granger and Joyeux (1980) that the aggregation of first order Markov processes leads to a long memory process. CPI is therefore a natural candidate since it is an aggregation of several separate time series.

Bos, Koopman, and Ooms (2014a) successfully used a Monte Carlo maximum likelihood procedure to fit a LM-SV model to US CPI. The data set is composed of monthly CPI observations from January 1965 to May 2011. Although more data is available now, we deliberately use the same observation window as the original authors. The model structure which we will use is exactly the same one as proposed by the original authors, which includes an AR(1) parameter with constant, seasonal AR factors at lags $11, 12, 13$, and a outlier variable for the month of July, 1980 and we denote this vector of coefficients as $\boldsymbol{\beta}^* = [\phi, \phi_{11}, \phi_{12}, \phi_{13}, b]$ respectively.

We construct the same series by considering the log percentage change $\pi_t = 100 \log(P_t/P_{t-1})$ where P_t is the CPI index at time t. We then de-seasonalize the data by regressing π_t onto fixed seasonal dummies without a constant as $\pi_t = D\boldsymbol{\beta} + r_t$ where D are seasonal dummies. Instead of using the inflation dataset π_t, the authors use $\hat{y}_t = \hat{r}_t + \bar{\pi}_t$ where $\hat{r}_t$ are the residuals after regressing out statistically significant seasonal factors, and $\bar{\pi}_t$ is the average level of inflation equal to 0.34. As shown in Figure 3.3, the sample autocorrelation plot exhibits a very slow rate of hyperbolic decay. This is highly typical of ARFIMA models.

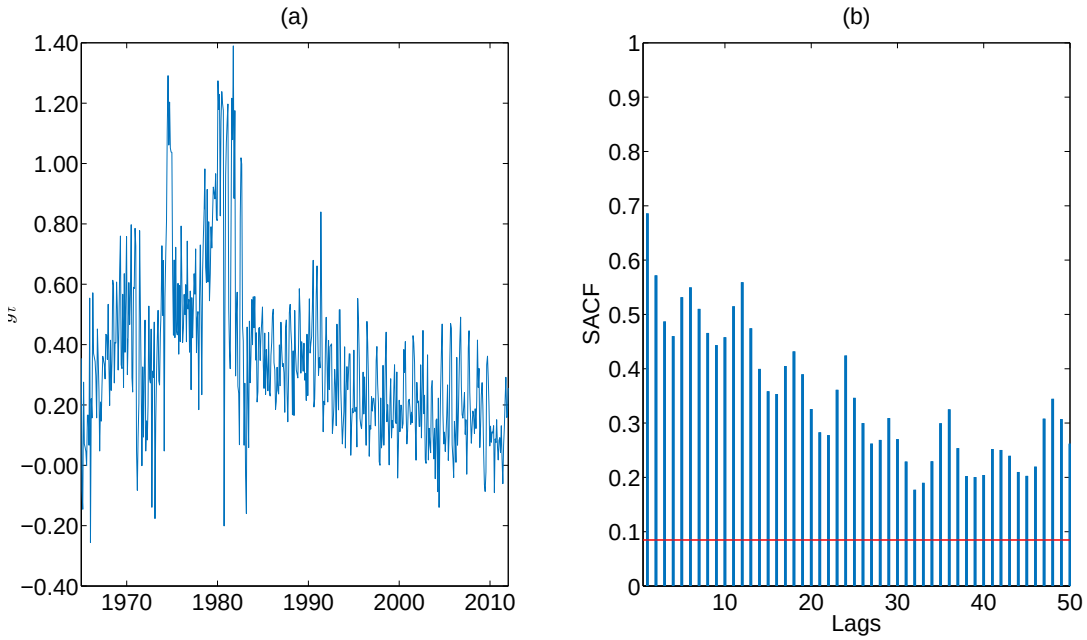

FIGURE 3.3: (a): 100 $\times$ log difference of deseasonalized U.S. CPI plot from January 1965 to May 2011. (b): Sample autocorrelation plot.

Our main focus here is not to model the complexities of U.S. CPI, so we refer the reader to the original paper for an excellent and detailed analysis. The main aim of this paper is to assess the effects of including generalised long memory. We use the same priors as before, and the priors of $[\phi_{11}, \phi_{12}, \phi_{13}]$ are each respectively U$[-1, 1]$, and the prior of b is $N(0, 10)$. We also increase the number of loops to $100,000$ and follow the advice of Gelman et al. (2013) once again and use a burn-in period of $50,000$. It is duly important to note here after taking into consideration the main findings of our simulation study, a sample size of 567 will have a material impact on the parameter estimates. We however continue with the analysis being mindful that sample sizes greater than $1,500$ are ideal. There are no parallelization issues here as there is only one data set that is run on a single core. Our main findings are summarized in Table 3.4.

The first striking and arguably most interesting parameter in the table is $\hat{u}$ which is estimated as 1, so that the GARFIMA filter has collapsed to the regular ARFIMA model specification. As the theory would dictate, the Gegenbauer $\hat{d}^{\text{GARFIMA}}$ should be half of the long memory parameter $\hat{d}^{\text{ARFIMA}}$, which is indeed the case here. Our Bayesian sampler estimated a long memory parameter of 0.274, which is similar to the value of 0.287 found using Monte Carlo maximum likelihood in Bos, Koopman, and Ooms (2014a). All remaining parameters are also very close, except for σ^2 which differs from 0.0172.

The highest posterior density (HPD) for each parameter does not arouse suspicion, apart from ϕ_1 which includes 0. This also is in tune with the work of the original authors as ϕ_1 was found to be statistically insignificant at a conventional level of significance. The Gelman-Rubin (GR) convergence statistics are all close to 1, which indicates all parameters have converged successfully. The MATLAB script which computes the GR statistic calculation is provided by Simo Sarkka and can be downloaded from the Aalto University School of Science website.[4]

More importantly, we note the log-likelihood for both models are equivalent (as expected), and our method reports a slightly higher log-likelihood of 278 compared to that found by the original authors of 252. The errors are not normally distributed, and serially correlated up to 24 lags as was found to be the case in Bos, Koopman, and Ooms (2014a) also.

3.5.2 Out-of-sample fit: U.S. Equity Risk Premium (ERP)

The U.S. equity risk premium is the excess return that equities provide over and above risk-free assets. It is the premium that investors are earning in compensation for holding onto riskier assets. Invariably, this translates into riskier stocks earning a higher risk premium. Understanding the ERP is important for several reasons. From a practical point of view, practitioners consider the ERP as a forward looking metric of the future state of the economy. Quarterly ERP figures turned negative 3 months before The Great Depression and right before the Global Financial Crisis (GFC). It is therefore a relevant metric to provide further scope to justify any claims of a looming global recession. Practically, the ERP is important as it can be viewed as a receptacle

[4] http://becs.aalto.fi/en/research/bayes/mcmcdiag/

	Bos, Koopman, and Ooms (2014a) estimates		Our ARFIMA(1,0)-SV estimates				Our GARFIMA(1,0)-SV estimates		
	$\hat{\boldsymbol{\theta}}$	s.e.	$\hat{\boldsymbol{\theta}}$	H.P.D.	s.e.	GR	$\hat{\boldsymbol{\theta}}$	H.P.D.	s.e.	GR
u							1.000	(1.000, 1.000)	(0.001)	1.000
d	0.287	(0.06)	0.274	(0.161, 0.388)	(0.0580)	1.000	0.134	(0.089, 0.182)	(0.024)	1.001
μ	0.193	(0.12)	0.089	(0.041, 0.142)	(0.026)	1.000	0.089	(0.041, 0.137)	(0.025)	1.000
ϕ	0.052	(0.07)	0.069	(−0.069, 0.210)	(0.071)	1.000	0.075	(−0.054, 0.202)	(0.066)	1.001
ϕ_{11}	0.130	(0.04)	0.142	(0.069, 0.216)	(0.038)	1.000	0.142	(0.069, 0.217)	(0.038)	1.000
ϕ_{12}	0.378	(0.04)	0.372	(0.297, 0.452)	(0.040)	1.000	0.371	(0.293, 0.449)	(0.040)	1.000
ϕ_{13}	0.067	(0.05)	0.085	(−0.007, 0.172)	(0.045)	1.000	0.083	(−0.004, 0.170)	(0.045)	1.000
b	-1.245	(0.24)	−1.114	(−1.668, −0.524)	(0.287)	1.000	−1.112	(−1.678, −0.531)	(0.288)	1.000
α	-3.862	*	−3.695	(−4.554, −2.845)	(0.510)	1.000	−3.699	(−4.523, −2.834)	(0.501)	1.000
β	0.985	(0.01)	0.965	(0.933, 0.997)	(0.017)	1.001	0.965	(0.933, 0.997)	(0.016)	1.001
σ^2	0.017	*	0.059	(0.032, 0.092)	(0.016)	1.001	0.059	(0.032, 0.092)	(0.016)	1.002
LL	252.086		278.459				278.482			
N	41.517	[0.00]	153.846		[0.00]		40.516		[0.00]	
Q	56.439	[0.00]	153.846		[0.00]		335.338		[0.00]	
AR_d	-		39.87%				36.83%			
AR_u	-						39.89%			

TABLE 3.4: MCMC estimates for the parameters of the ARFIMA(1,0)-SV as estimated by Bos, Koopman, and Ooms (2014a), and as estimated by our procedure and the GARFIMA(1,0)-SV model. H.P.D: 95% highest posterior density. s.e: Standard error. GL: Gelman-Rubin convergence statistic. LL: posterior mean of the log-likelihood function of y_t. N: Normality tests of (Doornik and Hansen, 2008). Q: Ljung-Box test statistic for 24 lags. AR: Acceptance rate of the relevant long-memory parameter(s). * indicates that it is not possible to find the equivalent standard error due to alternative transformations of the SV model which are reported by Bos, Koopman, and Ooms (2014a). The relevant p-value of each test is reported in brackets.

of investor sentiment.

Unfortunately, there is no consensus on what constituents should be used to construct the ERP. We instead favour popular measures of "risky" and "risk-free" assets. For risky assets, we use the percentage daily returns of the SP500. For "risk free" assets, we use the daily percentage change of US 1 year constant treasuries with a constant maturity rate. We first calculate the return series for each series and then map both onto a date vector. The ERP is then calculated as the difference between two observations for all relevant dates.

We further postulate that including the GFC is not a true representation of the data generating process, and begin our observation period from the 2^{nd} of March, 2009 when the market reached its low. We use data available up to the 31^{st} of October, 2016, which gives a total of $1,836$ data points. For our forecasting exercise, we fix an observation window of $1,500$ and slide forward in time. We therefore have a forecast horizon of 62 days, which corresponds from the 9^{th} of March to the 31^{st} of May.

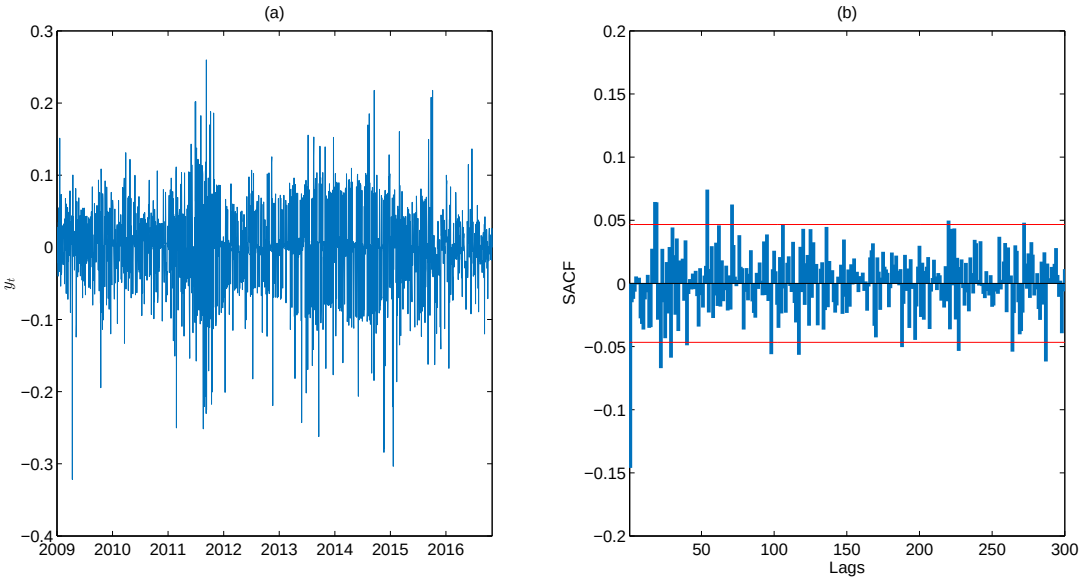

FIGURE 3.4: (a): U.S. equity risk premium from $02/03/2009$ to $31/10/2016$. (b): Sample autocorrelation plot.

Figure 3.4(a) depicts clear bursts of volatility which seem highly persistent and are strongly reminiscent of time varying volatility. There is also evidence of persistent autocorrelations which exist at higher order lags in Figure 3.4(b). Clearly, if we compare Figure 3.4(b) with Figure 3.1, we see the SACF of the ERP data is suggestive of a GARFIMA-SV model with a large and negative value of u, and a small value of d. The clear defining attributes which set this a part from a MA(1) model are the

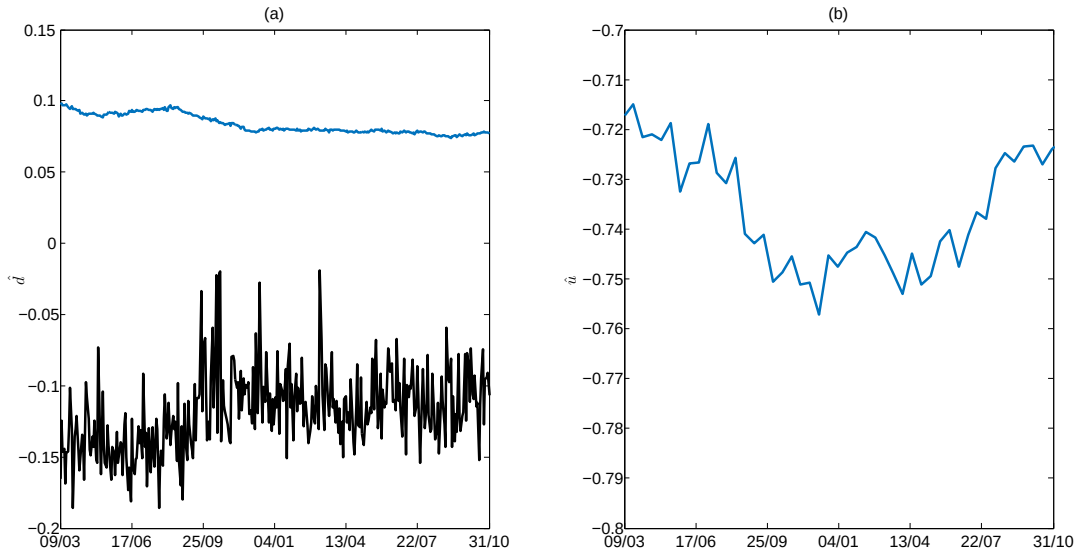

FIGURE 3.5: Rolling window long memory parameter estimates across forecast horizon. (a): ARFIMA-SV estimate of $\hat{d}$ (black) and GARFIMA-SV estimate of $\hat{d}$ (blue). (b): GARFIMA-SV estimate of $\hat{u}$.

persistent and statistically significant autocorrelations at higher lags. We assume the same priors for all remaining hyperparameters as before, $10,000$ iterates after a burn-in period of $10,000$ and forecast one-step ahead. As shown in Figure 3.5 (a), the long-memory parameter $\hat{d}$ of the ARFIMA-SV model is estimated as -0.11, which suggests that it is not long-memory. The acceptance rate of the long memory parameter is 16% on average across all time periods.

The Gegenbauer parameter estimates however read a different story and clearly suggest long-memory stationarity. The value of $\hat{d}$ is estimated as 0.08 on average, $\hat{u}$ is also estimated as -0.73, and the acceptance rate is 32%. Evidently, the ARFIMA-SV model could not detect a presence of long memory, where as the GARFIMA-SV model did so. The long-memory parameter estimates of the GARFIMA-SV model are consistent with the suggestions of the SACF.

We compare our model with the MA(1)-SV for two reasons. First, the SACF of the ERP data may be suggestive of a moving-average lagged-1 model. The second motivation stems from the work of Stock and Watson (2007) who found the MA(1) model to be superior to the model of Atkeson and Ohanian (2001), AR(p) where p is estimated according to the AIC, AR(4), the Nelson-Schwert model, an unobserved components stochastic volatility model, and two fixed MA coefficient models when forecasting US GDP inflation one-step ahead from 1970-1983 and 1984-2004. The estimation

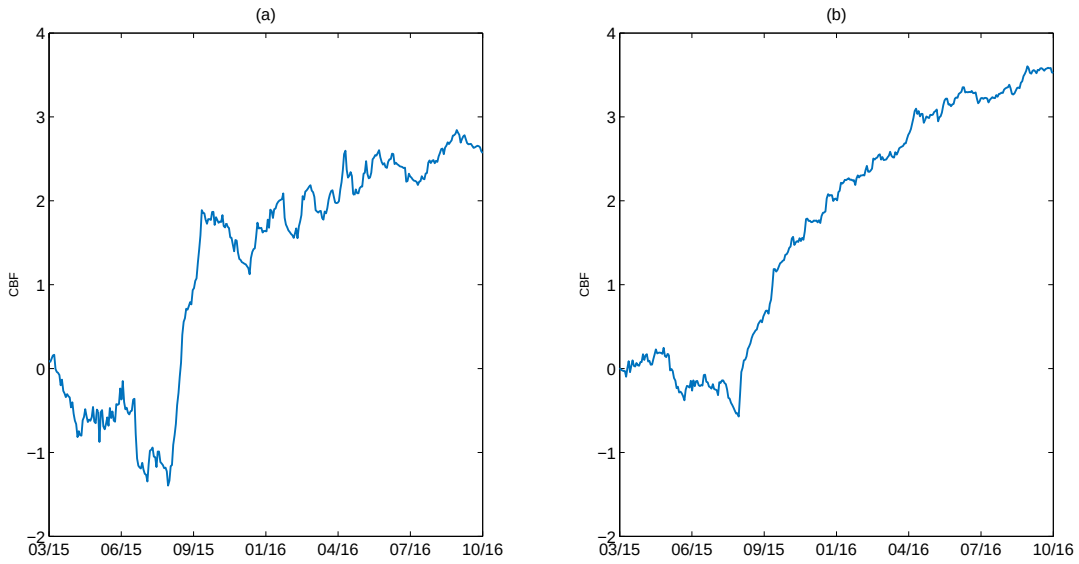

FIGURE 3.6: (a): U.S. equity risk premium (2005-2014) one-step ahead cumulative log Bayes factor: GARFIMA-SV Vs. ARFIMA-SV. (b): U.S. equity risk premium (2005-2014) one-step ahead cumulative log Bayes factor: GARFIMA-SV Vs. MA(1)-SV.

procedure is cited in Chan (2013).

As shown in Figure 3.6 the relative one-step ahead cumulative log Bayes factor CBF_{T+s}, given in (3.11) across forecast periods, further reaffirms our rolling parameter estimates and reveal the preferred model the GARFIMA-SV. As per Figure 3.6 (b) the GARFIMA-SV model is also superior compared to the MA(1)-SV model In both cases, we see a sharp increase in the CBF around August of 2015, and a gradual increase soon afterward. The reason for this is due to the data window moving further away from the GFC, and more towards a consistent regime. This intriguing behavior is highly suggestive of threshold effects.

3.6 Conclusion and future research

High profile economists have notably pointed out that due to a downward trend in global commodity prices, together with looming talks of deflation and subdued company profits, the potential of a global recession is indeed a reality. This is coupled with unusually high bouts of persistent volatility in global equity markets. We provide a statistical handle which measures this by discussing the estimation of the GARFIMA-SV model. We take several approaches to speed up our work. First, by exploiting

the banded structure of the covariance matrix we greatly speed up the evaluation of the likelihood function. Second, due to the nature of MCMC we use parallel computing which needs some detail to ensure is being conducted correctly. To validate our method, a comprehensive in and out of sample simulation study was performed, and good results are found. Finally, we apply our model to the US CPI and the US ERP which have attracted attention as of late. The GARFIMA-SV model is found to be equivalent to the plain long memory stochastic volatility model when forecasting inflation, but found to be superior when considering the equity risk premium. Potential avenues for future research include incorporating generalised error distributions, leverage effects and in particular switching regimes.

Chapter 4

Extensions to leverage and heavy tails for Cryptocurrency modelling

"An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem"

John Tukey

Chapter 3 proposed the GMA-SV model by combining generalised long memory with stochastic error processes as

$$\begin{aligned} y_t|h_t &= \mu + \sum_{j=0}^{J_t^*} \lambda_j \exp(h_{t-j}/2)\epsilon_{t-j}^*, \quad \epsilon_t^* \sim \mathrm{N}(0,1), \\ h_{t+1}|h_t &= \alpha + \beta(h_t - \alpha) + \sigma\eta_t^*, \quad \eta_t^* \sim \mathrm{N}(0,1), \\ h_1 &= \alpha + \sigma/\sqrt{1-\beta^2}\eta_1^*, \quad \eta_1^* \sim \mathrm{N}(0,1), \end{aligned}$$

where $J_t^* = \min(t, J)$. A natural contender to extend this model is to include other commonly discussed financial effects such as leverage (Bensoussan, Crouhy, and Galai, 1994; Bouchaud, Matacz, and Potters, 2001; Yu, 2005) and heavy tailed distributions (Liesenfeld and Jung, 2000; Asai, 2008; Chan and Hsiao, 2014). However, before discussing these extended models (see Appendix E for a list of these models) and their application to Cryptocurrency modelling in Chapter 5, this chapter first presents some model structures and Bayesian MCMC sampling techniques for estimating the parameters of these extended models. These techniques are applied and tested vigorously through several simulation studies to provide evidence for the extended models in Chapter 5 based on our second paper, which has been invited for review in *Econometrics and Statistics*.

Cryptocurrencies are a very popular topic that stem from the computer science literature and we believe they will shape the world moving forward. As such, we conclude this chapter by formally providing a small and intuitive introduction on Cryptocurrencies for the purpose of understanding the remainder of this thesis. We begin with the modelling structure of the preliminary SV model with leverage (SV-LVG) in the next section.

4.1 Stochastic volatility model with leverage

The first important SV model to extend is the SV-LVG model without long memory. It would be intuitive to assume that Cryptocurrencies are heavily dependent on news, since they are financial time series at their core. There are a few approaches to model the asymmetric dependency between the return of yesterday and the volatility of today. They are approximately divided into a conditional and a marginal approach and each approach has its advantages and disadvantages. We provide a review of the conditional approach in Chapter 7.1 and in Chapters 4 and 5, we adopt the marginal approach. In essence, the marginal approach models the asymmetric dependency through a correlation parameter which links the distributions of returns and volatilities together. This approach has been widely celebrated in the past decade and was popularized by Yu (2005). The model of Yu (2005) is given by

$$
\begin{aligned}
y_t|h_t &= \exp\left(h_t/2\right)\varepsilon_t, && (4.1)\\
h_{t+1}|h_t,\alpha,\beta,\sigma^2,\rho &= \alpha+\beta(h_t-\alpha)+\sigma\eta_t, && (4.2)\\
\text{Corr}(\varepsilon_t,\eta_t) &= \rho. && (4.3)
\end{aligned}
$$

This basic SV-LVG is presented in Appendix E as Model 2.

4.1.1 Factorization of the bivariate model

The following lemma shows how a bivariate distribution can be expressed as a marginal and a conditional distribution. This factorization facilitates model implementation in Bayesian MCMC samplers.

Lemma 1

We first recall the result that in general, the multivariate vector $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be partitioned as

$$\boldsymbol{X} = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

such that:

1. $X^{(1)}$ and $X^{(2)}$ are independent if and only if $\Sigma_{12} = 0$; and
2. The conditional distribution of $X^{(1)}$ given $X^{(2)} = x^{(2)}$ is a multivariate normal with conditional mean vector

$$\mathbb{E}(X^{(1)}|X^{(2)} = x^{(2)}) = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}), \tag{4.4}$$

and covariance matrix

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \tag{4.5}$$

Using the results in (4.4) and (4.5) with

$$\begin{aligned} X^{(1)} &= y_t|h_t, X^{(2)} = h_{t+1}, \\ \mu^{(1)} &= 0, \mu^{(2)} = \alpha + \beta(h_t - \alpha), \\ \Sigma_{11} &= \exp(h_t); \quad \Sigma_{12} = \Sigma_{21} = \rho\sigma\exp(h_t/2); \quad \Sigma_{22} = \sigma^2, \end{aligned}$$

we have

$$\begin{aligned} \mathbb{E}(X^{(1)}|X^{(2)}) &= \mathbb{E}(y_t|h_{t+1}, h_t) \\ &= 0 + \rho\sigma\exp(h_t/2)\sigma^{-2}[h_{t+1} - \alpha - \beta(h_t - \alpha)] \\ &= \rho/\sigma\exp(h_t/2)[h_{t+1} - \alpha - \beta(h_t - \alpha)], \\ \mathbb{V}\text{ar}(y_t|h_{t+1}, h_t) &= \exp(h_t) - [\rho\sigma\exp(h_t/2)]^2/\sigma^2 \\ &= \exp(h_t)(1 - \rho^2), \end{aligned}$$

which is the same result as proposed in Yu (2005). This result can also be obtained from Lemma two.

Lemma 2

We note here the continuous time process briefly. It is more convenient to convert correlated Wiener processes in the observation and latent equation to uncorrelated Wiener processes so they are easier to sample. A transformation is sought such that

$$\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix},$$

where $z_i(t)$ are correlated Wiener processes with correlation ρ_{ij}, the $w_i(t)$ are uncorrelated Wiener processes and the (a_{ij}) are to be chosen such that the correlation structures of $z_i(t)$ are preserved. Since $\mathbb{E}[z_i(t)] = 0$ and $\mathbb{E}[(z_i)^2] = 1$ then it is easy to see that $\sum_{j=1}^{2} a_{ij}^2 = 1$ for $i = 1, 2$ so that their modulus' are preserved. Also, the condition $E[z_i(t)z_j(t)] = \rho_{ij}$ for $i \neq j$ imposes the condition $a_{ik}a_{jk} = \rho_{ij}$ for $i = 1, 2$ and $j = 1, 2$. This reduces to the conditions

$$\begin{aligned} a_{11}^2 + a_{12}^2 &= 1, \\ a_{21}^2 + a_{22}^2 &= 1, \\ a_{11}a_{21} + a_{12}a_{22} &= \rho. \end{aligned}$$

One possible set of solutions are $a_{11} = 1, a_{12} = 0, a_{21} = \rho, a_{22} = \sqrt{1-\rho^2}$. Thus

$$\begin{aligned} z_1(t) &= w_1(t), \\ z_2(t) &= \rho w_1(t) + \sqrt{1-\rho^2} dw_2(t). \end{aligned} \tag{4.6}$$

4.1.2 Model specification

If equation (4.6) is applied to the standard SV model in (4.1) and (4.2), then

$$\begin{aligned} y_t|h_{t+1},h_t &= \exp(h_t/2)(\rho\eta_t+\sqrt{1-\rho^2}\varepsilon_t) \\ &= \rho\exp(h_t/2)\eta_t+\exp(h_t/2)\sqrt{1-\rho^2}\varepsilon_t \\ &= \frac{\rho}{\sigma}\exp(h_t/2)[h_{t+1}-\alpha-\beta(h_t-\alpha)]+\exp(h_t/2)\sqrt{1-\rho^2}\varepsilon_t \\ &\sim \mathrm{N}\left[\frac{\rho}{\sigma}\exp(h_t/2)[h_{t+1}-\alpha-\beta(h_t-\alpha)],\exp(h_t)(1-\rho^2)\right]. \end{aligned} \quad (4.7)$$

Upon inspection of (4.7), it is clear the marginal distribution $y_t|h_t$, after taking expectation of $y_t|h_{t+1},h_t$ over h_{t+1}, has $E[y_t|h_t]=0$ and $V\mathrm{ar}[y_t|h_t]=\rho^2\exp(h_t)+\exp(h_t)(1-\rho^2)=\exp(h_t)$ so that $y_t|h_t\sim\mathrm{N}(0,\exp(h_t))$.

The two main differences between this model and the standard SV model are:

1. The observations y_t are now conditional on h_{t+1}. This means the posterior distribution of each parameter will also need to condition this.
2. Under the standard SV model, the two vectors $\boldsymbol{h}=(h_1,\ldots,h_T)$ and $\boldsymbol{y}=(y_1,\ldots,y_T)$ are modeled independently. Now, the SV-LVG model considers the joint distribution for the order pair (y_t,h_{t+1}) based on $\boldsymbol{h}=(h_2,\ldots,h_{T+1})$ and $\boldsymbol{y}=(y_1,\ldots,y_T)$ where h_1 follows a marginal distribution.

We denote further $\boldsymbol{h}_{1:T}=(h_1,\ldots,h_T)$.

The next section provides some estimation methodologies for the central Gegenbauer long memory stochastic volatility model with leverage (GMA-SV-LVG), where the long memory component is added back to SV-LVG model.

4.2 Bayesian inference for the GMA-SV-LVG model

4.2.1 Extension of the SV-LVG model to the GMA-SV-LVG model

As will be discussed later, Cryptocurrencies show evidence of long memory effects, therefore, coupling this with leverage and SV is a sensible choice to fully capture their

dynamics. The GMA-SV-LVG model is given by

$$
\begin{aligned}
y_t|h_t &= \sum_{j=0}^{J_t^*} \lambda_j \exp(h_{t-j}/2)\epsilon_{t-j}^*, && (4.8)\\
h_{t+1}|h_t &= \alpha + \beta(h_t - \alpha) + \sigma\eta_t^*, && (4.9)\\
h_1 &= \alpha + \frac{\sigma}{\sqrt{1-\beta^2}}\eta_1^*, \\
\begin{pmatrix} \epsilon_t^* \\ \eta_t^* \end{pmatrix} &\sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right). && (4.10)
\end{aligned}
$$

This model is described in detail in Chapters 5.2 and 5.3 and is referred to as Model 4 in Appendix E.

As it is less efficient to work with a bivariate distribution, the factorization in Section 4.1.1 is again considered here. Therefore, the bivariate distribution in (4.10) can be factorized into the marginal component $h_{t+1}|h_t$ and the conditional component $y_t|h_{t+1}, h_t$. We consider the conditional distribution of $\epsilon_t = \rho\eta_t^* + \sqrt{1-\rho^2}\epsilon_t^*$. Hence, the new expression for the conditional observation equation is

$$
\begin{aligned}
y_t|h_{t+1}, h_t =& \sum_{j=0}^{J_t^*} \lambda_j \exp(h_{t-j}/2)\epsilon_{t-j} \\
=& \sum_{j=0}^{J_t^*} \lambda_j \exp(h_{t-j}/2)(\rho\eta_{t-j}^* + \sqrt{1-\rho^2}\epsilon_{t-j}^*) \\
=& \sum_{j=0}^{J_t^*} \lambda_j \exp(h_{t-j}/2)\rho\eta_{t-j}^* + \sum_{j=0}^{J_t^*} \lambda_j \exp(h_{t-j}/2)\sqrt{1-\rho^2}\epsilon_{t-j}^* \\
=& \sum_{j=0}^{J_t^*} \lambda_j \exp(h_{t-j}/2)\frac{\rho}{\sigma}[h_{t+1-j} - \alpha - \beta(h_{t-j} - \alpha)] + \sum_{j=0}^{J_t^*} \lambda_j \exp(h_{t-j}/2)\sqrt{1-\rho^2}\epsilon_{t-j}^*,
\end{aligned} \tag{4.11}
$$

such that

$$
y_t|h_{t+1}, h_t \sim \mathrm{N}\left(\sum_{j=0}^{J_t^*} \lambda_j e^{h_{t-j}/2}\frac{\rho}{\sigma}[h_{t+1-j} - \alpha - \beta(h_{t-j} - \alpha)], (1-\rho^2)\sum_{j=0}^{J_t^*} \lambda_j^2 e^{h_{t-j}}\right). \tag{4.12}
$$

The paper of Yu (2005) uses OpenBUGS to estimate the model parameters. As such, there was no need to derive the posterior distributions of each parameter. However,

in various extensions of the GMA-SV-LVG model that are considered in this thesis, we encounter many non-standard posterior distributions in which OpenBUGS may not be very efficient. In light of this, the full conditional distributions for each parameter of the GMA-SV-LVG model are derived to facilitate our programming of the MCMC sampler using MATLAB.

4.2.2 Observational likelihood function of the GMA-SV-LVG model

The log-likelihood of a T-dimensional multivariate normal random variable $\boldsymbol{x}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is

$$\ell = \log L = -\frac{1}{2}\left(\log(|\boldsymbol{\Sigma}|) + (\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) + T\log(2\pi)\right).$$

The density of y_t can be written in matrix notation as

$$\boldsymbol{Y}|\boldsymbol{h}, \boldsymbol{G}_{J^*} \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}), \tag{4.13}$$

where $\boldsymbol{\Gamma} = (1-\rho^2)\boldsymbol{G}_{J^*}\boldsymbol{V}\boldsymbol{G}'_{J^*}$, and $|\boldsymbol{G}_{J^*}| = 1$ such that $|\boldsymbol{\Gamma}| = (1-\rho^2)\exp(\sum_{t=1}^{T} h_t)$. Chapter 3.3.1 defines $\boldsymbol{\mu}$ and gives a detailed description of this matrix representation. Therefore, the log-likelihood function is

$$\log f(\boldsymbol{Y}|\boldsymbol{h}, \boldsymbol{G}_{J^*}) = -\frac{T}{2}\log(2\pi(1-\rho^2)) - \frac{1}{2}\sum_{t=1}^{T} h_t - \frac{1}{2}(\boldsymbol{Y}-\boldsymbol{\mu})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}), \tag{4.14}$$

which will be used to derive the full posterior distribution for the purposes of inference in the next section.

4.2.3 Sampling the return level parameter μ in the GMA-SV-LVG model

We begin with deriving the posterior distribution for the non-zero mean μ of the observation equation. We assume models 4.8 and 4.9 are modified to include a constant term μ. Although this constant term is not considered in the return equation in the following chapters, it is still nonetheless a common addition to most SV models with leverage. As such, the complete posterior derivation including the non-zero mean μ is presented for completeness. After adding μ to (4.12), we have

$$y_t|h_{t+1}, h_t \sim \mathrm{N}\left(\mu + \sum_{j=0}^{J_t^*} \lambda_j e^{h_{t-j}/2} \frac{\rho}{\sigma}(h_{t+1-j} - \alpha - \beta(h_{t-j} - \alpha)), (1-\rho^2) \sum_{j=0}^{J_t^*} \lambda_j^2 e^{h_{t-j}}\right).$$

Then we have

$$\begin{aligned} p(\mu|\boldsymbol{Y}, \boldsymbol{h}, \alpha, \beta, \sigma^2, \rho, u, d) &\propto f(\boldsymbol{Y}|\boldsymbol{h}, \mu, u, d, \alpha, \beta, \sigma^2, \rho) f(h_1|\alpha, \beta, \sigma^2) f(\boldsymbol{h}|\boldsymbol{h}_{1:T}, \alpha, \beta, \sigma^2) \times p(\mu) \\ &\propto f(\boldsymbol{Y}|\boldsymbol{h}, \mu, u, d, \alpha, \beta, \sigma^2, \rho) \times p(\mu), \end{aligned}$$

where the prior $p(\mu) \sim \mathrm{N}(\mu_s, \sigma_s^2)$. Defining $V_t = \sum_{j=0}^{J_t^*} \lambda_j e^{h_{t-j}/2}$ and $Z_t = \sum_{j=0}^{J_t^*} \lambda_j^2 e^{h_{t-j}}$, the posterior distribution of μ is

$$\begin{aligned} &\propto \exp\left\{\frac{-1}{2(1-\rho^2)} \sum_{t=1}^{T} \frac{\left[y_t - \mu - \frac{\rho}{\sigma} V_t[h_{t+1} - \alpha - \beta(h_t - \alpha)]\right]^2}{Z_t}\right\} \times \exp\left\{-\frac{(\mu - \mu_s)^2}{2\sigma_s^2}\right\} \\ &\propto \exp\left\{\frac{-1}{2(1-\rho^2)} \left[\sum_{t=1}^{T} \frac{(y_t - \mu)^2}{Z_t} - 2\sum_{t=1}^{T} \left(\frac{y_t - \mu}{Z_t}\right) \frac{\rho}{\sigma} V_t[h_{t+1} - \alpha - \beta(h_t - \alpha)]\right]\right\} \\ &\quad \times \exp\left\{-\frac{\mu^2 - 2\mu\mu_s}{2\sigma_s^2}\right\} \\ &\propto \exp\left\{\frac{-1}{2(1-\rho^2)} \left[\sum_{t=1}^{T} \frac{-2y_t\mu + \mu^2}{Z_t} + \mu \sum_{t=1}^{T} \left(\frac{1}{Z_t}\right) \frac{2\rho}{\sigma} V_t[h_{t+1} - \alpha - \beta(h_t - \alpha)]\right]\right\} \\ &\quad \times \exp\left\{-\frac{\mu^2 - 2\mu\mu_s}{2\sigma_s^2}\right\} \\ &= \exp\left\{-\frac{1}{2}\left[\mu^2 \left(\sum_{t=1}^{T} \frac{1}{Z_t(1-\rho^2)} + \frac{1}{\sigma_s^2}\right)\right.\right. \\ &\qquad \left.\left. - 2\mu \left(\sum_{t=1}^{T} \frac{y_t - \frac{\rho}{\sigma} V_t[h_{t+1} - \alpha - \beta(h_t - \alpha)]}{Z_t(1-\rho^2)} + \frac{\mu_s}{\sigma_s^2}\right)\right]\right\}. \end{aligned}$$

Note that it can easily be shown that

$$\exp\left\{Ax^2 - 2Bx\right\} \quad \sim \quad \mathrm{N}(VM, V) \quad \text{where } M = B \text{ and } V = A^{-1}.$$

Therefore, we have

$$\mu|\boldsymbol{Y}, \boldsymbol{h}, \alpha, \beta, \sigma^2, \rho, u, d \sim \mathrm{N}\left(V_\mu M_\mu, V_\mu\right),$$

where

$$M_\mu = \left(\sum_{t=1}^{T} \frac{y_t - \frac{\rho}{\sigma} V_t (h_{t+1} - \alpha - \beta(h_t - \alpha))}{Z_t(1-\rho^2)} + \frac{\mu_s}{\sigma_s^2} \right) \text{ and } V_\mu \quad = \left(\sum_{t=1}^{T} \frac{1}{Z_t(1-\rho^2)} + \frac{1}{\sigma_s^2} \right)^{-1}.$$

Since the constant term μ is not considered in the remaining chapters, it is a straightforward exercise to replace it with zero in the remaining sections to derive the posterior conditional distribution for other parameters.

4.2.4 Sampling the volatility level parameter α in the GMA-SV-LVG model

Without the constant μ, the matrix in (4.13) becomes

$$\boldsymbol{Y}|\boldsymbol{h}, \boldsymbol{G}_{J^*} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Gamma}), \tag{4.15}$$

where the covariance matrix can be expressed as $\boldsymbol{\Gamma} = (1-\rho^2)\boldsymbol{G}_{J^*}\boldsymbol{V}\boldsymbol{G}'_{J^*}$ where $\boldsymbol{V} = \text{diag}(\boldsymbol{W} \circ \boldsymbol{W})$, $\boldsymbol{W} = (e^{h_1/2}, \ldots, e^{h_T/2})$ and $A \circ B$ refers to the Hadamard product of vectors A and B. To simplify the model structure, we consider the transformation

$$\boldsymbol{Y}^* = \boldsymbol{G}_{J^*}^{-1}\boldsymbol{Y}, \tag{4.16}$$

where $\boldsymbol{Y}^* = (y_1^*, \ldots, y_T^*)$ is now independent of the Gegenbauer long memory parameters u and d. From (4.7), the density of y_t^* is

$$\begin{aligned} f(y_t^*) &= \frac{1}{\sqrt{2\pi\xi_t(1-\rho^2)\exp(\frac{h_t}{2})}} \exp\left[-\frac{\left\{ y_t^* - \rho\frac{\exp(\frac{h_t}{2})}{\sigma}[h_{t+1} - \alpha - \beta(h_t - \alpha)] \right\}^2}{2(1-\rho^2)\xi_t \exp(h_t)} \right] \\ &= \frac{1}{\sqrt{2\pi\xi_t(1-\rho^2)\exp(\frac{h_t}{2})}} \exp\left[-\frac{\left\{ \frac{y_t^*}{\exp(h_t/2)} - \frac{\rho}{\sigma}[h_{t+1} - \alpha - \beta(h_t - \alpha)] \right\}^2}{2(1-\rho^2)\xi_t} \right] \\ \ln f(\boldsymbol{Y}^*) &= -\frac{1}{2}\sum_{t=1}^{T} \ln[2\pi\xi_t(1-\rho^2)] - \sum_{t=1}^{T}\frac{h_t}{2} - \sum_{t=1}^{T} \frac{\left\{ \frac{y_t^*}{\exp(h_t/2)} - \frac{\rho}{\sigma}[h_{t+1} - \alpha - \beta(h_t - \alpha)] \right\}^2}{2(1-\rho^2)\xi_t}. \end{aligned}$$

The conditional posterior distribution for α is given by

$$\begin{aligned}
& p(\alpha|\boldsymbol{Y}^*, \boldsymbol{h}, \beta, \sigma^2, \rho) \\
\propto\ & f(\boldsymbol{Y}^*|\boldsymbol{h}, \alpha, \beta, \sigma^2, \rho) f(h_1|\alpha, \beta, \sigma^2) f(\boldsymbol{h}|\boldsymbol{h}_{1:T}, \alpha, \beta, \sigma^2) \times p(\alpha) \\
=\ & \frac{1}{(\sqrt{2\pi})^T (\sqrt{1-\rho^2})^T \prod_{t=1}^T \exp(\frac{h_t}{2})} \exp\left\{ -\sum_{t=1}^T \frac{(\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha)))^2}{2(1-\rho^2)} \right\} \times \\
& \frac{\sqrt{1-\beta^2}}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(h_1 - \alpha)^2(1-\beta^2)}{2\sigma^2} \right\} \times \\
& \frac{1}{(\sqrt{2\pi}\sigma)^T} \exp\left\{ -\frac{\sum_{t=1}^T (h_{t+1} - \alpha - \beta(h_t - \alpha))^2}{2\sigma^2} \right\} \times \frac{1}{\sqrt{2\pi}\sigma_\alpha} \exp\left\{ -\frac{(\alpha - \mu_\alpha)^2}{2\sigma_\alpha^2} \right\}
\end{aligned} \tag{4.17}$$

where the prior $\alpha \sim \mathrm{N}(\mu_\alpha, \sigma_\alpha^2)$. We first consider the first product term in the posterior distribution (compared to the standard SV model), which is the contribution of $y_t^*|h_{t+1}, h_t$ due to the leverage effect

$$\begin{aligned}
& \frac{1}{(\sqrt{2\pi})^T (\sqrt{1-\rho^2})^T \prod_{t=1}^T \exp(\frac{h_t}{2})} \exp\left\{ -\sum_{t=1}^T \frac{(\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha)))^2}{2(1-\rho^2)} \right\} \\
\propto\ & \exp\left\{ -\frac{1}{2(1-\rho^2)} \sum_{t=1}^T \left[-2\frac{y_t^*}{e^{h_t/2}} \frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha)) + \frac{\rho^2}{\sigma^2}(h_{t+1} - \alpha - \beta(h_t - \alpha))^2 \right] \right\} \\
\propto\ & \exp\left\{ -\frac{1}{2(1-\rho^2)} \sum_{t=1}^T \left[2\frac{y_t^*}{e^{h_t/2}} \frac{\rho(1-\beta)}{\sigma}\alpha + \frac{\rho^2}{\sigma^2}(h_{t+1} - \alpha - \beta(h_t - \alpha))^2 \right] \right\} \\
\propto\ & \exp\left\{ -\frac{\rho}{2\sigma(1-\rho^2)} \sum_{t=1}^T \left[2\frac{y_t^*}{e^{h_t/2}}(1-\beta)\alpha + \frac{\rho}{\sigma}[\alpha^2(1-\beta)^2 - 2\alpha(h_{t+1} - \beta h_t)(1-\beta)] \right] \right\} \\
=\ & \exp\left\{ -\frac{1}{2}\left(\alpha^2 \frac{T(1-\beta)^2\rho^2}{(1-\rho^2)\sigma^2} - 2\alpha \sum_{t=1}^T \frac{(1-\beta)\rho}{(1-\rho^2)\sigma} \left[(h_{t+1} - \beta h_t)\frac{\rho}{\sigma} - \frac{y_t^*}{e^{h_t/2}} \right] \right) \right\},
\end{aligned}$$

since

$$\begin{aligned}
(h_{t+1} - \alpha - \beta(h_t - \alpha))^2) &= (h_{t+1} - \beta h_t - \alpha(1-\beta))^2 \\
&= (h_{t+1} - \beta h_t)^2 - 2(h_{t+1} - \beta h_t)\alpha(1-\beta) + \alpha^2(1-\beta)^2 \\
&= \alpha^2(1-\beta)^2 - 2\alpha(h_{t+1} - \beta h_t)(1-\beta) + \text{terms independent of } \alpha.
\end{aligned}$$

Next, we consider the initial volatility h_1 which is the second term inside the exponent in (4.17).

Term 2:

$$\begin{aligned} -2T_2 &= \frac{(h_1-\alpha)^2(1-\beta^2)}{\sigma^2} \\ &= \frac{(h_1^2-2h_1\alpha+\alpha^2)(1-\beta^2)}{\sigma^2} \\ &= \frac{(-2h_1\alpha+\alpha^2)(1-\beta^2)}{\sigma^2} + \text{terms independent of } \alpha \\ &= \frac{1-\beta^2}{\sigma^2}\alpha^2 - \frac{2(1-\beta^2)h_1}{\sigma^2}\alpha + \text{terms independent of } \alpha. \end{aligned} \tag{4.18}$$

Next, we consider the volatilities $\boldsymbol{h}$ which are the third term in (4.17).

Term 3:

$$\begin{aligned} -2T_3 &= \sum_{t=1}^{T} \frac{(h_{t+1}-\alpha(1-\beta)-\beta h_t)^2}{\sigma^2} \\ &= \sum_{t=1}^{T} \frac{-\alpha(1-\beta)h_{t+1}-\alpha(1-\beta)(h_{t+1}-\alpha(1-\beta)+\beta h_t)+\alpha(1-\beta)\beta h_t}{\sigma^2} \\ &= -\sum_{t=1}^{T} \frac{\alpha h_{t+1}+\alpha h_{t+1}-\alpha^2(1-\beta)-\alpha\beta h_t-\alpha\beta h_t}{\sigma^2(1-\beta)^{-1}} + \text{terms independent of } \alpha \\ &= -\sum_{t=1}^{T} \frac{2\alpha h_{t+1}-\alpha^2(1-\beta)-2\alpha\beta h_t}{\sigma^2(1-\beta)^{-1}} + \text{terms independent of } \alpha \\ &= \sum_{t=1}^{T} \frac{(1-\beta)}{\sigma^2(1-\beta)^{-1}}\alpha^2 - \sum_{t=1}^{T-1} \frac{2(h_{t+1}-\beta h_t)}{\sigma^2(1-\beta)^{-1}}\alpha + \text{terms independent of } \alpha. \end{aligned} \tag{4.19}$$

Finally, we consider the prior density $p(\alpha)$.

Term 4:

$$\begin{aligned} -2T_4 &= \frac{(\alpha-\mu_\alpha)^2}{\sigma_\alpha^2} \\ &= \frac{\alpha^2-2\alpha\mu_\alpha}{\sigma_\alpha^2} + \text{terms independent of } \alpha \\ &= \frac{1}{\sigma_\alpha^2}\alpha^2 - \frac{2\mu_\alpha}{\sigma_\alpha^2}\alpha + \text{terms independent of } \alpha, \end{aligned} \tag{4.20}$$

summing over these terms, we have

$$\begin{aligned} T_2+T_3+T_4 &= -\frac{1}{2}\left[\left(\frac{1-\beta^2}{\sigma^2} + \sum_{t=1}^{T}\frac{(1-\beta)^2}{\sigma^2} + \frac{1}{\sigma_\alpha^2}\right)\alpha^2 \right. \\ &\quad \left. -2\left(\frac{(1-\beta^2)h_1}{\sigma^2} + \frac{(1-\beta)}{\sigma^2}\sum_{t=1}^{T}(h_{t+1}-\beta h_t) + \frac{\mu_\alpha}{\sigma_\alpha^2}\right)\alpha\right] \end{aligned} \tag{4.21}$$

Thus, we have

$$\alpha|\boldsymbol{Y}^*, h_1, \boldsymbol{h}, \beta, \sigma^2, \rho \sim N(V_\alpha M_\alpha, V_\alpha), \tag{4.22}$$

where

$$M_\alpha = \frac{(1-\beta)\rho}{(1-\rho^2)\sigma}\sum_{t=1}^{T}\left[(h_{t+1}-\beta h_t)\frac{\rho}{\sigma} - \frac{y_t^*}{e^{h_t/2}}\right] + \frac{(1-\beta^2)h_1}{\sigma^2} + \frac{(1-\beta)}{\sigma^2}\sum_{t=1}^{T}[h_{t+1}-\beta h_t] + \frac{\mu_\alpha}{\sigma_\alpha^2}, \tag{4.23}$$

$$V_\alpha = \left(\frac{T(1-\beta)^2\rho^2 + T(1-\beta)^2 + (1-\beta^2)}{\sigma^2} + \frac{1}{\sigma_\alpha^2}\right)^{-1}. \tag{4.24}$$

It is clear when $\rho = 0$, the posterior distribution collapses to the standard SV model in (2.2).

4.2.5 Sampling the volatility persistence parameter β in the GMA-SV-LVG model

As previously mentioned, the method of Chib and Greenberg (1994) is used to facilitate the sampling of β in Chapter 2.2.2. Recall this method relied upon the distribution of h_1 being used as the target density. Therefore, the proposal density of interest for β in this scenario is given by

$$\begin{aligned}
&p(\beta|\boldsymbol{Y}^*, h_1, \boldsymbol{h}, \alpha, \sigma^2, \rho) \propto f(\boldsymbol{Y}^*|h_1, \boldsymbol{h}, \alpha, \beta, \sigma^2, \rho)\, f(\boldsymbol{h}|\boldsymbol{h}_{1:T}, \alpha, \beta, \sigma^2) \times p(\beta) \\
&= \frac{1}{(\sqrt{2\pi})^T(\sqrt{1-\rho^2})^T\prod_{t=1}^{T}\exp(\frac{h_t}{2})}\exp\left\{-\sum_{t=1}^{T}\frac{(\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1}-\alpha-\beta(h_t-\alpha)))^2}{2(1-\rho^2)}\right\} \times \\
&\frac{1}{(\sqrt{2\pi}\sigma)^T}\exp\left\{-\frac{\sum_{t=1}^{T}(h_{t+1}-\alpha-\beta(h_t-\alpha))^2}{2\sigma^2}\right\} \times \frac{1}{\sqrt{2\pi}\sigma_\beta}\exp\left\{-\frac{(\beta-\mu_\beta)^2}{2\sigma_\beta^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T}\frac{(y_t^* - \frac{\rho}{\sigma}(h_{t+1}-\alpha-\beta(h_t-\alpha)))^2}{e^{h_t}(1-\rho^2)} + \frac{\sum_{t=1}^{T}(h_{t+1}-\alpha-\beta(h_t-\alpha))^2}{\sigma^2} + \frac{(\beta-\mu_\beta)^2}{\sigma_\beta^2}\right]\right\},
\end{aligned}$$

where the prior for β is $\mathrm{N}(\mu_\beta, \sigma_\beta^2)$. Note that

$$\begin{aligned}
&[h_{t+1}-\alpha-\beta(h_t-\alpha)]^2 \\
= \;& (h_{t+1}-\alpha)^2 - 2\beta(h_t-\alpha)(h_{t+1}-\alpha) + \beta^2(h_t-\alpha)^2 + \text{ terms independent of } \beta \\
= \;& \beta^2(h_t-\alpha)^2 - 2\beta(h_t-\alpha)(h_{t+1}-\alpha) + \text{ terms independent of } \beta.
\end{aligned}$$

Now, we consider each term in the exponent:

Term 1:

$$\begin{aligned}
&\sum_{t=1}^{T} \frac{(\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha)))^2}{(1-\rho^2)} \\
= &\sum_{t=1}^{T} \frac{\frac{y_t^{*2}}{e^{h_t}} - 2\frac{y_t^*}{e^{h_t/2}}\frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha)) + \frac{\rho^2}{\sigma^2}(h_{t+1} - \alpha - \beta(h_t - \alpha))^2}{(1-\rho^2)} \\
= &\sum_{t=1}^{T} \frac{2\frac{y_t^*}{e^{h_t/2}}\frac{\rho}{\sigma}\beta(h_t - \alpha) + \frac{\rho^2}{\sigma^2}(\beta^2(h_t - \alpha)^2 - 2\beta(h_t - \alpha)(h_{t+1} - \alpha))}{(1-\rho^2)} + \text{ terms independent of } \beta \\
= &\frac{\rho^2}{\sigma^2(1-\rho^2)}\sum_{t=1}^{T}(h_t - \alpha)^2\beta^2 + \frac{2\rho}{\sigma(1-\rho^2)}\sum_{t=1}^{T}(h_t - \alpha)\left[\frac{y_t^{*2}}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1} - \alpha)\right]\beta \\
&+ \text{terms independent of } \beta.
\end{aligned}$$

Term 2:

$$\begin{aligned}
&\sum_{t=1}^{T} \frac{(h_{t+1} - \alpha - \beta(h_t - \alpha))^2}{\sigma^2} \\
=&\sum_{t=1}^{T} \frac{1}{\sigma^2}(h_t - \alpha)^2\beta^2 - \sum_{t=1}^{T} \frac{2}{\sigma^2}(h_t - \alpha)(h_{t+1} - \alpha)\beta + \text{terms independent of } \beta. \quad (4.25)
\end{aligned}$$

Term 3:

$$\begin{aligned}
\frac{(\beta - \mu_\beta)^2}{\sigma_\beta^2} &= \frac{\beta^2}{\sigma_\beta^2} - 2\frac{\beta\mu_\beta}{\sigma_\beta^2} + \frac{\mu_\beta^2}{\sigma_\beta^2} \\
&= \frac{1}{\sigma_\beta^2}\beta^2 - 2\frac{\mu_\beta}{\sigma_\beta^2}\beta + \text{terms independent of } \beta. \quad (4.26)
\end{aligned}$$

Hence, the terms in the exponent can be expressed as

$$\begin{aligned}
&\left(\frac{\rho^2}{\sigma^2(1-\rho^2)}\sum_{t=1}^{T}(h_t - \alpha)^2 + \sum_{t=1}^{T}\frac{1}{\sigma^2}(h_t - \alpha)^2 + \frac{1}{\sigma_\beta^2}\right)\beta^2 \\
-2&\left(\frac{\rho\sum_{t=1}^{T}(h_t - \alpha)}{\sigma(1-\rho^2)}\left[\frac{\rho}{\sigma}(h_{t+1} - \alpha) - \frac{y_t^*}{e^{h_t/2}}\right] + \sum_{t=1}^{T}\frac{(h_t - \alpha)(h_{t+1} - \alpha)}{\sigma^2} + \frac{\mu_\beta}{\sigma_\beta^2}\right)\beta. \quad (4.27)
\end{aligned}$$

Therefore, we have

$$\beta|\boldsymbol{Y}^*, h_1, \boldsymbol{h}, \alpha, \sigma^2, \rho \sim N(V_\beta M_\beta, V_\beta), \quad (4.28)$$

where

$$M_\beta = \frac{\rho \sum_{t=1}^{T}(h_t - \alpha)}{\sigma(1-\rho^2)} \left[\frac{\rho}{\sigma}(h_{t+1} - \alpha) - \frac{y_t^*}{e^{h_t/2}} \right] + \sum_{t=1}^{T} \frac{(h_t - \alpha)(h_{t+1} - \alpha)}{\sigma^2} + \frac{\mu_\beta}{\sigma_\beta^2}, \quad (4.29)$$

$$V_\beta = \left(\frac{1}{\sigma^2(1-\rho^2)} \sum_{t=1}^{T} (h_t - \alpha)^2 + \frac{1}{\sigma_\beta^2} \right)^{-1}. \quad (4.30)$$

Again, in order to implement the MH step, we first sample β^* from the proposal density in (4.28), and use the marginal density of h_1 as the target density

$$q(\beta) = \frac{\sqrt{1-\beta^2}}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(h_1 - \alpha)^2 (1-\beta^2)}{2\sigma^2} \right\}. \quad (4.31)$$

We work with $\log q(\beta)$ so that

$$\log q(\beta) = \frac{1}{2}\log(1-\beta^2) - \frac{1}{2}\log(2\pi\sigma^2) - \frac{(h_1 - \alpha)^2(1-\beta^2)}{2\sigma^2}.$$

Given the current value $\beta^{(m-1)}$ at the $(m-1)$th MCMC loop, β^* is accepted for $\beta^{(m)}$ with probability $\min\left\{\exp[q(\beta^*) - q(\beta^{(m-1)})], 1\right\}$.

4.2.6 Sampling the volatility of volatility parameter σ^2 in the GMA-SV-LVG model

The conditional posterior distribution of σ^2 is

$$\begin{aligned}
& p(\sigma^2 | \boldsymbol{Y}^*, h_1, \boldsymbol{h}, \alpha, \beta, \rho) \propto f(\boldsymbol{Y}^* | h_1, \boldsymbol{h}, \alpha, \beta, \sigma^2, \rho)\, f(h_1 | \alpha, \beta, \sigma^2)\, f(\boldsymbol{h} | \boldsymbol{h}_{1:T}, \alpha, \beta, \sigma^2) \times p(\sigma^2) \\
= \; & \frac{1}{(\sqrt{2\pi})^T (\sqrt{1-\rho^2})^T \prod_{t=1}^{T} \exp(\frac{h_t}{2})} \exp\left\{ -\sum_{t=1}^{T} \frac{[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha))]^2}{2(1-\rho^2)} \right\} \times \\
& \frac{\sqrt{1-\beta^2}}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(h_1 - \alpha)^2(1-\beta^2)}{2\sigma^2} \right\} \times \\
& \frac{1}{(\sqrt{2\pi}\sigma)^T} \exp\left\{ -\frac{\sum_{t=1}^{T}(h_{t+1} - \alpha - \beta(h_t - \alpha))^2}{2\sigma^2} \right\} \times \frac{(\frac{b}{2})^{\frac{a}{2}}}{\Gamma(\frac{a}{2})} \sigma^{-2(\frac{a}{2}+1)} \exp\left(-\frac{(\frac{b}{2})}{\sigma^2} \right),
\end{aligned}$$

where the prior $\sigma^2 \sim \Gamma(\frac{a}{2}, \frac{b}{2})$. It is clear that we cannot continue as usual to find the posterior of σ^2, because we require the exponent term to be in terms of σ^2 only, and not σ. This problem stems from including the conditional distribution of $y_t^* | h_{t+1}, h_t$, which did not exist when considering the plain SV model. Therefore, the procedure set out in Chib and Greenberg (1994) is used once more, and the conditional density

of $y_t^*|h_{t+1}, h_t$ is set as the target density for the MH scheme. Thus, the proposal density of σ^2 of interest is

$$\begin{aligned} & p(\sigma^2|h_1, \boldsymbol{h}, \alpha, \beta, \rho) \propto f(h_1|\alpha, \beta, \sigma^2)\, f(\boldsymbol{h}|\alpha, \beta, \sigma^2) \times p(\sigma^2) \\ = \quad & \frac{(1-\beta^2)^{\frac{1}{2}}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2}\right\} \times \\ & \frac{1}{(\sqrt{2\pi}\sigma)^T} \exp\left\{-\frac{\sum_{t=1}^T (h_{t+1}-\alpha-\beta(h_t-\alpha))^2}{2\sigma^2}\right\} \times \frac{(\frac{b}{2})^{\frac{a}{2}}}{\Gamma(\frac{a}{2})} \sigma^{-2(\frac{a}{2}+1)} \exp\left(-\frac{(\frac{b}{2})}{\sigma^2}\right), \end{aligned}$$

which is the conditional density of the standard SV model in (2.4) of Section 2.2.3. Hence a proposed value $(\sigma^2)^*$ can be sampled from

$$\sigma^2|\boldsymbol{h}, \alpha, \beta, \rho \sim \text{IG}\left(\frac{T+a}{2}, \frac{b+(h_1-\alpha)^2(1-\beta^2)+\sum_{t=1}^T [h_{t+1}-\alpha-\beta(h_t-\alpha)]^2}{2}\right), \tag{4.32}$$

and the target density is

$$\begin{aligned} & p(\boldsymbol{Y}^*|h_1, \boldsymbol{h}, \alpha, \beta, \sigma^2, \rho) \\ = & \frac{1}{[2\pi(1-\rho^2)]^{\frac{T}{2}} \prod_{t=1}^T \exp(\frac{h_t}{2})} \exp\left\{-\sum_{t=1}^T \frac{[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1}-\alpha-\beta(h_t-\alpha))]^2}{2(1-\rho^2)}\right\}. \end{aligned} \tag{4.33}$$

We work on the log density

$$\begin{aligned} \log q(\sigma^2) &= \log f(\boldsymbol{Y}^*|h_1, \boldsymbol{h}, \alpha, \beta, \sigma^2, \rho) \\ &= -\sum_{t=1}^T \frac{[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1}-\alpha-\beta(h_t-\alpha))]^2}{2(1-\rho^2)} + \text{terms independent of } \sigma^2, \\ &= \frac{\rho}{1-\rho^2} \sum_{t=1}^T \frac{y_t^*[h_{t+1}-\alpha-\beta(h_t-\alpha)]e^{-h_t/2}}{\sigma} \\ &\quad -\frac{\rho^2}{2(1-\rho^2)} \sum_{t=1}^T \frac{[h_{t+1}-\alpha-\beta(h_t-\alpha)]^2}{\sigma^2} + \text{terms independent of } \sigma^2. \end{aligned} \tag{4.34}$$

Once again, given the current value $(\sigma^2)^{(m-1)}$ at the $(m-1)$th MCMC loop, $(\sigma^2)^*$ is accepted for $(\sigma^2)^{(m)}$ with probability $\min\left\{\exp[q((\sigma^2)^*) - q((\sigma^2)^{(m-1)})], 1\right\}$.

4.2.7 Sampling the leverage parameter ρ in the GMA-SV-LVG model

Assuming a Gaussian prior for ρ such that $\rho \sim N(\mu_\rho, \sigma_\rho^2)$, the posterior distribution for ρ is

$$
\begin{aligned}
& f(\rho|\boldsymbol{Y}^*, h_1, \boldsymbol{h}, \alpha, \beta, \sigma^2) \propto f(\boldsymbol{Y}^*|h_1, \boldsymbol{h}, \alpha, \beta, \sigma^2, \rho) f(h_1|\alpha, \beta, \sigma^2) f(\boldsymbol{h}|\boldsymbol{h}_{1:T}, \alpha, \beta, \sigma^2) \times p(\rho) \\
&= \frac{1}{(\sqrt{2\pi})^T (\sqrt{1-\rho^2})^T \prod_{t=1}^T \exp(\frac{h_t}{2})} \exp\left\{ -\sum_{t=1}^T \frac{[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha))]^2}{2(1-\rho^2)} \right\} \times \\
& \quad \frac{(1-\beta^2)^{\frac{1}{2}}}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(h_1 - \alpha)^2 (1-\beta^2)}{2\sigma^2} \right\} \times \\
& \quad \frac{1}{(\sqrt{2\pi}\sigma)^T} \exp\left\{ -\frac{\sum_{t=1}^T (h_{t+1} - \alpha - \beta(h_t - \alpha))^2}{2\sigma^2} \right\} \times \frac{1}{\sqrt{2\pi\sigma_\rho^2}} \exp\left\{ -\frac{(\rho - \mu_\rho)^2}{2\sigma_\rho^2} \right\} \\
& \propto \frac{1}{(1-\rho^2)^{T/2}} \exp\left\{ -\sum_{t=1}^T \frac{[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha))]^2}{2(1-\rho^2)} - \frac{(\rho - \mu_\rho)^2}{2\sigma_\rho^2} \right\} \\
& \propto \frac{1}{(1-\rho^2)^{T/2}} \exp\left\{ -\sum_{t=1}^T \frac{[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha))]^2}{2(1-\rho^2)} - \frac{\rho^2}{2\sigma_\rho^2} + \frac{2\mu_\rho \rho}{2\sigma_\rho^2} \right\}. \qquad (4.35)
\end{aligned}
$$

Since ρ is bounded, a grid-based method to sample from the conditional posterior distribution is used. A Griddy-Gibbs approach (Ritter and Tanner, 1992) is used for the sampling. The algorithm essentially uses the inverse-CDF method on the empirical CDF. Formally:

1. Form a grid $\{\rho_1, \ldots, \rho_n\}$ where $\rho_i \in [-1 + c + u_\rho, 1 - c - u_\rho]$ where c is a small constant chosen to avoid boundary problems, and $u_\rho \sim U[0, 0.01]$ is a small random number. The purpose of u_ρ is to "shuffle" the random grid for each loop such that estimates of ρ can be explored between decimal places. We consider a large enough grid length of $n = 500$ points.

2. Evaluate $p(\rho_i|\boldsymbol{Y}^*, h_1, \boldsymbol{h}, \alpha, \beta, \sigma^2)$ in (4.35) at $\rho_i \in \{\rho_1, \ldots, \rho_n\}$ and the current set of parameter estimates at iteration m to obtain $W_m = \{w_1, \ldots, w_n\}$.

3. Use W_m to obtain the empirical CDF of ρ which is found by a cumulative sum.

4. Sample $u \sim U[0, 1]$ to make a draw ρ^* from the empirical inverse CDF.

4.2.8 Sampling the latent volatilities h and the long memory parameters u and d of the GMA-SV-LVG model

The sampling of the latent volatilities $\boldsymbol{h}$ as well as the long memory parameters d and u can be found in Appendix F. It is important to note the method to sample $\boldsymbol{h}$ is different from the method proposed in Appendix B for the GMA-SV model. The first method derives the posterior distribution of $\boldsymbol{h}$ via the likelihood of $\boldsymbol{Y}^* = \log \boldsymbol{G}_J^{-2}\boldsymbol{Y}^2$ which follows a log chi-square distribution expressed as ten component mixture of normals and utilizes Gibbs sampling. This new method however considers the likelihood of $\boldsymbol{Y}^* = \boldsymbol{G}_J^{-1}\boldsymbol{Y}$, the idea of MAP and utilizes an acceptance/rejection MH scheme. Our experience dictates that acceptance/rejection MH schemes avoid the problems of samplers not being able to explore the parameter space fully and can subsequently become "stuck".

4.2.9 Simulation studies of the SV-LVG model

Although theoretically pleasing, MCMC techniques are notorious for being practically burdensome. The SV-LVG model is especially important, because this basic model serves as a cornerstone for the following chapters. Since MCMC sampling can be a cumbersome task, it is advantageous to ensure the SV-LVG model alone (with no long memory effects) is being estimated correctly. This ensures that any future potential issues that arise from other parameters of some extended models can be easily detected. To achieve this, we conduct several simulation studies. Synthetic SV-LVG data is generated from the model and is simulated with $T = 2,000$ and the true parameters equal to $\mu = 0, \alpha = 0, \beta = 0.975, \sigma^2 = 0.01$. The SV-LVG model is then estimated using this data, and this process is repeated $1,000$ times with the averages of parameter estimates reported in Table 4.1; where $\widehat{AR}_\theta$% represents the percentage points of the acceptance rates for parameter θ. The total number of loops is $20,000$, and we discard the first $10,000$ as the burn-in. The same priors are used as in Chapter 3, with the addition of $\rho \sim \text{N}(-0.5, 0.05)$.

ρ	$\hat{\rho}$	RMSE($\hat{\rho}$)	$\hat{\mu}$	RMSE($\hat{\mu}$)	$\hat{\alpha}$	RMSE($\hat{\alpha}$)	$\hat{\beta}$	RMSE($\hat{\beta}$)	$\hat{\sigma}^2$	RMSE($\hat{\sigma}^2$)	$\widehat{AR}_\rho\%$	$\widehat{AR}_h\%$	$\widehat{AR}_\beta\%$	$\widehat{AR}_{\sigma^2}\%$
0.0	0.001	0.001	0.001	0.016	0.020	0.020	0.975	0.010	0.031	0.011	3.225	61.572	93.128	95.965
		(0.105)		(0.024)		(0.243)		(0.008)		(0.007)				
-0.1	-0.094	0.006	-0.002	0.031	-0.004	0.004	0.976	0.009	0.031	0.011	3.138	61.873	93.101	95.161
		(0.103)		(0.024)		(0.250)		(0.008)		(0.007)				
-0.5	-0.466	0.034	-0.003	0.000	-0.005	0.005	0.977	0.008	0.030	0.010	2.296	67.922	93.850	83.153
		(0.086)		(0.024)		(0.244)		(0.007)		(0.007)				
-0.9	-0.874	0.026	-0.005	0.017	-0.003	0.003	0.982	0.003	0.026	0.006	0.585	85.452	96.316	52.252
		(0.020)		(0.022)		(0.234)		(0.004)		(0.004)				

TABLE 4.1: Parameter estimates for SV-LVG model in an initial simulation study.

From this initial simulation study in Table 4.1, it is clear that ρ is estimated reasonably well using the MH sampler, however the acceptance rate of $\boldsymbol{h}$ is too high, and the acceptance rate of ρ is too low.

ρ	$\hat{\rho}$	RMSE($\hat{\rho}$)	$\hat{\mu}$	RMSE($\hat{\mu}$)	$\hat{\alpha}$	RMSE($\hat{\alpha}$)	$\hat{\beta}$	RMSE($\hat{\beta}$)	$\hat{\sigma}^2$	RMSE($\widehat{\sigma^2}$)	$\widehat{AR}_\rho\%$	$\widehat{AR}_h\%$	$\widehat{AR}_\beta\%$	$\widehat{AR}_{\sigma^2}\%$
0.0	-0.009	0.009	-0.000	0.019	-0.008	0.008	0.975	0.010	0.031	0.011	80.618	61.422	93.145	96.944
		(0.096)		(0.024)		(0.243)		(0.008)		(0.007)				
-0.1	-0.090	0.010	-0.001	0.006	-0.020	0.020	0.975	0.010	0.031	0.011	80.622	61.902	93.206	96.318
		(0.096)		(0.024)		(0.247)		(0.008)		(0.007)				
-0.5	-0.465	0.035	-0.004	0.001	0.004	0.004	0.977	0.008	0.030	0.010	80.905	68.289	93.741	84.027
		(0.085)		(0.024)		(0.243)		(0.007)		(0.006)				
-0.9	-0.874	0.026	-0.005	0.001	-0.006	0.006	0.981	0.004	0.026	0.006	81.311	85.938	96.474	51.594
		(0.041)		(0.022)		(0.232)		(0.004)		(0.004)				

TABLE 4.2: Parameter estimates for SV-LVG model after tuning the posterior precision of ρ using Griddy Gibbs

Following the results of Table 4.1, the next step is to improve the acceptance rates of ρ and $\boldsymbol{h}$. Therefore, we follow the tuning procedure in the exact fashion as discussed in Section 4.2.7. However, Table 4.2 shows the acceptance rate of ρ is now too high. As such, it seems the next natural step is to use an alternative method to sample the posterior of ρ.

ρ	$\hat{\rho}$	RMSE($\hat{\rho}$)	$\hat{\mu}$	RMSE($\hat{\mu}$)	$\hat{\alpha}$	RMSE($\hat{\alpha}$)	$\hat{\beta}$	RMSE($\hat{\beta}$)	$\hat{\sigma}^2$	RMSE($\hat{\sigma}^2$)	$\widehat{AR}_\rho\%$	$\widehat{AR}_h\%$	$\widehat{AR}_\beta\%$	$\widehat{AR}_{\sigma^2}\%$
0.0	0.005	0.005	-0.002	0.067	-0.018	0.018	0.976	0.009	0.031	0.011	30.126	61.502	92.899	95.916
		(0.110)		(0.024)		(0.249)		(0.008)		(0.007)				
-0.1	-0.093	0.007	0.000	0.024	0.002	0.002	0.975	0.010	0.031	0.011	30.475	61.700	93.139	95.215
		(0.109)		(0.024)		(0.244)		(0.008)		(0.007)				
-0.5	-0.461	0.039	-0.001	0.024	-0.014	0.014	0.978	0.007	0.030	0.010	32.358	67.947	93.801	83.330
		(0.096)		(0.024)		(0.246)		(0.007)		(0.006)				
-0.9	-0.868	0.032	-0.005	0.032	-0.004	0.004	0.981	0.004	0.026	0.006	31.331	85.483	96.455	52.408
		(0.046)		(0.022)		(0.231)		(0.004)		(0.004)				

TABLE 4.3: Parameter estimates for SV-LVG model after changing the sampler of ρ from Griddy Gibbs to the MAP sampler.

In Table 4.3, the method of estimating ρ is changed from the Griddy Gibbs sampler to the MAP sampler in Section 1.2.4. As evidenced, the acceptance rate of ρ is now at an

appropriate level. Therefore, we instead adopt the MAP sampler with tuning in order to sample ρ instead of using the Griddy Gibbs sampler. The main issue with Table 4.3 however is the uncomfortably large acceptance rate of σ^2. This is a particularly worrisome result since σ^2 is close to a boundary.

ρ	$\hat{\rho}$	RMSE($\hat{\rho}$)	$\hat{\mu}$	RMSE($\hat{\mu}$)	$\hat{\alpha}$	RMSE($\hat{\alpha}$)	$\hat{\beta}$	RMSE($\hat{\beta}$)	$\hat{\sigma^2}$	RMSE($\hat{\sigma^2}$)	$\widehat{AR}_\rho\%$	$\widehat{AR}_h\%$	$\widehat{AR}_\beta\%$	$\widehat{AR}_{\sigma^2}\%$
0.0	0.003	0.003	0.001	0.012	-0.011	0.011	0.976	0.009	0.031	0.011	29.585	62.167	93.119	30.717
		(0.110)		(0.024)		(0.249)		(0.008)		(0.007)				
-0.1	-0.092	0.008	-0.002	0.027	-0.018	0.018	0.976	0.009	0.031	0.011	30.410	62.057	92.901	30.504
		(0.109)		(0.024)		(0.253)		(0.008)		(0.007)				
-0.5	-0.472	0.028	-0.004	0.011	-0.003	0.003	0.978	0.007	0.029	0.009	32.469	68.940	93.774	31.655
		(0.096)		(0.024)		(0.244)		(0.007)		(0.006)				
-0.9	-0.870	0.030	-0.004	0.010	-0.009	0.009	0.982	0.003	0.026	0.006	31.368	85.696	96.463	31.745
		(0.046)		(0.022)		(0.233)		(0.004)		(0.004)				

TABLE 4.4: Parameter estimates for SV-LVG model using MAP for ρ and independent Gaussian proposal for σ^2.

The next iteration to optimise the MCMC algorithm is to change the sampler of σ^2 from a Random Walk Metropolis Hastings algorithm in Section 1.2.4 (Case 2) to an adaptive independent Metropolis Hastings algorithm (Case 1) with a Gaussian proposal. The results of this change are depicted in Table 4.4, and the acceptance rate of σ^2 is now satisfactory.

ρ	$\hat{\rho}$	RMSE($\hat{\rho}$)	$\hat{\mu}$	RMSE($\hat{\mu}$)	$\hat{\alpha}$	RMSE($\hat{\alpha}$)	$\hat{\beta}$	RMSE($\hat{\beta}$)	$\hat{\sigma^2}$	RMSE($\hat{\sigma^2}$)	$\widehat{AR}_\rho\%$	$\widehat{AR}_h\%$	$\widehat{AR}_\beta\%$	$\widehat{AR}_{\sigma^2}\%$
0.0	0.000	0.000	0.013	0.013	-0.144	0.144	0.969	0.016	0.038	0.018	30.380	73.900	95.030	32.610
		(0.010)		(0.023)		(0.185)		(0.009)		(0.008)				
-0.1	-0.101	0.001	-0.029	0.029	-0.236	0.236	0.967	0.018	0.028	0.008	28.460	84.680	95.370	39.890
		(0.010)		(0.022)		(0.150)		(0.010)		(0.006)				
-0.5	-0.500	0.000	-0.004	0.004	0.151	0.151	0.985	0.000	0.025	0.005	26.420	88.180	94.010	38.040
		(0.010)		(0.025)		(0.309)		(0.005)		(0.005)				
-0.9	-0.886	0.014	-0.007	0.007	-0.139	0.139	0.978	0.007	0.033	0.013	29.310	82.540	97.410	34.530
		(0.039)		(0.021)		(0.203)		(0.005)		(0.005)				

TABLE 4.5: Parameter estimates for SV-LVG model by reducing error tolerance in the MATLAB objective function.

The next step is to obtain even more accurate results for ρ in order to obtain the best model. Although the error is small, it should be noted that using a synthetic time series of $T = 2,000$ with $1,000$ repeats should yield very accurate results. In a real data setting, especially with very volatile data, this error can be significantly inflated. As such, in order to obtain more accurate results for ρ, we modify the optimisation objective function (i.e. the log likelihood) constraints in MATLAB. By default, MATLAB will stop iterating when there is less than a 10^{-4} difference around the solution. It is possible however to modify this to be even stricter to 10^{-6}, and the results of this are outputted in Table 4.5. Clearly, the results of ρ are now superior

and close to the theoretical values. However, the acceptance rates of $\boldsymbol{h}$ are now too high. The potential reason for this could be due to the fact the conditional variance of $\boldsymbol{h}$ has reduced, and therefore proposed values of $\boldsymbol{h}$ are being sampled too close to the currently proposed value. As such, this experiment is repeated again but with a slightly less conservative tolerance.

ρ	$\hat{\rho}$	RMSE($\hat{\rho}$)	$\hat{u}$	RMSE($\hat{u}$)	$\hat{d}$	RMSE($\hat{d}$)	$\hat{\mu}$	RMSE($\hat{\mu}$)	$\hat{\alpha}$	RMSE($\hat{\alpha}$)	$\hat{\beta}$	RMSE($\hat{\beta}$)	$\hat{\sigma}^2$	RMSE($\hat{\sigma}^2$)	$\widehat{AR}_\rho\%$	$\widehat{AR}_u\%$	$\widehat{AR}_d\%$	$\widehat{AR}_h\%$	$\widehat{AR}_\beta\%$	$\widehat{AR}_{\sigma^2}\%$
0.0	0.000	0.000	0.500	0.020	0.300	0.015	-0.000	0.000	-0.004	0.004	0.976	0.009	0.031	0.011	24.694	44.767	28.206	28.856	92.691	30.659
		(0.045)		(0.008)		(0.017)		(0.027)		(0.252)		(0.008)		(0.007)						
-0.1	-0.099	0.001	0.500	0.010	0.299	0.001	-0.002	0.040	-0.014	0.014	0.976	0.009	0.031	0.011	24.779	45.124	29.037	29.499	93.042	30.909
		(0.045)		(0.008)		(0.017)		(0.027)		(0.247)		(0.008)		(0.007)						
-0.5	-0.496	0.004	0.500	0.000	0.300	0.012	-0.007	0.021	0.014	0.014	0.978	0.007	0.029	0.009	29.026	50.730	29.505	30.337	93.683	31.721
		(0.043)		(0.008)		(0.018)		(0.026)		(0.249)		(0.007)		(0.006)						
-0.9	-0.892	0.008	0.500	0.004	0.301	0.013	-0.026	0.002	0.165	0.165	0.981	0.004	0.025	0.005	32.660	66.635	33.757	31.562	96.722	31.136
		(0.030)		(0.007)		(0.017)		(0.026)		(0.233)		(0.004)		(0.004)						

TABLE 4.6: Parameter estimates for SV-LVG model by setting tolerance to 10^{-5}.

Finally, the tolerance is increased slightly to 10^{-5} and the Gegenbauer long memory parameters are also included. It is clear from Table 4.6 the results are satisfactory. As such, we accept this framework and utilize this as the procedure for all future models involving leverage.

Next we consider extending the GMA-SV-LVG model to incorporate heavy tails (GMA-SV-LVG-HC). Again, we first consider the simpler GMA-SV-HC model, which excludes the leverage component.

4.3 Bayesian inference for GMA-SV-HC model

Many SV models are assumed to follow a normal distribution. However, increasing evidence in real applications has shown the normal distribution is not an appropriate choice, as many returns or volatilities show leptokurtic shapes. The ability of the Student's t-distribution to provide flexible tails is an alternative choice to the normal distribution. Cryptocurrencies have wild volatility characteristics, and the consideration of heavy tailed distributions is a progressive step to measure their unique dynamics. Although a theoretically pleasing idea, the practicalities of implementing the Student's t-distribution is not a straight forward exercise; as such, we rely upon scale mixtures.

The scale mixture representation has received substantial attention in Bayesian robustness (Box and Tiao, 2011). In this section, the scale mixtures representation

of the Student's t-distribution is discussed in detail. The two most common choices to express the Student's t-distribution is either through a scale mixture of normals (SMN) or a scale mixture of uniforms (SMU); see Choy and Chan (2008) for examples.

4.3.1 SMU representation

Walker and Gutiérrez-Pena (1999) proposed a new class of scale mixtures, known as the SMU distribution. We choose the SMN representation for the Student's t-distribution but we also provide a brief overview on how the SMU works. Let X be a continuous random variable with location μ and scale σ. The pdf of X is said to have a SMU representation if it can be expressed as

$$f_X(x|\mu,\sigma) = \int_0^\infty \mathrm{U}(x|\mu-\kappa(w)\sigma, \mu+\kappa(u)\sigma)\pi(w)dw,$$

where $\mathrm{U}(x|a,b)$ is the uniform density function with support $[a,b]$, $\kappa(\cdot)$ is a positive function, and $\pi(\cdot)$ is a density function on the positive real domain.

The EP distribution with mean μ and variance σ^2 has the density function

$$f_X(x|\mu,\sigma^2,\beta) = \frac{c_1}{\sigma}\exp\left(-\left|\frac{c_0^{1/2}(x-\mu)}{\sigma}\right|^{2/\beta}\right),$$

where

$$\beta\in(0,2], \quad c_0 = \frac{\Gamma(3\beta/2)}{\Gamma(\beta/2)}, \quad c_1 = \frac{c_0^{1/2}}{\beta\Gamma(\beta/2)}.$$

Two special cases here are the normal distribution $(\beta=1)$ and the double-exponential distribution $(\beta=2)$. The EP density has the SMU representation

$$f_X(x|\mu,\sigma^2,\beta) = \int_0^\infty \mathrm{U}\left(x|\mu-\frac{\sigma}{\sqrt{2c_0}}w^{\beta/2}, \mu+\frac{\sigma}{\sqrt{2c_0}}w^{\beta/2}\right)\times \mathrm{Ga}\left(w|1+\frac{\beta}{2}, 2^{-1/\beta}\right)dw,$$

where $\mathrm{Ga}(\cdot,\cdot)$ represent the Gamma density function. This can be expressed in hierarchical form as

$$\begin{aligned} X|\mu,\sigma^2,\beta,w &\sim \mathrm{U}\left(\mu-\frac{\sigma}{\sqrt{2c_0}}w^{\beta/2}, \mu+\frac{\sigma}{\sqrt{2c_0}}w^{\beta/2}\right), \\ w|\beta &\sim \mathrm{Ga}\left(1+\frac{\beta}{2}, 2^{-1/\beta}\right). \end{aligned}$$

Therefore, we can rewrite the Normal distribution in terms of a SMU distribution as:

$$\begin{aligned} X|\mu,\sigma^2,w &\sim \text{U}\left(\mu-\sigma\sqrt{w},\mu+\sigma\sqrt{w}\right), \\ w &\sim \text{Ga}\left(\frac{3}{2},\frac{1}{2}\right). \end{aligned}$$

And also the Student's t-distribution with ν degrees of freedom as

$$\begin{aligned} X|\mu,\sigma^2,\nu,w &\sim \text{U}\left(\mu-\sigma\sqrt{\xi w},\mu+\sigma\sqrt{\xi w}\right), \\ w &\sim \text{Ga}\left(\frac{3}{2},\frac{1}{2}\right), \\ \xi|\nu &\sim IG(\nu/2,\nu/2). \end{aligned}$$

As the density of a uniform is constant, the use of the SMU can facilitate standard truncated posterior distributions if the data distribution is non-standard with respect to certain parameters of interest. Examples of popular SMU distribution applications include:

1. **The Exponential Power distribution**
   If $X \sim \text{U}\left(\mu-\frac{\sigma}{\sqrt{2c_0}}w^{\beta/2},\mu+\frac{\sigma}{\sqrt{2c_0}}w^{\beta/2}\right)$ where $w|\beta \sim \text{Ga}(1+\frac{\beta}{2},2^{-1/\beta})$, then X comes from the Exponential Power distribution with parameters μ,σ^2,β.

2. **The uniform power distribution**
   If $X \sim \text{U}\left(\mu-\sigma w^{\beta/2},\mu+\sigma w^{\beta/2}\right)$ where $w \sim \text{Ga}(\frac{3}{2},\frac{1}{2})$, then X comes from the uniform power distribution with parameters μ,σ^2,β.

3. **The generalised t-distribution**
   If $X \sim \text{U}\left(\mu-q^{\frac{1}{p}}s^{-\frac{1}{2}}w^{\frac{1}{p}}\sigma,\mu+q^{\frac{1}{p}}s^{-\frac{1}{2}}w^{\frac{1}{p}}\sigma\right)$ where $w \sim \text{Ga}(1+\frac{1}{p},1)$, and $s \sim \text{GG}(q,1,\frac{p}{2})$, then X comes from the GT distribution with parameters μ, σ^2 and two shape parameters p and q.

4.3.2 SMN representation

We consider the SMN for the Student's t-distribution in our proposed GMA-SV-LVG-HC model as it can make use of some normal conjugates (Sections 2.2 and 4.2). Examples of popular SMN distribution applications include:

1. **The Student's t-distribution**

   If $X \sim \text{N}(\mu, \xi\sigma^2)$ where $\xi \sim \text{IG}(\frac{\nu}{2}, \frac{\nu}{2})$ and $\text{IG}(\cdot,\cdot)$ represents the Inverse Gamma distribution, then $X \sim t_\nu(\mu, \sigma)$.

2. **The Pearson Type VII family**

   If $X \sim \text{N}(\mu, \xi\sigma^2)$ where $\xi \sim \text{IG}(\frac{\nu}{2}, \frac{\delta}{2})$, then X follows the Pearson Type VII distribution with parameters μ, σ^2, δ and ν.

3. **The Variance Gamma distribution**

   If $X \sim \text{N}(\mu, \xi\sigma^2)$ where $\xi \sim \text{Ga}(\frac{\nu}{2}, \frac{\nu}{2})$, then X follows the Variance Gamma distribution with parameters μ, σ^2 and ν.

4.3.3 Gegenbauer long memory SV model with Student's t-distribution for returns (GMA-SV-HC)

This model drops the leverage effect but considers heavy tailed Student's t-innovations in the observation equation. We introduce a mixing variable ξ_t to $\exp(h_t)$ as

$$\begin{aligned} y_t|h_t &= \sum_{j=0}^{J_t^*} \lambda_j \xi_{t-j}^{\frac{1}{2}} \exp(h_{t-j}/2)\epsilon_{t-j}^*, \quad \epsilon_{t-j}^* \sim \text{N}(0,1), \qquad (4.36) \\ h_{t+1}|h_t &= \alpha + \beta(h_t - \alpha) + \sigma\eta_t^*, \quad \eta_t \sim \text{N}(0,1), \\ h_1 &= \alpha + \sigma/\sqrt{1-\beta^2}\eta_1^*, \quad \eta_1 \sim \text{N}(0,1), \qquad (4.37) \end{aligned}$$

where the mixing variable $\xi_t|\nu \sim \text{IG}(\frac{\nu}{2}, \frac{\nu}{2})$. Note that this is model 3 in Appendix E. The model can be written in matrix form as in (4.15) where the covariance matrix can be expressed as $\boldsymbol{\Gamma} = \boldsymbol{G}_{J^*}\boldsymbol{V}\boldsymbol{G}'_{J^*}$ where $\boldsymbol{V} = \text{diag}(\boldsymbol{W} \circ \boldsymbol{W} \circ \boldsymbol{\xi})$, $\boldsymbol{W} = (e^{h_1/2}, \ldots, e^{h_T/2})$ and $\boldsymbol{\xi} = (\xi_2, \ldots, \xi_{T+1})$. Again, we consider the transformation $\boldsymbol{Y}^* = \boldsymbol{G}_J^{-1}\boldsymbol{Y}$ as in (4.16) where $\boldsymbol{Y}^* = (y_1^*, \ldots, y_T^*)$ is independent of the Gegenbauer long memory parameters u and d. Note that (4.36) corresponds to

$$y_t^*|h_t = \xi_t^{\frac{1}{2}} \exp(h_t/2)\epsilon_t, \quad \epsilon_t \sim \text{N}(0,1).$$

The next two subsections discuss the sampling of the shape parameter ν and mixing variable ξ_t for the Student's t-innovations of returns.

4.3.4 Sampling the mixing variable ξ_t in the GMA-SV-HC model

Each ξ_t is independent, and can be sampled separately. The posterior distribution of ξ_t is

$$\begin{aligned}
p(\xi_t|y_t^*, h_t, \nu) &\propto f(y_t^*|h_t, \xi_t) \times p(\xi_t|\nu) \\
&= \frac{1}{\sqrt{2\pi\xi_t e^{h_t}}} \exp\left(-\frac{y_t^{*2}}{2\xi_t e^{h_t}}\right) \times \frac{(\frac{\nu}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \xi_t^{-(\frac{\nu}{2}+1)} \exp\left(-\frac{\frac{\nu}{2}}{\xi_t}\right) \\
&\propto \xi_t^{-\frac{1}{2}} \exp\left(-\frac{y_t^{*2}}{2\xi_t e^{h_t}}\right) \times \xi_t^{-(\frac{\nu}{2}+1)} \exp\left(-\frac{\frac{\nu}{2}}{\xi_t}\right) \\
&= \xi_t^{-(\frac{1}{2}+\frac{\nu}{2}+1)} \exp\left[-\left(\frac{y_t^{*2}}{2\xi_t e^{h_t}} + \frac{\frac{\nu}{2}}{\xi_t}\right)\right] \\
&= \xi_t^{-(\frac{1}{2}+\frac{\nu}{2}+1)} \exp\left[-\left(\frac{\frac{1}{2}y_t^{*2}e^{-h_t} + \frac{\nu}{2}}{\xi_t}\right)\right] \\
&\sim \text{IG}\left(\frac{\nu+1}{2}, \frac{\nu + y_t^{*2}e^{-h_t}}{2}\right).
\end{aligned}$$

4.3.5 Sampling the shape parameter ν in the GMA-SV-HC model

In order to sample ν, an independence-chain MH algorithm is implemented. Assuming the prior $\nu \sim \text{U}[0, \tilde{\nu}]$, the posterior distribution of ν is

$$\begin{aligned}
p(\nu|\boldsymbol{\xi}) &\propto p(\nu) \times p(\boldsymbol{\xi}|\nu) \\
&= \prod_{t=1}^{T} \frac{(\frac{\nu}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \xi_t^{-\frac{\nu}{2}-1} \exp\left(\frac{-\nu/2}{\xi_t}\right) I_{0,\tilde{\nu}} \\
&= \frac{(\frac{\nu}{2})^{\frac{T\nu}{2}}}{\Gamma(\frac{\nu}{2})^T} \left(\prod_{t=1}^{T} \xi_t^{-\frac{\nu}{2}-1}\right) \exp\left(\sum_{t=1}^{T} \frac{-\nu/2}{\xi_t}\right) I_{0,\tilde{\nu}},
\end{aligned}$$

where $I_{0,\bar{\nu}}$ is an indicator function such that it is equal to one if $0 < \nu < \tilde{\nu}$ and zero otherwise and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_T)$. Then we have

$$\log p(\nu|\xi_t) = \frac{T\nu}{2} \log\left(\frac{\nu}{2}\right) + T \log\Gamma\left(\frac{2}{\nu}\right) + \left(\frac{\nu}{2}+1\right) \sum_{t=1}^{T} \log \xi_t^{-1} - \nu/2 \sum_{t=1}^{T} \xi_t^{-1},$$

where

$$\begin{aligned}
\frac{d \log p(\nu|\boldsymbol{\xi})}{d\nu} &= \frac{T}{2} \log\left(\frac{\nu}{2}\right) + \frac{T}{2} - \frac{T}{2}\psi\left(\frac{\nu}{2}\right) - \frac{1}{2}\sum_{t=1}^{T} \log \xi_t - \frac{1}{2}\sum_{t=1}^{T} \xi_t^{-1}, \\
\frac{d^2 \log p(\nu|\boldsymbol{\xi})}{d\nu^2} &= \frac{T}{2\nu} - \frac{T}{4}\psi_1\left(\frac{\nu}{2}\right),
\end{aligned}$$

where ψ and ψ_1 are the digamma and trigamma functions respectively. We maximise the density in order to find the mode $\bar{\nu}$ using the Newton-Raphson method and also evaluate the inverse of the Fishers information evaluated at the mode denoted as $V_{\bar{\nu}}$. We then sample from $\nu^* \sim N(\bar{\nu}, V_{\bar{\nu}})$, and accept or reject ν^* using the MH algorithm with a normal proposal.

The sampling procedures of α, β and σ^2 are given in Section 2.2, however instead use y_t^*. Also, the sampling of the latent volatilities h_1 and $\boldsymbol{h}$ as well as the long memory parameters d and u can be found in Appendix F.

4.3.6 Simulation study of the SV-HC model

Once again, we conclude this section by testing our proposed estimators at various values of ν utilizing all of the MCMC samplers from previous sections. The length of observations is $T = 2,000$. As before, the experiment is repeated $1,000$ times and the average values are recorded in Table 4.7. Evidently, as ν becomes larger, the estimation error increases. This is an anticipated result due to the fact that differences in percentiles between Student's t-distributions becomes smaller as ν increases. After testing the accuracy of estimating the heavy tail component in the SV-MC model, we consider in the next section the GMA-SV-LVG-HC model. This model has both leverage and heavy observational tails by combining the GMA-SV-LVG model in Section 4.2 and the GMA-SV-HC model in Section 4.3.

ν	$\hat{\nu}$	RMSE($\hat{\nu}_1$)	$\hat{\mu}$	RMSE($\hat{\mu}$)	$\hat{\alpha}$	RMSE($\hat{\alpha}$)	$\hat{\beta}$	RMSE($\hat{\beta}$)	$\hat{\sigma}^2$	RMSE($\hat{\sigma}^2$)	$\widehat{AR}_h\%$	$\widehat{AR}_\beta\%$	$\widehat{AR}_\nu\%$	Time (min.)
3.0	3.092	0.092	0.000	0.000	0.057	0.057	0.973	0.012	0.026	0.006	40.566	46.653	25.165	243
		(0.286)		(0.000)		(0.263)		(0.024)		(0.007)				
4.0	4.235	0.235	0.000	0.000	0.026	0.026	0.978	0.007	0.027	0.007	40.309	46.801	31.057	147
		(0.401)		(0.000)		(0.225)		(0.007)		(0.007)				
5.0	5.229	0.229	0.000	0.000	-0.001	0.001	0.978	0.007	0.027	0.007	40.620	46.881	31.974	109
		(0.579)		(0.000)		(0.216)		(0.007)		(0.007)				
6.0	6.859	0.859	0.000	0.000	0.036	0.036	0.976	0.009	0.027	0.007	40.752	46.607	30.126	73
		(0.993)		(0.000)		(0.225)		(0.012)		(0.007)				
7.0	7.961	0.961	0.000	0.000	0.041	0.041	0.976	0.009	0.026	0.006	41.129	46.722	29.546	51
		(1.276)		(0.000)		(0.235)		(0.014)		(0.007)				
8.0	10.374	2.374	0.000	0.000	0.044	0.044	0.978	0.007	0.027	0.007	40.361	47.133	33.092	40
		(2.138)		(0.000)		(0.219)		(0.007)		(0.007)				
9.0	12.465	3.465	0.000	0.000	0.046	0.046	0.978	0.007	0.028	0.008	39.795	46.931	32.802	33
		(2.751)		(0.000)		(0.225)		(0.007)		(0.007)				
10.0	14.580	4.580	0.000	0.000	0.066	0.066	0.979	0.006	0.027	0.007	40.380	46.620	34.625	30
		(3.432)		(0.000)		(0.234)		(0.007)		(0.007)				
15.0	20.147	5.147	0.000	0.000	0.017	0.017	0.980	0.005	0.026	0.006	41.236	46.896	34.958	22
		(4.169)		(0.000)		(0.228)		(0.006)		(0.006)				
20.0	22.416	2.416	0.000	0.000	0.024	0.024	0.978	0.007	0.025	0.005	42.553	46.827	34.948	18
		(4.195)		(0.000)		(0.212)		(0.007)		(0.006)				

TABLE 4.7: Parameter estimates of the SV-HC model for various levels of ν

4.4 Bayesian inference for GMA-SV-LVG-HC model

4.4.1 Model specification

Finally, the GMA-SV-LVG-HC model is a culmination of the previously discussed effects which have been iteratively built up. This model contains features that are powerful to deal with extremely volatile time series', such as those of Cryptocurrencies. The GMA-SV-LVG-HC model is derived by first conditioning ϵ_t^*, then introducing the mixing variable ξ_t which is combined with $\exp(h_t)$ and σ^2. To begin with, the GMA-SV-LVG-HC model is given by

$$\begin{aligned} y_t &= \sum_{j=0}^{J_t^*} \lambda_j \varepsilon_{t-j}^* \\ h_{t+1} &= \alpha + \beta(h_t - \alpha) + \eta_t^* \\ h_1 &= \alpha + \frac{\sigma}{\sqrt{1-\beta^2}}\eta_1 \\ \begin{pmatrix} \varepsilon_t^* \\ \eta_t^* \end{pmatrix} &\sim t_\nu \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \exp(h_t) & \rho\sigma\exp(h_t/2) \\ \rho\sigma\exp(h_t/2) & \sigma^2 \end{pmatrix} \right). \end{aligned} \tag{4.38}$$

The equivalent SMN representation of (4.38) is

$$\begin{pmatrix} \varepsilon_t^* \\ \eta_t^* \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \xi_t \begin{pmatrix} \exp(h_t) & \rho\sigma\exp(h_t/2) \\ \rho\sigma\exp(h_t/2) & \sigma^2 \end{pmatrix} \right), \text{ where } \xi_t \sim \mathrm{IG}\left(\frac{\nu}{2}, \frac{\nu}{2}\right).$$

Therefore, this model can be written as

$$\begin{aligned} \begin{pmatrix} y_t \\ h_{t+1} \end{pmatrix} &\sim \mathrm{N}\left(\begin{pmatrix} \sum_{j=0}^{J_t^*} \lambda_j \varepsilon_{t-j}^* \\ \alpha + \beta(h_t - \alpha) \end{pmatrix}, \xi_t \begin{pmatrix} \exp(h_t) & \rho\sigma\exp(h_t/2) \\ \rho\sigma\exp(h_t/2) & \sigma^2 \end{pmatrix} \right), \\ h_1 &\sim \mathrm{N}\left(\alpha, \frac{\sigma^2}{1-\beta^2}\right), \\ \xi_t &\sim \mathrm{IG}\left(\frac{\nu}{2}, \frac{\nu}{2}\right). \end{aligned}$$

Again, by expressing the bivariate model as a product of a marginal volatility and conditional observation component, and using the conditional normal distribution formula

$$X_1 | x_2 = x_2 \sim \mathrm{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), (1-\rho^2)\sigma_1^2\right),$$

we have

$$\begin{aligned}\varepsilon_t^*|\eta_t^* &= \mathrm{N}\left(\rho\frac{\exp(h_t/2)}{\sigma}(\eta_t^*-0),(1-\rho^2)\xi_t\exp(h_t)\right)\\ &= \mathrm{N}\left(\rho\frac{\exp(h_t/2)}{\sigma}[h_{t+1}-\alpha-\beta(h_t-\alpha)],(1-\rho^2)\xi_t\exp(h_t)\right). \quad (4.39)\end{aligned}$$

Hence

$$\begin{aligned}y_t|h_{t+1},h_t &= \sum_{j=0}^{J_t^*}\lambda_j\varepsilon_{t-j}^*|\eta_{t+1-j}^*\\ &= \sum_{j=0}^{\infty}\lambda_j\left[\rho\frac{\exp(h_t/2)}{\sigma}[h_{t+1}-\alpha-\beta(h_t-\alpha)]+\sqrt{(1-\rho^2)}\xi_t^{\frac{1}{2}}\exp(h_t/2)\varepsilon_t\right],\end{aligned}$$

where $\varepsilon_t\sim\mathrm{N}(0,1)$ and $\xi_t\sim\mathrm{IG}\left(\frac{\nu}{2},\frac{\nu}{2}\right)$. The model can also be written as

$$\begin{aligned}y_t|h_{t+1},h_t &\sim \mathrm{N}\left(\sum_{j=0}^{J_t^*}\lambda_j\exp(h_{t-j}/2)\frac{\rho}{\sigma}[h_{t+1-j}-\alpha-\beta(h_{t-j}-\alpha)],(1-\rho^2)\sum_{j=0}^{J_t^*}\lambda_j^2\xi_{t-j}\exp(h_{t-j})\right),\\ h_{t+1}|h_t &\sim \mathrm{N}(\alpha+\beta(h_t-\alpha),\xi_t\sigma^2)=\mathrm{N}(\alpha(1-\beta)+\beta h_t,\xi_t\sigma^2),\\ h_1 &\sim \mathrm{N}\left(\alpha,\xi_1\frac{\sigma^2}{1-\beta^2}\right),\\ \xi_{t+1} &\sim \mathrm{IG}\left(\frac{\nu}{2},\frac{\nu}{2}\right),\end{aligned}$$

in which the first line of the model agrees with (4.11) and (4.12) when the mixing variable ξ_t is attached to $\exp(h_t)$ and σ^2. Using the transformation in (4.16), we have

$$\begin{aligned}y_t^*|h_{t+1},h_t &\sim \mathrm{N}\left(\exp(h_t/2)\frac{\rho}{\sigma}[h_{t+1}-\alpha-\beta(h_t-\alpha)],(1-\rho^2)\xi_t\exp(h_t)\right),\\ h_{t+1}|h_t &= \alpha+\beta(h_t-\alpha)+\xi_t\sigma\eta_t,\quad \eta_t\sim\mathrm{N}(0,1),\end{aligned}$$

where $\xi_t|\nu\ \sim\mathrm{IG}(\frac{\nu}{2},\frac{\nu}{2})$.

4.4.2 Comparison with Choy and Chan (2000) and Wang et al. (2012)

It is worthwhile to discuss the applications for two pre-existing models without long memory, that is, the SV-LVG-HC model. Choy and Chan (2000) used the same approach to first represent the bivariate Student's t-distribution in the standard SV model in (4.1) to (4.3) as a scale mixtures of bivariate normals and factorized the

model using (4.7) with a common mixing variable as

$$\begin{aligned} y_t|h_t &\sim \mathrm{N}\left(\frac{\rho}{\tau}\exp(h_t/2)(h_{t+1}-\mu-\phi(h_t-\mu)), \xi_t\exp(h_t)(1-\rho^2)\right), \\ h_{t+1}|h_t &\sim \mathrm{N}\left(\alpha+\beta(h_t-\alpha), \xi_t\sigma^2\right), \\ \xi_t|\nu &\sim \mathrm{IG}(\nu/2,\nu/2). \end{aligned}$$

This can be hierarchically expressed as

$$\begin{aligned} \begin{pmatrix} y_t \\ h_{t+1} \end{pmatrix} &\sim \mathrm{N}\left(\begin{pmatrix} 0 \\ \alpha+\beta(h_t-\alpha) \end{pmatrix}, \xi_t \begin{pmatrix} e^{h_t} & \rho\sigma e^{h_t/2} \\ \rho\sigma e^{h_t/2} & \sigma^2 \end{pmatrix}\right), \\ \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} &\sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \xi_t \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \\ h_1 &\sim \mathrm{N}\left(\alpha, \frac{1-\beta^2}{\xi_1\sigma^2}\right), \\ \xi_t &\sim \mathrm{IG}\left(\frac{\nu}{2}, \frac{\nu}{2}\right). \end{aligned}$$

Wang, Chan, and Choy (2011) considered an alternative approach and stated, "the main difference between our [Wang's] approach and Choy's approach lies in the order of conditioning on the mixing parameter and the conditional density of $y_t|h_{t+1}$". They begin with the bivariate t-distribution and factorize it into a conditional t-distribution by dividing the joint density with a marginal density

$$f(y_t|h_{t+1}) = \frac{f(y_t, h_{t+1})}{f(h_{t+1})},$$

where

$$\begin{aligned} f(y_t|h_{t+1}, h_t) &= \frac{\Gamma(\frac{\nu}{2}+1)}{\Gamma(\frac{\nu}{2})\nu\pi\sigma\exp(\frac{h_t}{2})\sqrt{1-\rho^2}}\left[1+\frac{1}{\nu}\frac{M_t^2-2\rho M_t N_t+N_t^2}{D_t}\right]^{(\frac{\nu}{2}+1)}, \\ D_t &= \sigma^2(1-\rho^2)\exp(h_t), \\ M_t &= y_t\sigma, \\ N_t &= \exp\left(\frac{h_t}{2}\right)\{h_{t+1}-[\alpha+\beta(h_t-\mu)]\}, \end{aligned}$$

and the density of the marginal t-distribution for $h_{t+1}|h_t$ is

$$f(h_{t+1}|h_t) = \frac{\Gamma(\frac{\nu}{2}+1)}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}\sigma}\left[1+\frac{1}{\nu\sigma^2}\{h_{t+1}-[\alpha+\beta(h_t-\mu)]\}^2\right]^{-(\frac{\nu+1}{2})}.$$

Then using (4.7), it can be shown that

$$\begin{aligned} y_t|h_{t+1},h_t &\sim t_{\nu+1}\left(\frac{\rho}{\sigma}\exp\left(\frac{h_t}{2}\right)\{h_{t+1}-[\alpha+\beta(h_t-\alpha)]\},\right. \\ &\qquad \left.\left(\frac{\nu}{\nu+1}\right)(1-\rho^2)\exp(h_t)\left[1+\frac{1}{\nu\sigma^2}\{h_{t+1}-[\alpha+\beta(h_t-\alpha)]\}^2\right]\right), \\ h_{t+1}|h_t &\sim t_\nu(\mu+\phi(h_t-\mu),\sigma^2). \end{aligned}$$

Thereafter they consider a SMN representation for each component of the conditional and marginal t-distributions given by

$$\begin{aligned} p(y_t|h_{t+1},h_t) &\sim \mathrm{N}\left(\frac{\rho}{\sigma}\exp\left(\frac{h_t}{2}\right)\{h_{t+1}-[\alpha+\beta(h_t-\alpha)]\},\right. \\ &\qquad \left.\frac{1}{\lambda_{y_t}}\left(\frac{\nu}{\nu+1}\right)(1-\rho^2)\exp(h_t)\left[1+\frac{1}{\nu\sigma^2}\{h_{t+1}-[\alpha+\beta(h_t-\alpha)]\}^2\right]\right), \\ h_{t+1}|h_t &\sim \mathrm{N}(\alpha+\beta(h_t-\alpha),\frac{\sigma^2}{\lambda_{h_{t+1}}}), \\ \lambda_{y_t} &\sim \mathrm{Ga}\left(\frac{\nu+1}{2},\frac{\nu+1}{2}\right), \\ \lambda_{h_t} &\sim \mathrm{Ga}\left(\frac{\nu}{2},\frac{\nu}{2}\right). \end{aligned}$$

However, it is important to note this model is different from that of Choy and Chan (2000) because of a difference in the order of factorization and conditioning of the mixing variable. This model offers extra flexibility as it contains two mixing variables for each component allowing the level of dispersion to vary across components at each time point. This is the clear benefit of using this model. However, both models still rely on the same shape parameter ν. Since our aim is to extend the standard SV-LVG model to adopt long memory effect, we consider the approach of Choy and Chan (2000) for simplicity.

4.4.3 Sampling the volatility level parameter α in the GMA-SV-LVG-HC model

The posterior distribution of α is given by

$$
\begin{aligned}
& p(\alpha|\boldsymbol{Y}^*, h_1, \boldsymbol{h}, \beta, \sigma^2, \rho, \xi_0, \boldsymbol{\xi}) \propto f(\boldsymbol{Y}^*|h_1, \boldsymbol{h}, \alpha, \beta, \sigma^2, \rho, \boldsymbol{\xi}) f(h_1|\alpha, \beta, \sigma^2, \xi_0) f(\boldsymbol{h}|\boldsymbol{h}_{1:T}, \alpha, \beta, \sigma^2, \boldsymbol{\xi}) p(\alpha) \\
= & \frac{1}{(\sqrt{2\pi})^T (\sqrt{1-\rho^2})^T \prod_{t=1}^T \xi_t^{\frac{1}{2}} \exp(h_t/2)} \exp\left\{ -\sum_{t=1}^T \frac{[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha))]^2}{2(1-\rho^2)\xi_t} \right\} \times \\
& \frac{(1-\beta^2)^{\frac{1}{2}}}{\sqrt{2\pi}\sigma\xi_0^{\frac{1}{2}}} \exp\left\{ -\frac{(h_1 - \alpha)^2(1-\beta^2)}{2\sigma^2} \right\} \times \\
& \frac{1}{(\sqrt{2\pi}\sigma)^T \prod_{t=1}^T \xi_t^{\frac{1}{2}}} \exp\left\{ -\frac{\sum_{t=1}^T (h_{t+1} - \alpha - \beta(h_t - \alpha))^2}{2\sigma^2} \right\} \times \frac{1}{\sqrt{2\pi}\sigma_\alpha} \exp\left\{ -\frac{(\alpha - \mu_\alpha)^2}{2\sigma_\alpha^2} \right\}.
\end{aligned}
$$

We consider the first term:

$$
\begin{aligned}
& \frac{1}{(\sqrt{2\pi})^T (\sqrt{1-\rho^2})^T \prod_{t=1}^T \xi_t^{\frac{1}{2}} \exp(h_t/2)} \exp\left\{ -\sum_{t=1}^T \frac{[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha))]^2}{2(1-\rho^2)\xi_t} \right\} \\
\propto & \exp\left\{ -\sum_{t=1}^T \frac{[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha))]^2}{2(1-\rho^2)\xi_t} \right\} \\
= & \exp\left\{ -\frac{1}{2(1-\rho^2)} \sum_{t=1}^T \xi_t^{-1} \left[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}[h_{t+1} - \alpha - \beta(h_t - \alpha)] \right]^2 \right\} \\
\propto & \exp\left\{ -\frac{1}{2(1-\rho^2)} \sum_{t=1}^T \xi_t^{-1} \left[-2\frac{y_t^*}{e^{h_t/2}} \frac{\rho}{\sigma}[h_{t+1} - \alpha - \beta(h_t - \alpha)] + \frac{\rho^2}{\sigma^2}[h_{t+1} - \alpha - \beta(h_t - \alpha)]^2 \right] \right\} \\
\propto & \exp\left\{ -\frac{1}{2(1-\rho^2)} \sum_{t=1}^T \xi_t^{-1} \left[2\frac{y_t^*}{e^{h_t/2}} \frac{\rho(1-\beta)}{\sigma}\alpha + \frac{\rho^2}{\sigma^2}[h_{t+1} - \alpha - \beta(h_t - \alpha)]^2 \right] \right\} \\
\propto & \exp\left\{ -\frac{\rho}{2\sigma(1-\rho^2)} \sum_{t=1}^T \xi_t^{-1} \left[2\frac{y_t^*}{e^{h_t/2}}(1-\beta)\alpha + \frac{\rho}{\sigma}[\alpha^2(1-\beta)^2 - 2\alpha(h_{t+1} - \beta h_t)(1-\beta)] \right] \right\} \\
= & \exp\left\{ -\frac{1}{2} \left(\alpha^2 \frac{(1-\beta)^2\rho^2}{(1-\rho^2)\sigma^2} \sum_{t=1}^T \xi_t^{-1} - 2\alpha \frac{(1-\beta)\rho}{(1-\rho^2)\sigma} \sum_{t=1}^T \xi_t^{-1} \left[(h_{t+1} - \beta h_t)\frac{\rho}{\sigma} - \frac{y_t^*}{e^{h_t/2}} \right] \right) \right\}.
\end{aligned}
$$

The derivation of the remaining terms can be found in (4.18) to (4.21) for the GMA-SV-LVG model, by replacing σ^2 with $\xi_t\sigma^2$. We then have the result

$$
\alpha|\boldsymbol{Y}^*, h_1, \boldsymbol{h}, \beta, \sigma^2, \rho, \xi_0, \boldsymbol{\xi} \sim N\left(V_\alpha M_\alpha, V_\alpha\right),
$$

where

$$
\begin{aligned}
V_\alpha &= \left(\frac{(1-\beta)^2\rho^2}{(1-\rho^2)\sigma^2} \sum_{t=1}^{T} \xi_t^{-1} + \frac{(1-\beta^2)}{\sigma^2} \sum_{t=1}^{T} \xi_t^{-1} + \frac{(1-\beta^2)\xi_0^{-1}}{\sigma^2} + \frac{1}{\sigma_\alpha^2} \right)^{-1}, \\
M_\alpha &= \frac{(1-\beta)\rho}{(1-\rho^2)\sigma} \sum_{t=1}^{T} \xi_t^{-1} \left[(h_{t+1} - \beta h_t) \frac{\rho}{\sigma} - \frac{y_t^*}{e^{h_t/2}} \right] + \frac{(1-\beta^2)}{\sigma^2} \xi_0^{-1} h_1 \\
&\quad + \frac{(1-\beta)}{\sigma^2} \sum_{t=1}^{T} \xi_t^{-1} [h_{t+1} - \beta h_t] + \frac{\mu_\alpha}{\sigma_\alpha^2}.
\end{aligned}
$$

4.4.4 Sampling the volatility persistence parameter β in the GMA-SV-LVG-HC model

We follow the method of Chib and Greenberg (1994) to use the density of h_1 as the proposal density. Therefore, we do not consider the density of h_1 when deriving the proposal density of interest for β as

$$
\begin{aligned}
& p(\beta|\boldsymbol{Y}^*, h_1, \boldsymbol{h}, \alpha, \sigma^2, \rho, \boldsymbol{\xi}) \propto f(\boldsymbol{Y}^*|h_1, \boldsymbol{h}, \alpha, \beta, \sigma^2, \rho, \boldsymbol{\xi}) f(\boldsymbol{h}|\boldsymbol{h}_{1:T}, \alpha, \beta, \sigma^2, \boldsymbol{\xi}) \times p(\beta) \\
= \;& \frac{1}{(\sqrt{2\pi})^T (\sqrt{1-\rho^2})^T \prod_{t=1}^{T} \xi_t^{\frac{1}{2}} \exp(h_t/2)} \exp\left\{ - \sum_{t=1}^{T} \frac{[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha))]^2}{2(1-\rho^2)\xi_t} \right\} \times \\
& \frac{1}{(\sqrt{2\pi}\sigma)^T \prod_{t=1}^{T} \xi_t^{\frac{1}{2}}} \exp\left\{ - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \xi_t^{-1} [h_{t+1} - \alpha - \beta(h_t - \alpha)]^2 \right\} \times \\
& \frac{1}{\sqrt{2\pi}\sigma_\alpha} \exp\left\{ - \frac{(\beta - \mu_\beta)^2}{2\sigma_\beta^2} \right\} \\
\propto \;& \exp\left\{ -\frac{1}{2} \left[\sum_{t=1}^{T} \frac{[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1} - \alpha - \beta(h_t - \alpha))]^2}{(1-\rho^2)\xi_t} \right.\right. \\
& \left.\left. + \sum_{t=1}^{T} \frac{[h_{t+1} - \alpha - \beta(h_t - \alpha)]^2}{\sigma^2 \xi_t} + \frac{(\beta - \mu_\beta)^2}{\sigma_\beta^2} \right] \right\}.
\end{aligned}
$$

Now, we consider the first term in the exponent as

$$\begin{aligned}
&\sum_{t=1}^{T} \frac{[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}[h_{t+1} - \alpha - \beta(h_t - \alpha)]^2}{(1-\rho^2)\xi_t} \\
= &\sum_{t=1}^{T} \frac{\frac{y_t^{*2}}{e^{h_t}} - 2\frac{y_t^*}{e^{h_t/2}}\frac{\rho}{\sigma}[h_{t+1} - \alpha - \beta(h_t - \alpha)] + \frac{\rho^2}{\sigma^2}[h_{t+1} - \alpha - \beta(h_t - \alpha)]^2}{(1-\rho^2)\xi_t} \\
= &\sum_{t=1}^{T} \frac{2\frac{y_t^*}{e^{h_t/2}}\frac{\rho}{\sigma}\beta(h_t - \alpha) + \frac{\rho^2}{\sigma^2}[\beta^2(h_t - \alpha)^2 - 2\beta(h_t - \alpha)(h_{t+1} - \alpha)]}{(1-\rho^2)\xi_t} + \text{terms independent of } \beta \\
= &\frac{\rho^2}{\sigma^2(1-\rho^2)}\sum_{t=1}^{T}\xi_t^{-1}(h_t - \alpha)^2\beta^2 + \frac{2\rho}{\sigma(1-\rho^2)}\sum_{t=1}^{T}\xi_t^{-1}\left[\frac{y_t^*}{e^{h_t/2}}(h_t - \alpha) - \frac{\rho}{\sigma}(h_t - \alpha)(h_{t+1} - \alpha)\right]\beta \\
&+ \text{terms independent of } \beta.
\end{aligned}$$

The derivation of the remaining terms can be found in (4.25) and (4.26), by replacing σ^2 with $\xi_t\sigma^2$. Hence, the terms in the exponent can be expressed as

$$\begin{aligned}
&\left(\frac{\rho^2}{\sigma^2(1-\rho^2)}\sum_{t=1}^{T}\xi_t^{-1}(h_t - \alpha)^2 + \frac{1}{\sigma^2}\sum_{t=1}^{T}\xi_t^{-1}(h_t - \alpha)^2 + \frac{1}{\sigma_\beta^2}\right)\beta^2 \\
&-2\left(\frac{\rho}{\sigma(1-\rho^2)}\sum_{t=1}^{T}\xi_t^{-1}(h_t - \alpha)\left[\frac{\rho}{\sigma}(h_{t+1} - \alpha) - \frac{y_t^*}{e^{h_t/2}}\right] + \frac{1}{\sigma^2}\sum_{t=1}^{T}\xi_t^{-1}(h_t - \alpha)(h_{t+1} - \alpha) + \frac{\mu_\beta}{\sigma_\beta^2}\right)\beta,
\end{aligned}$$

compared to (4.27) for the GMA-SV-LVG model. Therefore,

$$\beta|\boldsymbol{Y}^*, h_1, \boldsymbol{h}, \alpha, \sigma^2, \rho, \boldsymbol{\xi} \sim N(V_\beta M_\beta, V_\beta), \tag{4.40}$$

where

$$\begin{aligned}
M_\beta &= \frac{\rho}{\sigma(1-\rho^2)}\sum_{t=1}^{T}\xi_t^{-1}(h_t - \alpha)\left[\frac{\rho}{\sigma}(h_{t+1} - \alpha) - \frac{y_t^*}{e^{h_t/2}}\right] + \frac{1}{\sigma^2}\sum_{t=1}^{T}\xi_t^{-1}(h_t - \alpha)(h_{t+1} - \alpha) + \frac{\mu_\beta}{\sigma_\beta^2}, \\
V_\beta &= \left(\frac{1}{\sigma^2(1-\rho^2)}\sum_{t=1}^{T}\xi_t^{-1}(h_t - \alpha)^2 + \frac{1}{\sigma_\beta^2}\right)^{-1},
\end{aligned}$$

Again, in order to implement the MH step, we first sample β^* from the proposal density in (4.40) and use the density of h_1

$$q(\beta) = \frac{\sqrt{1-\beta^2}}{\sqrt{2\pi\sigma^2\xi_0}}\exp\left\{-\frac{(h_1 - \alpha)^2(1-\beta^2)}{2\sigma^2\xi_0}\right\},$$

as the target density. We work with $\log q(\beta)$ so that

$$\log q(\beta) = \frac{1}{2}\log(1-\beta^2) - \frac{1}{2}\log(2\pi\sigma^2\xi_0) - \frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2\xi_0},$$

and β^* is accepted for $\beta^{(m)}$ with probability $\min\left\{\exp[q(\beta^*) - q(\beta^{(m-1)})], 1\right\}$.

4.4.5 Sampling the volatility of volatility parameter σ^2 in the GMA-SV-LVG-HC model

The conditional posterior distribution of σ^2 is

$$\begin{aligned}
& p(\sigma^2|\boldsymbol{Y}^*, h_1, \boldsymbol{h}, \alpha, \beta, \rho, \xi_0, \boldsymbol{\xi}) \\
\propto\ & f(\boldsymbol{Y}^*|h_1, \boldsymbol{h}, \alpha, \beta, \sigma^2, \rho, \boldsymbol{\xi}) f(h_1|\alpha, \beta, \sigma^2, \xi_0) f(\boldsymbol{h}|\boldsymbol{h}_{1:T}, \alpha, \beta, \sigma^2, \boldsymbol{\xi}) p(\sigma^2) \\
=\ & \frac{1}{[\sqrt{2\pi(1-\rho^2)}]^T \prod_{t=1}^{T} \xi_t^{\frac{1}{2}} \exp(\frac{h_t}{2})} \exp\left\{-\sum_{t=1}^{T} \frac{(\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}[h_{t+1} - \alpha - \beta(h_t - \alpha)])^2}{2(1-\rho^2)\xi_t}\right\} \times \\
& \frac{\sqrt{1-\beta^2}}{\sqrt{2\pi\sigma^2\xi_0}} \exp\left\{-\frac{1}{2\sigma^2}\xi_0^{-1}(h_1-\alpha)^2(1-\beta^2)\right\} \times \\
& \frac{1}{(\sqrt{2\pi}\sigma)^T \prod_{t=1}^{T} \xi_t^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T} \xi_t^{-1}[h_{t+1} - \alpha - \beta(h_t - \alpha)]^2\right\} \frac{(\frac{b}{2})^{\frac{a}{2}}}{\Gamma(\frac{a}{2})} \sigma^{-2(\frac{a}{2}+1)} \exp\left(-\frac{(\frac{b}{2})}{\sigma^2}\right).
\end{aligned}$$

It is clear that we cannot continue as usual to find the posterior distribution of σ^2, because we require the exponent term to be in terms of σ^2 only, and not σ. This problem originates from including the conditional of $y_t^*|h_{t+1}, h_t$. Therefore, we once again follow the procedure set out in Chib and Greenberg (1994), and use the conditional density of $y_t^*|h_{t+1}, h_t$ as the proposal density for the MH sampling scheme. Therefore, the proposal density of interest for σ^2 is

$$\begin{aligned}
& p(\sigma^2|h_1, \boldsymbol{h}, \alpha, \beta, \xi_0, \boldsymbol{\xi}) \propto f(h_1|\alpha, \beta, \sigma^2, \xi_0) f(\boldsymbol{h}|\boldsymbol{h}_{1:T}, \alpha, \beta, \sigma^2, \boldsymbol{\xi}) \times p(\sigma^2) \\
=\ & \frac{\sqrt{1-\beta^2}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\xi_0^{-1}(h_1-\alpha)^2(1-\beta^2)\right\} \times \\
& \frac{1}{(\sqrt{2\pi}\sigma)^T} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T} \xi_t^{-1}[h_{t+1} - \alpha - \beta(h_t - \alpha)]^2\right\} \times \frac{(\frac{b}{2})^{\frac{a}{2}}}{\Gamma(\frac{a}{2})} \sigma^{-2(\frac{a}{2}+1)} \exp\left(-\frac{(\frac{b}{2})}{\sigma^2}\right),
\end{aligned}$$

which is the conditional density of the standard SV model when σ^2 is replaced by $\sigma^2\xi_t$ for the first two product terms. We use this proposal density

$$\sigma^2|h_1,\boldsymbol{h},\alpha,\beta,\rho,\xi_0,\boldsymbol{\xi}\sim \text{IG}\left(\frac{T+a}{2},\frac{b+\xi_0^{-1}(h_1-\alpha)^2(1-\beta^2)+\sum_{t=1}^{T}\xi_t^{-1}[h_{t+1}-\alpha-\beta(h_t-\alpha)]^2}{2}\right), \tag{4.41}$$

to draw $(\sigma^2)^*$ and use

$$f(\boldsymbol{Y}^*|h_1,\boldsymbol{h},\alpha,\beta,\sigma^2,\rho,\boldsymbol{\xi})=\frac{1}{[\sqrt{2\pi(1-\rho^2)}]^T\prod_{t=1}^{T}\xi_t^{\frac{1}{2}}\exp(\frac{h_t}{2})}\exp\left\{-\sum_{t=1}^{T}\frac{[\frac{y_t^*}{e^{h_t/2}}-\frac{\rho}{\sigma}(h_{t+1}-\alpha-\beta(h_t-\alpha))]^2}{2(1-\rho^2)\xi_t}\right\},$$

as the target density. We work on the log density

$$\begin{aligned}
\log q(\sigma^2) &= \log f(\boldsymbol{Y}^*|h_1,\boldsymbol{h},\alpha,\beta,\sigma^2,\rho,\boldsymbol{\xi})\\
&= -\sum_{t=1}^{T}\frac{[\frac{y_t^*}{e^{h_t/2}}-\frac{\rho}{\sigma}(h_{t+1}-\alpha-\beta(h_t-\alpha))]^2}{2(1-\rho^2)\xi_t}+\text{ terms independent of }\sigma^2\\
&= \frac{\rho}{1-\rho^2}\sum_{t=1}^{T}\frac{y_t^*e^{-h_t/2}[h_{t+1}-\alpha-\beta(h_t-\alpha)]}{\sigma\xi_t}\\
&\quad -\frac{\rho^2}{2(1-\rho^2)}\sum_{t=1}^{T}\frac{[h_{t+1}-\alpha-\beta(h_t-\alpha)]^2}{\sigma^2\xi_t}+\text{ terms independent of }\sigma^2.We\,not\,e\,this\,method\,prove
\end{aligned}$$

Once again, $(\sigma^2)^*$ is accepted for $(\sigma^2)^{(m)}$ with probability $\min\left\{\exp[q((\sigma^2)^*)-q((\sigma^2)^{(m-1)})],1\right\}$.

4.4.6 Sampling the leverage parameter ρ in the GMA-SV-LVG-HC model

Since ρ is bounded, a grid-based method is used to sample from the posterior distribution

$$\begin{aligned}
&p(\rho|\boldsymbol{Y}^*,h_1,\boldsymbol{h},\alpha,\beta,\sigma^2,\boldsymbol{\xi})\propto f(\boldsymbol{Y}^*|h_1,\boldsymbol{h},\alpha,\beta,\sigma^2,\rho,\boldsymbol{\xi})\times p(\rho)\\
=\;& \frac{1}{(\sqrt{2\pi})^T(\sqrt{1-\rho^2})^T\prod_{t=1}^{T}\xi_t\exp(\frac{h_t}{2})}\exp\left\{-\sum_{t=1}^{T}\frac{(\frac{y_t^*}{e^{h_t/2}}-\frac{\rho}{\sigma}(h_{t+1}-\alpha-\beta(h_t-\alpha)))^2}{2(1-\rho^2)\xi_t}\right\}\\
&\frac{1}{\sqrt{2\pi}\sigma_\rho}\exp\left\{-\frac{(\rho-\mu_\rho)^2}{2\sigma_\rho^2}\right\}\\
\propto\;& \frac{1}{(1-\rho^2)^{T/2}}\exp\left\{-\sum_{t=1}^{T}\frac{[\frac{y_t^*}{e^{h_t/2}}-\frac{\rho}{\sigma}(h_{t+1}-\alpha-\beta(h_t-\alpha))]^2}{2(1-\rho^2)\xi_t}-\frac{(\rho-\mu_\rho)^2}{2\sigma_\rho^2}\right\}\\
\propto\;& \frac{1}{(1-\rho^2)^{T/2}}\exp\left\{-\sum_{t=1}^{T}\frac{[\frac{y_t^*}{e^{h_t/2}}-\frac{\rho}{\sigma}(h_{t+1}-\alpha-\beta(h_t-\alpha))]^2}{2(1-\rho^2)\xi_t}-\frac{\rho^2}{2\sigma_\rho^2}+\frac{2\mu_\rho\rho}{2\sigma_\rho^2}\right\}.
\end{aligned}$$

Hence we have

$$\log p(\rho|\boldsymbol{Y},\boldsymbol{h},\alpha,\beta,\sigma^2,\boldsymbol{\xi}) = -\frac{T}{2}\log(1-\rho^2) - \sum_{t=1}^{T}\frac{[\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}(h_{t+1}-\alpha-\beta(h_t-\alpha))]^2}{2(1-\rho^2)\xi_t} - \frac{\rho^2}{2\sigma_\rho^2} + \frac{2\mu_\rho\rho}{2\sigma_\rho^2} + \text{ terms independent of } \rho.$$

As discussed in Section 4.2.9, ρ can be efficiently estimated using MAP sampler with tolerance level 10^{-5}.

4.4.7 Sampling the mixing variable ξ_t in the GMA-SV-LVG-HC model

To sample $(\xi_1, \boldsymbol{\xi})$ first note that each element is independent and $\xi_t \sim \text{IG}(\frac{\nu}{2}, \frac{\nu}{2})$. When $t = 0$,

$$p(\xi_1|h_1,\alpha,\beta,\sigma^2,\nu) \propto f(h_1|\xi_1,\alpha,\beta,\sigma^2)\, f(\xi_1)$$

so that:

$$\xi_1|h_1,\alpha,\beta,\sigma^2,\nu \ \sim \text{IG}\left(\frac{\nu+1}{2}, \frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2} + \frac{\nu}{2}\right).$$

Similarly, when $t > 0$, we have

$$p(\xi_{t+1}|h_{t+1},h_t,\alpha,\beta,\sigma^2,\rho,\nu,y_t^*) \propto f(h_{t+1}|h_t,\xi_{t+1},\alpha,\beta,\sigma^2) f(y_t^*|h_{t+1},h_t,\xi_{t+1},\alpha,\beta,\sigma,\rho)\, f(\xi_{t+1})$$

$$\xi_{t+1}|h_{t+1},h_t,\alpha,\beta,\sigma^2,\rho,\nu,y_t^* \ \sim \text{IG}\left(\frac{\nu}{2}+1, S_\xi\right)$$

where $S_\xi = \dfrac{[h_{t+1}-\alpha-\beta(h_t-\alpha)]^2}{2\sigma^2} + \dfrac{\{\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}[h_{t+1}-\alpha-\beta(h_t-\alpha)]\}^2}{2(1-\rho^2)} + \dfrac{\nu}{2}$.

4.4.8 Sampling the latent volatilities h and long memory parameters u and d in the GMA-SV-LVG-HC model

The sampling of the latent volatilities h_1 and $\boldsymbol{h}$ as well as the long memory parameters d and u can be found in Appendix F. Next, we look at aspects of Cryptocurrencies.

4.5 Background for Cryptocurrencies

4.5.1 Initial important concepts

Before discussing Cryptocurrencies, we introduce some initial important concepts.

- **Mining**: To mine is to discover and solve blocks which are along the blockchain. A financial reward, in the form of a Cryptocurrency, can be obtained for solving the algorithm, called a mining reward.
- **Hashrate**: This refers to the speed at which a block is discovered and the rate at which the associated mathematical problem is solved.
- **Blockchain**: A system which allows the creation of a digital ledger of transactions. This digital ledger is decentralized since all users of the blockchain carry the same information. It can be envisaged as a large set of computers working together to create a network, instead of one computer (e.g. a bank's server) to store a ledger. The blockchain records every single transaction that has ever happened on it.
- **Block**: Many blocks make a blockchain. It can be interpreted as pages in the ledger.
- **Smart contract**: An unalterable agreement stored on the blockchain that has a specific logic operation. Once signed, it can never be altered.

4.5.2 History of cryptocurrencies

Cryptocurrencies have a rich history and date back to the early 1980s which was a few years after the first commercial uses of the internet began. The initial proponents of Cryptocurrencies began within the cryptographic community and was spearheaded by Chaum (1981). Chaum (1981) explains the use of a so-called "binding" algorithm, which enables decentralized web-based encryption for money transfer. This algorithm facilitated secure exchange between parties, and therefore laid the foundation of electronic currency transfers. This process is known as "blinded money". Thereafter, the idea of Cryptocurrencies reached an inner circle of enthusiasts who began to commercialize the ideas of anonymous money. The first Cryptocurrency created as a result of this was DigiCash, which shared similar traits to Cryptocurrencies today,

except DigiCash itself had a monopoly on supply - similar to central banks today. The company responsible for creating DigiCash was based in the Netherlands, and initially dealt with individuals. However, the Netherlands central bank soon objected to this and subsequently DigiCash was only allowed to be used by licensed financial institutions.

During the same period, a host of other payment systems also emerged which later failed, including DigiCash. In fact, PayPal is the only payment system which emanated from the history of Cryptocurrencies that is still being used today.

The deep underlying philosophies of DigiCash were later resurrected by Nakamoto (2008) via Bitcoin. Nakamoto (2008) outlined the foundation of how Bitcoin would operate and solve the so called "double-spending problem". Once again, a small group of enthusiasts began to exchange Bitcoin with one another and also mine the currency. Slowly thereafter, large online merchants began to accept Bitcoin as a medium of exchange such as Newegg and Microsoft.

Bitcoin is the most widely used Cryptocurrency with the largest market capitalization. It is a fully digital currency that can be exchanged in a worldwide peer-to-peer (P2P) network, which is intended to share data amongst its users. Traditional applications of P2P networking traditionally involved music and video sharing platforms. However, Bitcoin is not a string of data that can be duplicated outside of the network and therefore be synthetically created outside of the Blockchain. A Bitcoin is an entry on a huge global ledger (called the Blockchain). As of July 2018, the size of the Bitcoin blockchain is 164.4 Gigabytes.

To date, there are 17.2 million Bitcoin in circulation, and it is estimated that one third have been lost due to people losing their password and login information. It has a programmed supply limit of 21 million Bitcoin, and is forecasted to reach its supply limit by the year 2140. Approximately every 4 years, the reward for mining Bitcoin is halved, until it will reach 0. Interestingly, Bitcoin will never theoretically reach the 21 million supply limit, since the exact value is in fact $20,999,999.9769$. When there is no longer any reward for miners, the transaction fees will be the reason why mining will continue.

In general, when Cryptocurrencies are transacted, there is no packet of data which is being sent from the seller to the receiver, rather, there is a re-entry on the blockchain.

There are advantages and disadvantages for transacting with Cryptocurrencies and Table 4.8 provides a comparison.

Advantages of Cryptocurrencies	**Disadvantages of Cryptocurrencies**
Most Cryptocurrencies have limited supply and may therefore uphold their value such as precious metals do.	It facilitates black market activity.
Its control is decentralized.	It leads to tax evasion.
Miners are financially rewarded. This is important since miners upkeep the ledger, and therefore the authenticity of all transactions. This is in contrast to financial institutions who are not directly paid for authenticating their ledger.	Data loss will lead to financial loss.
It ensures privacy when transacting through complete anonymity.	It has potential for price manipulation.
Near impossible for governments to freeze assets. This is particularly relevant for citizens of repressive states.	It is not readily available to be transacted into fiat currency.
It has very low transaction fees.	Money can be virtually "lost" without passwords, with no way to recover them.
It has low barriers to international transactions	It has no refunds, or oversight for disputes.

TABLE 4.8: Advantages and disadvantages of transacting with Cryptocurrencies.

4.5.3 Comparison with fiat currencies

In essence, Cryptocurrencies use cryptographic encryption techniques to regulate the generation of units and verify transfers whilst operating independent of a central bank. This is in stark contrast to traditional fiat currencies whereby their value is derived from macroeconomic elements such as terms of trade, interest rates and inflation. These aspects are further summarized in Table 4.9 as below.

Aspect	**Fiat Currency**	**Cryptocurrency**
Generation of units	Monetary policy	Awarded to users for mining
Controlled by	Central bank	Decentralized
Settlement time	Three days	Varied however typically between one second to 48 hours.
Ledger authentication	Financial institutions	Many computers

TABLE 4.9: Comparison of fiat and Cryptocurrencies.

4.6 Conclusion

This chapter derives the estimation methodologies for the GMA-SV-LVG-HC model and its sub-models and reviews some background of Cryptocurrencies. The next two chapters are two papers which analyze Cryptocurrency returns using the GMA-SV-LVG-HC model and investigate the properties of Cryptocurrencies. Chapter 5 is our second publication (which is currently under review after a revision was invited) in *Econometrics and Statistics,* and Chapter 6 is our third publication accepted in *Finance Research Letters.*

Chapter 5

Bivariate Student-t long memory stochastic volatility models with leverage

"The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data"

John Tukey

This paper proposes a Gegenbauer long memory stochastic volatility model with leverage and a bivariate Student's t-error distribution to model the innovations of the observation and latent volatility jointly for cryptocurrency time series. This paper is inspired by the deep rooted characteristics found in cryptocurrencies. To date, little to no work has been done within the econometrics literature to understand their properties. To do this, a rigorous in-sample simulation is conducted to assess the performance of the model with its nested alternatives and study the behavior of many cryptocurrencies - in particular Bitcoin. The data analysis is initiated with a broad scope of 114 cryptocurrencies, then a more detailed understanding of five of the most popular cryptocurrencies and followed up with a specific focus on Bitcoin. The model parameters are estimated with Bayesian approach using Markov Chain Monte Carlo sampling. In order to implement model selection, the Deviance Information Criterion (DIC) is used. The models are compared with many popular models including those commonly used in industry. Proposed models are applied in a Value-at-Risk (VaR) context and several measures are used to assess model performance.

5.1 Introduction

Academic interests in anonymous communications research date back to the early 1980s (Chaum, 1981), and the first digital currency, *DigiCash* was launched in 1990, which offered anonymity through cryptographic protocols. This was later resurrected by Nakamoto (2008) who adopted the philosophies of (Chaum, 1981) with the addition of crowd sourcing and peer-to-peer networking which avoided centralized control. In essence, all cryptocurrencies are digital ledgers which contain names and balances. The philosophy of cryptocurrencies is that there is no central bank where the currency derives its value from, and each person transacting has faith in the system. One of the underlying goals is to avoid central control so each person owns their copy of the ledger. To date, this has manifested itself into a growing cryptocurrency community who are accepting cryptocurrencies as a means of exchange. Banks such as UBS and Credit Suisse have now developed a streamline payment mechanism for institutional investors. It is estimated that more than $\$2.3$ billion USD is invested in Cryptocurrency hedge funds globally. Governments around in the world have also started legislative proceedings for regulation and consumer safety.

Our work is deeply motivated by the unique characteristics found in cryptocurrency data, which have gained large media attention as of late. Interestingly, the term 'cryptocurrency' has been trending upwards as a Google.com search word since 2015. Although still in its infancy, there has been a lot of attention surrounding this topic with regulators, investors and governments weighing in on the discussion. Unfortunately, little work to date has been done in the statistical literature on understanding the properties of cryptocurrencies in general, and we endeavor for our work to be pioneering in this respect.

The most popular and largest cryptocurrency by market capitalization undoubtedly is Bitcoin. A $1,000 USD investment in Bitcoin in July of 2010 would have returned $81,000,000 just 7 years later. Due to these extremely strong returns, coupled with their nature, Bitcoin or cryptocurrencies in general face scrutiny as being speculative (Cheah and Fry, 2015). Skeptics argue this is reminiscent of a tech-bubble of the early 2000s or even the Tulip mania of the seventeenth century. Conversely, there is evidence to suggest the cryptocurrency market is still in its infancy and is inefficient (Urquhart, 2016), with properties such as price clustering (Urquhart, 2017). There is,

however, a strong growing network of Bitcoin users and academics who are shedding light on this new technology. In this work, we discuss a large investable sample of cryptocurrencies, and conditionally measure some important stylized facts. It is due to these unique stylized facts that we contribute a model which is designed to capture these effects.

To date, there are more than $2,000$ investable cryptocurrencies. Unlike their fiat counterparts, the differences are not due to sovereign macroeconomic factors. Their nuances are due to more technical factors such as transaction times, and the supporting infrastructure that facilitate their trade. These unique factors can have statistical interpretations as is discussed later in Chapter 5.5.

A commonly observed property found in financial time series is the long memory effect, see, for example, Granger and Joyeux (1980) and Hosking (1981). A stationary time series y_t is said to be long memory if $\sum_{k=0}^{\infty}|\delta(k)|$ diverges, where $\delta(k)$ is the k^{th}-lag autocovariance. This class of time series generalizes the usual Box-Jenkins ARIMA model by modelling long term correlation structures as suggested by Mandelbrot and Ness (1968). An appealing generalization of traditional long memory models are generalised autoregressive fractional integrated moving average (GARFIMA) models; whereby Gegenbauer polynomials replace the plain long memory fractional differencing operator. Gegenbauer polynomials were first introduced to the time-series community by Gray, Zhang, and Woodward (1989b). The novelty in such polynomials lie in their orthogonality and recursion properties.

Cryptocurrencies face different issues compared to their fiat counterparts which can be better understood with familiar statistical tools. One of these issues is a delay in their transaction times whereby future volatility can have an affect on the currently observed price. This phenomena is closely related to the familiar leverage effect. The leverage effect has its roots in the asymmetric return-volatility relationship attributed to financial leverage or debt-to-equity ratios (Christie, 1982). Put simply, it is the claim that one day ahead volatility is negatively correlated to currently observed returns (Nelson, 1991). This is purported to occur due to traditional stock prices negatively reacting to bad news thereby causing an increase in the debt-to-equity ratio of the firm leading to higher expected future volatility (Engle and Ng, 1993).

The most heated debate about cryptocurrencies are their extreme variability characteristics. The statistical literature is full of meaningful ways to explain this. A

common procedure found in the literature is to modify the observation and/or the latent equation to include a non-Gaussian heavy-tailed distributions (Chib, Nardari, and Shephard, 2002; Berg, Meyer, and Yu, 2004; Yu, 2005; Omori et al., 2007; Omori and Watanabe, 2008), and this would be a natural step to model cryptocurrencies.

An alternative way to measure such variability is modelling the errors stochastically. The Stochastic Volatility model (SV) was first introduced by Taylor (1986) to describe the time varying nature of volatility typically found in financial returns. It is widely viewed as a competitor to the generalised autoregressive conditional heteroscedastic (GARCH) model of Bollerslev (1986) because it adequately displays the main so called stylized-facts found in the daily returns of financial returns (Carnero, Peña, and Ruiz, 2004). See Engle (1995) and Shephard (2005) for a detailed account and comparison of the two approaches. To the best of our knowledge, our proposed Gegenbauer long memory stochastic volatility model with leverage and bivariate Student's t error distribution is the first in the literature that generalise many other popular models such as the standard SV model of Taylor (1986), the ARFIMA model of Granger and Joyeux (1980), the SV-L model of Meyer and Yu (2000) and the ARFIMA-SV model of Bos, Koopman, and Ooms (2014b).

Inference using the SV model using the classical approach proved difficult due to intractable likelihood functions which involve high dimensional integrations. Examples of estimation attempts using the classical approach include Melino and Turnbull (1990) using a generalised method of moments to price currency options, Harvey, Ruiz, and Shephard (1994) who applied a quasi maximum likelihood approach in a multivariate context, and also Ait-Sahalia and Kimmel (2007) who used Monte Carlo simulations to estimate short-dated at-the-money options. Over the last twenty years, the Bayesian approach has also become popular due to cheap computing power with many variations. For example, the Metropolis algorithm (Jacquier, Polson, and Rossi, 1994), importance sampling (Shephard and Pitt, 1997), normal mixtures (Kim, Shephard, and Chib, 1998) and more recently Chan and Grant (2016a) who used the Metropolis-Hastings algorithm and the Acceptance-Rejection Metropolis Algorithm instead of Kalman filter-based algorithms.

There are a host of applications which are traditionally used to deduce inference from such models, and the most relevant within recent history is risk control through the use of Value-at-Risk (VaR). For more than twenty years, VaR has been a widely used

technique to measure portfolio tail-risk by financial controllers. The importance of VaR has became more relied upon in the last ten years due to the Global Financial Crisis. In this paper, we also conduct a VaR forecasting exercise to compare the relative performance of our model to alternative popular VaR models including the RiskMetrics approach. Cryptocurrency risk modelling in general is an important aspect not only for modelling cryptocurrencies, but also for practical reasons due to financial institutions increased risk exposure to these new financial assets (Hotz-Behofsits, Huber, and Zörner, 2018; Catania, Grassi, and Ravazzolo, 2018; Hencic and Gouriéroux, 2015).

This chapter proposes for the first time an efficient Bayesian estimation procedure that models Gegenbauer long memory, stochastic volatility, leverage and a bivariate Student's t-distribution. Further, our work sheds light on a large investable sample of cryptocurrencies which have been otherwise neglected. These inferences are extended to real world applications and shed light on the relative merits of particular cryptocurrencies over one another that have otherwise been deemed controversial whether or not they are true.

The remainder of this chapter is organised as follows: the model is introduced in Section 2, and its estimation procedure is detailed in Section 3. Our results are discussed in Sections 4 and 5, and we conclude with Section 6.

5.2 The model

Let $y_t, t = 1, 2, ..., T$ be a stochastic process satisfying the equations

$$\phi(B)(1 - 2uB + B^2)^d y_t = \psi(B)\varepsilon_t^*, \tag{5.1}$$

$$h_{t+1} = \alpha + \beta(h_t - \alpha) + \eta_{t+1}^*, \tag{5.2}$$

$$\begin{pmatrix} \varepsilon_t^* \\ \eta_{t+1}^* \end{pmatrix} \sim t_\nu \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} e^{h_t} & \sigma\rho e^{h_t/2} \\ \sigma\rho e^{h_t/2} & \sigma^2 \end{pmatrix} \right), \tag{5.3}$$

where the autoregressive (AR) and moving average (MA) polynomials are $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$, $\psi(B) = 1 + \psi_1 B + \ldots + \psi_q B^q$ respectively and B is the backshift operator.

We assume that y_t is stationary and invertible such that the zeros of $\phi(z)$ and $\psi(z)$ lie outside the unit circle with no common zeros. It is known that y_t is causal when $(\{|u|< 1, d < 0.5\} \cup \{|u|= 1, d < 0.25\})$, invertible when $(\{|u|< 1, d > -0.5\} \cup \{|u|= 1, d > -0.25\})$ and long memory when $(\{|u|< 1, 0 < d < 0.5\} \cup \{|u|= 1, 0 < d < 0.25\})$. See Dissanayake, Peiris, and Proietti (2016) for details. The class of time series generated by (5.1-5.3) is similar to the GARFIMA(p, q)-SV time series process of Phillip, Chan, and Peiris (2017) but includes additional important features.

We also assume a leverage effect between the errors of the observation equation (5.1) and the latent equation (5.2) such that $\mathbb{E}[\varepsilon_t^* \eta_{t+1}^*] = \rho$. Further, they are assumed to follow a bivariate t-distribution. Clearly, h_t is the log-volatility, which evolves according to the state equation (5.2) for $t = 1, \ldots, T$, α is the constant level of the volatility, β is the persistence of the volatility process and η_{t+1}^* is the volatility of volatility. We assume $|\beta|< 1$ so h_t is stationary.

For simplicity, we discuss the generalised fractional stochastic volatility noise process when $\phi(B) = \psi(B) = 1$ such that $(1 - 2uB + B^2)^d y_t = \varepsilon_t^*$. Under the assumption that y_t is causal, we have the following MA(∞) representation

$$y_t = (1 - 2uB + B^2)^{-d} \varepsilon_t^* = \sum_{j=0}^{\infty} \lambda_j \varepsilon_{t-j}^*, \tag{5.4}$$

where λ_j are the Gegenbauer coefficients, initialized with $\lambda_0 = 1, \lambda_1 = 2ud$ and follow the recursion

$$\lambda_j = 2u \left(\frac{d-1}{j} + 1 \right) \lambda_{j-1} - \left(\frac{2(d-1)}{j} + 1 \right) \lambda_{j-2}, \quad j \geq 2. \tag{5.5}$$

Further details on the Gegenbauer polynomial and its properties can be found in Rainville (1960). A truncated moving average representation of the Wold representation in (5.4) arises from truncating at lag J so that

$$y_t = (1 - 2uB + B^2)^{-d} \varepsilon_t^* \approx \sum_{j=0}^{J_t^*} \lambda_j \varepsilon_{t-j}^*, \tag{5.6}$$

where $J_t^* = \min(t, J)$. For further discussion on the choice of J, see Phillip, Chan, and Peiris (2017). The power spectrum of (5.4), conditional of h_{t+1}, is given by

$$f_{y_t|h_{t+1}}(\omega) = C[4(\cos\ \omega - u)^2]^{-d} \quad -\pi < \omega < \pi,$$

where C is a suitable constant, and the Gegenbauer frequency is $\omega = \cos^{-1}(u)$. This proposed $(0, 0)$ order model is called the GARFIMA-SV leverage heavy common tail (GMA-SV-LVG-HC) model.

5.3 Bayesian inference

In order to estimate models in 5.1-6.3, we consider the fractional noise case in order not to detract from the main features of our model, noting that alternative mean structures such as ARMA(p,q) can easily be implemented. We present two important modifications in order to operationalize the model, and importantly note they are order invariant.

Firstly, Andrews and Mallows (1974) introduced Scale Mixtures of Normal (SMN) distributions, and the Student-t distribution as a SMN. Let $\boldsymbol{X}$ be a vector of continuous random variables with location $\boldsymbol{\mu}$ and scale matrix $\boldsymbol{\Sigma}$. If $\boldsymbol{X}$ can be represented as

$$f(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_0^\infty N(\boldsymbol{x}|\boldsymbol{\mu}, \kappa(\xi)\boldsymbol{\Sigma})\pi(\xi)d\xi,$$

where $N(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a multivariate normal pdf, $\kappa(\xi)$ is a positive function of ξ and $\pi(\cdot)$ is a pdf defined on $\Re^+$, then we say the pdf of $\boldsymbol{X}$ has a SMN representation. The quantity ξ is known as the mixing parameter, and $\pi(\cdot)$ is the mixing density. For a multivariate Student-t distribution with location $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν, $\kappa(\xi) = \xi$ and $\pi(\xi)$ is the pdf of the inverse gamma $IG(\frac{\nu}{2}, \frac{\nu}{2})$ distribution where

$$IG(\xi|a, b) = \frac{b^a}{\Gamma(a)}\xi^{-(a+1)}e^{-b/\xi}, \quad \xi, a, b > 0.$$

Then the PDF of the Student-t distribution can be expressed as

$$t(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \int_0^\infty N(\boldsymbol{x}|\boldsymbol{\mu}, \xi\boldsymbol{\Sigma})IG\left(\xi|\nu/2, \nu/2\right)d\xi$$

or hierarchically

$$\begin{aligned}\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \xi &\sim N(\boldsymbol{\mu}, \xi\boldsymbol{\Sigma}),\\ \xi|\nu &\sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right).\end{aligned}$$

We use such a SMN to redefine the bivariate t-error distribution, so that 6.3 can be rewritten as

$$\begin{pmatrix} \varepsilon_t^* \\ \eta_{t+1}^* \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \xi_{t+1} \begin{pmatrix} e^{h_t} & \sigma\rho e^{h_t/2} \\ \sigma\rho e^{h_t/2} & \sigma^2 \end{pmatrix} \right), \tag{5.7}$$

where $\xi_{t+1} \sim \mathrm{IG}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$.

Secondly, we derive the uncorrelated marginal distributions. This latter technique is commonly used in the financial mathematics literature, and was popularized into the econometrics literature by Meyer and Yu (2000). To see this, first recall that leverage is the negative relationship between y_t and h_{t+1}.Therefore by conditioning ϵ_t^* the bivariate normal distribution in Equation (5.7) can be expressed as a marginal and conditional

$$\varepsilon_t^* | \eta_{t+1}^* \sim \mathrm{N}\left(\frac{\rho}{\sigma} e^{h_t/2}(h_{t+1} - \alpha - \beta(h_t - \alpha)), \xi_{t+1} e^{h_t}(1-\rho^2)\right), \quad \eta_{t+1}^* \sim N(0, \xi_{t+1}\sigma^2).$$

Hence, y_t in Equation (5.6) can be expressed as

$$y_t | h_{t+1}, h_t = \sum_{j=0}^{J_t^*} \lambda_j e^{h_{t-j}/2} \frac{\rho}{\sigma}(h_{t+1-j} - \alpha - \beta(h_{t-j} - \alpha)) + \sum_{j=0}^{J_t^*} \lambda_j e^{h_{t-j}/2} \xi_{t+1-j}^{\frac{1}{2}} \sqrt{1-\rho^2}\varepsilon_{t-j}, \tag{5.8}$$

$$h_{t+1} = \alpha + \beta(h_t - \alpha) + \xi_{t+1}^{\frac{1}{2}} \sigma \eta_{t+1}, \tag{5.9}$$

where ε_t and η_{t+1} are the uncorrelated $\mathrm{N}(0,1)$ errors.

Let $\boldsymbol{Y} = [y_1, \ldots, y_T], \boldsymbol{h} = [h_1, \ldots, h_{T+1}]$ such that $\boldsymbol{Y}|\boldsymbol{h}, \boldsymbol{G}_J \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ where $\boldsymbol{\mu} = \frac{\rho}{\sigma}\boldsymbol{G}_J(\boldsymbol{W} \circ \boldsymbol{E})$, $\boldsymbol{W} = (e^{h_1/2}, \ldots, e^{h_T/2})$, $\boldsymbol{E} = [h_2 - \alpha - \beta(h_1 - \alpha), \ldots, h_{T+1} - \alpha - \beta(h_T - \alpha)]$ and $A \circ B$ refers to the Hadamard product of vectors A and B. The covariance matrix can be expressed as $\boldsymbol{\Gamma} = (1-\rho^2)\boldsymbol{G}_J \boldsymbol{V} \boldsymbol{G}_J'$ where $\boldsymbol{V} = \mathrm{diag}(\boldsymbol{W} \circ \boldsymbol{W} \circ \boldsymbol{\xi})$, $\boldsymbol{\xi} = (\xi_2, \ldots, \xi_{T+1})$ and $\boldsymbol{G}_J$ is a $T \times T$ lower triangular banded matrix with J Gegenbauer truncated moving average parameters in each column, and ones on the diagonal as

given below

$$\boldsymbol{G}_J = \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & \dots & \ddots & \dots & 0 \\ \lambda_1 & 1 & \dots & \dots & \dots & \dots & \ddots & \dots & 0 \\ \lambda_2 & \lambda_1 & \dots & \dots & \dots & \dots & \ddots & \dots & 0 \\ \vdots & \lambda_2 & \dots & \dots & \dots & \dots & \ddots & \dots & \vdots \\ \lambda_J & \vdots & \dots & \dots & \dots & \dots & \ddots & \dots & \vdots \\ 0 & \lambda_J & \dots & \dots & \dots & \dots & 0 & 0 & 0 \\ \vdots & 0 & \dots & \dots & \dots & \dots & 1 & 0 & 0 \\ \vdots & \vdots & \dots & \dots & \dots & \dots & \lambda_1 & 1 & 0 \\ 0 & 0 & \dots & \dots & \dots & \dots & \lambda_2 & \lambda_1 & 1 \end{bmatrix}.$$

Note $|\boldsymbol{G}_J| = 1$ such that $|\boldsymbol{\Gamma}| = (1-\rho^2)^T \exp(\sum_{t=1}^T h_t) \prod_{t=1}^T \xi_{t+1}$. Therefore, the observational log-likelihood is

$$\log f(\boldsymbol{Y}|\boldsymbol{h}, \boldsymbol{G}_J) = -\frac{T}{2}\log(2\pi(1-\rho^2)) - \frac{1}{2}\sum_{t=1}^{T}(h_t + \log \xi_{t+1}) - \frac{1}{2}(\boldsymbol{Y}-\boldsymbol{\mu})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}) \tag{5.10}$$

where the vector of all model parameters are $\boldsymbol{\theta} = (u, d, \boldsymbol{h}, \alpha, \beta, \sigma^2, \xi_1, \boldsymbol{\xi})$. The Bayesian analysis of each individual parameter can be found in Appendix F.

5.4 Simulation studies

We now outline a comprehensive simulation study in order to assess the performance of our proposed sampling scheme. The purpose of this section is to illustrate the proposed methodology outlined in Section 3, and explore the model performance for unknown $\boldsymbol{\theta}$.

5.4.1 Parameter estimation

Data is generated according to Equations (5.1-6.3) and the parameters are estimated subsequently. We use a value of $J = 1,000$ in order to simulate a time series as close as possible to an infinite moving average representation. We consider $\nu = [5, 6, 7, \dots, 15, 20]$ and $\rho = [-0.6, -0.3]$. The Gegenbauer parameters are set to $u =$

0.8 and $d = 0.4$ and the stochastic residuals are simulated according to the parameters $\alpha = 0, \beta = 0.97$ and $\sigma = 0.025$. Our simulated process has the expression

$$(1 - 1.6B + B^2)^{0.4} y_t = \varepsilon_t^*, \tag{5.11}$$

$$h_{t+1} = 0.97 h_t + \eta_{t+1}^*, \tag{5.12}$$

$$\begin{pmatrix} \varepsilon_t^* \\ \eta_{t+1}^* \end{pmatrix} \sim t_\nu \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} e^{h_t} & 0.025\rho e^{h_t/2} \\ 0.025\rho e^{h_t/2} & 0.025^2 \end{pmatrix} \right). \tag{5.13}$$

We simulate the process to be $T = 1,500$ to ensure the Gegenbauer parameters are estimated accurately (see Phillip, Chan, and Peiris (2017) for details).

The prior choices for the SV parameters are generally not influential since the likelihood holds most of the information, which is especially relevant since we assume $T = 1,500$. Our initial starting values are chosen arbitrarily and are purposely chosen to be far away from the true values. We also test with several fixed starting values to ensure that different MCMC chains converge within similar value ranges.

The hyperparameters are set as follows

$$u \sim N(0, 0.1)[-1, 1], d \sim N(0.125, 0.05)[0, 0.25]\mathbb{1}_{ud} + N(0.25, 0.0.05)[0, 0.5](1 - \mathbb{1}_{ud})$$

$$\nu \sim U[3, 23], \rho \sim N(-0.1, 0.05), \alpha \sim N(0, 0.05), \beta \sim N(0.99, 0.2).$$

The process in equations 5.11-5.13 is simulated $\Omega = 1,000$ times, and the parameters are estimated each time using the model. We report the estimated mean of each parameter, the root mean squared error (RMSE) and the mean of the standard errors in parentheses. We use $M = 10,000$ iterates after a burn-in period of $10,000$. We purposely choose a burn-in period of half the total number of iterations under the advice of Gelman et al. (2013).

As shown below in Tables 5.1 and 5.2, the estimates of $\hat{\nu}$ are more accurate with lower RMSE when the true value of ν is low. Conversley, as the true value of ν increases such that the error distribution approaches a Gaussian distribution, the error becomes larger. This is of course an anticipated result since the percentile difference between these different Student-t distributions becomes smaller as ν increases. Hence, the upward bias of $\hat{\nu}$ also increases with the true value ν. The leverage and long memory

parameters are also estimated with high accuracy. The constant term α of the SV model is estimated accurately, while the persistence parameter β and the volatility of volatility term σ^2 are both estimated well. The acceptance rate of $\boldsymbol{h}$ increases as the value of ν increases. This is due to the fact that draws of $\boldsymbol{h}$ from its proposal density are more likely to be accepted for a Gaussian distribution as opposed to a Student-t distribution. The acceptance rates of $\hat{u}$ and $\hat{d}$ are both close to the optimal acceptance rate which are set during tuning, and do not vary for different values of ν. The acceptance rate of $\hat{\beta}$ is high, which is the typical case since it is close to the boundary.

The average running time of the proposed MCMC scheme is also reported; in general it takes more time to estimate data with lower values of ν. This is because the sampler spends more time sampling $\boldsymbol{h}^*$ in the acceptance-rejection step due to outlier values associated with lower ν values. In rare circumstances, these outlier values are so extreme the sampler becomes "stuck" during this step, and as such the sampler will take an extremely long time to sample $\boldsymbol{h}^*$ (sometimes more than 30 hours). These rare outlier runs are purposely included in order to make the estimation as fair as possible, and not to bias the data by only including MCMC runs which were executed below a certain time.

ν_1	$\hat{\nu}_1$	RMSE($\hat{\nu}_1$)	$\hat{\rho}$	RMSE($\hat{\rho}$)	$\hat{u}$	RMSE($\hat{u}$)	$\hat{d}$	RMSE($\hat{d}$)	$\hat{\alpha}$	RMSE($\hat{\alpha}$)	$\hat{\beta}$	RMSE($\hat{\beta}$)	$\hat{\sigma}^2$	RMSE($\hat{\sigma}^2$)	$\widehat{AR}_h\%$	$\widehat{AR}_u\%$	$\widehat{AR}_d\%$	$\widehat{AR}_\beta\%$	$\widehat{AR}_\rho\%$	Time
5.0	5.203	0.269	-0.299	0.001	0.800	0.000	0.401	0.001	0.001	0.001	0.966	0.019	0.026	0.006	40.066	29.742	25.008	95.404	29.856	7h 18m
		(0.942)		(0.020)		(0.003)		(0.012)		(0.047)		(0.011)		(0.010)						
6.0	6.278	0.573	-0.300	0.000	0.800	0.000	0.400	0.000	0.002	0.002	0.965	0.020	0.027	0.007	42.599	29.454	24.968	95.159	29.211	6h 11m
		(1.620)		(0.021)		(0.003)		(0.013)		(0.046)		(0.011)		(0.010)						
7.0	7.413	0.857	-0.300	0.000	0.800	0.000	0.400	0.000	0.003	0.003	0.966	0.019	0.028	0.008	44.196	29.241	24.951	95.119	30.363	5h 37m
		(2.221)		(0.021)		(0.003)		(0.013)		(0.046)		(0.011)		(0.010)						
8.0	8.624	1.447	-0.300	0.000	0.800	0.000	0.400	0.000	0.003	0.003	0.966	0.019	0.026	0.006	45.593	29.210	25.644	94.867	31.605	7h 23m
		(2.846)		(0.022)		(0.003)		(0.013)		(0.046)		(0.011)		(0.010)						
9.0	9.518	1.412	-0.300	0.000	0.800	0.000	0.400	0.000	0.001	0.001	0.965	0.020	0.026	0.006	46.621	29.090	25.502	94.890	32.383	6h 12m
		(3.138)		(0.021)		(0.003)		(0.013)		(0.046)		(0.011)		(0.010)						
10.0	10.647	1.663	-0.300	0.000	0.800	0.000	0.400	0.000	0.002	0.002	0.966	0.019	0.026	0.006	47.380	29.135	25.535	94.790	33.043	5h 28m
		(3.378)		(0.022)		(0.004)		(0.013)		(0.046)		(0.012)		(0.010)						
11.0	11.783	1.947	-0.300	0.000	0.800	0.000	0.400	0.000	0.002	0.002	0.967	0.018	0.024	0.004	48.997	29.158	26.074	94.620	33.914	8h 18m
		(3.477)		(0.022)		(0.004)		(0.013)		(0.046)		(0.011)		(0.009)						
12.0	12.946	2.074	-0.300	0.000	0.800	0.000	0.401	0.001	0.000	0.000	0.966	0.019	0.025	0.005	48.637	29.404	25.690	94.688	34.650	6h 16m
		(3.517)		(0.023)		(0.004)		(0.013)		(0.046)		(0.011)		(0.009)						
13.0	14.288	2.172	-0.301	0.001	0.800	0.000	0.400	0.000	0.001	0.001	0.967	0.018	0.023	0.003	49.539	29.255	25.776	94.715	35.239	6h 24m
		(3.548)		(0.024)		(0.004)		(0.013)		(0.046)		(0.011)		(0.009)						
14.0	15.343	1.762	-0.300	0.000	0.800	0.000	0.400	0.000	-0.001	0.001	0.967	0.018	0.024	0.004	49.794	28.969	26.004	94.866	35.640	5h 38m
		(3.506)		(0.022)		(0.004)		(0.014)		(0.046)		(0.011)		(0.009)						
15.0	16.916	1.742	-0.300	0.000	0.800	0.000	0.400	0.000	0.000	0.000	0.967	0.018	0.023	0.003	50.930	29.442	25.764	94.588	35.992	4h 42m
		(3.486)		(0.022)		(0.004)		(0.014)		(0.046)		(0.011)		(0.009)						
20.0	20.694	1.128	-0.300	0.000	0.800	0.000	0.402	0.002	-0.003	0.003	0.968	0.017	0.021	0.001	53.112	29.624	26.204	94.622	36.197	6h 16m
		(3.314)		(0.024)		(0.004)		(0.014)		(0.046)		(0.011)		(0.008)						

TABLE 5.1: Simulation study results when true value of $\rho = -0.3$.

ν_1	$\hat{\nu}_1$	RMSE($\hat{\nu}_1$)	$\hat{\rho}$	RMSE($\hat{\rho}$)	$\hat{u}$	RMSE($\hat{u}$)	$\hat{d}$	RMSE($\hat{d}$)	$\hat{\alpha}$	RMSE($\hat{\alpha}$)	$\hat{\beta}$	RMSE($\hat{\beta}$)	$\hat{\sigma}^2$	RMSE($\hat{\sigma}^2$)	$\widehat{AR}_h\%$	$\widehat{AR}_u\%$	$\widehat{AR}_d\%$	$\widehat{AR}_\beta\%$	$\widehat{AR}_\nu\%$	Time
5.0	5.042	0.184	-0.599	0.001	0.800	0.000	0.400	0.000	0.001	0.001	0.970	0.015	0.023	0.003	48.256	31.437	26.513	95.745	29.149	6h 31m
		(0.845)		(0.020)		(0.003)		(0.012)		(0.046)		(0.008)		(0.007)						
6.0	6.237	0.423	-0.599	0.001	0.800	0.000	0.401	0.001	0.001	0.001	0.970	0.015	0.023	0.003	50.343	30.673	25.975	95.505	29.061	5h 44m
		(1.417)		(0.021)		(0.003)		(0.012)		(0.046)		(0.008)		(0.007)						
7.0	7.251	0.601	-0.599	0.001	0.800	0.000	0.401	0.001	0.002	0.002	0.971	0.014	0.023	0.003	52.451	30.617	25.393	95.555	30.680	5h 27m
		(1.977)		(0.022)		(0.003)		(0.013)		(0.046)		(0.008)		(0.007)						
8.0	8.208	0.814	-0.600	0.000	0.800	0.000	0.400	0.000	0.002	0.002	0.970	0.015	0.023	0.003	53.614	29.853	25.137	95.363	31.765	5h 33m
		(2.556)		(0.021)		(0.003)		(0.013)		(0.046)		(0.009)		(0.008)						
9.0	9.254	1.045	-0.600	0.000	0.800	0.000	0.400	0.000	0.002	0.002	0.970	0.015	0.022	0.002	54.242	30.048	25.120	95.310	32.330	5h 43m
		(2.974)		(0.022)		(0.003)		(0.013)		(0.046)		(0.008)		(0.008)						
10.0	10.450	1.408	-0.600	0.000	0.800	0.000	0.400	0.000	0.003	0.003	0.970	0.015	0.022	0.002	55.289	29.861	25.004	95.471	33.096	5h 35m
		(3.246)		(0.024)		(0.003)		(0.013)		(0.046)		(0.009)		(0.008)						
11.0	10.945	1.070	-0.600	0.000	0.800	0.000	0.400	0.000	0.002	0.002	0.970	0.015	0.023	0.003	55.665	30.201	25.167	95.169	33.496	6h 25m
		(3.337)		(0.023)		(0.004)		(0.013)		(0.046)		(0.009)		(0.008)						
12.0	12.483	1.489	-0.602	0.002	0.800	0.000	0.400	0.000	0.000	0.000	0.969	0.016	0.025	0.005	55.832	29.499	25.325	95.551	34.271	4h 38m
		(3.460)		(0.025)		(0.004)		(0.013)		(0.046)		(0.009)		(0.008)						
13.0	13.608	1.618	-0.602	0.002	0.800	0.000	0.401	0.001	0.001	0.001	0.970	0.015	0.022	0.002	56.878	29.448	25.306	95.125	35.034	6h 35m
		(3.499)		(0.026)		(0.004)		(0.013)		(0.046)		(0.009)		(0.008)						
14.0	14.586	1.579	-0.601	0.001	0.800	0.000	0.401	0.001	-0.000	0.000	0.970	0.015	0.024	0.004	56.914	29.655	24.818	95.245	35.377	5h 48m
		(3.495)		(0.025)		(0.004)		(0.013)		(0.045)		(0.009)		(0.008)						
15.0	15.955	1.300	-0.602	0.002	0.800	0.000	0.400	0.000	-0.001	0.001	0.969	0.016	0.023	0.003	57.204	29.860	25.012	95.308	35.676	5h 34m
		(3.456)		(0.025)		(0.004)		(0.014)		(0.045)		(0.009)		(0.008)						
20.0	20.466	1.239	-0.602	0.002	0.800	0.000	0.401	0.001	-0.003	0.003	0.970	0.015	0.023	0.003	58.826	29.579	25.650	95.428	35.969	2h 7m
		(3.320)		(0.025)		(0.004)		(0.014)		(0.045)		(0.009)		(0.008)						

TABLE 5.2: Simulation study results when true value of $\rho = -0.6$.

5.4.2 Methodology comparison

In Table 5.3, we compare our procedure with the methodologies of Wang, Chan, and Choy (2011) and Choy and Chan (2000). The authors propose model (5.1-5.3) without long memory effects, which is of course a special case of our model when $d = 0$. The authors in Wang, Chan, and Choy (2011) derived an alternative representation of the model by first deriving the conditional Student-t distribution for the observation equation and a marginal Student-t distribution for the latent process and then expressed these two distributions as a SMN. They also compared their methodology with that of Choy and Chan (2000) who modeled the bivariate Student-t distribution as a scale mixture of bivariate normal distributions in which we also adopt in this paper. We try to replicate their priors and parameters as much as possible so that all three methodologies are directly comparable. First, Wang, Chan, and Choy (2011) assigned a vague prior to α, so we assign our Normal prior to be $N(0, 10)$. The prior of $\beta^* = \frac{\beta+1}{2}$ is assumed to be Beta$(20, 1.5)$ so we assign our prior to be $N(0.95, 0.05)$ whereby the mean and variances are both close to the mean and variance of 0.86 and 0.012 approximately in Beta$(20, 1.5)$. A non-informative prior is used for σ^2, and this prior is also assumed in our proposed methodology. A uniform prior $U(-1, 1)$ is assigned to ρ, so we set the truncated Normal prior of ρ to $N(0, 10)\mathbb{I}(-1, 1)$. In both cases, ν is set to have a non-informative prior restricted within the domain $[1, 40]$. The authors considered synthetic time series of length $T = 500$ and replication of 100 times. We also follow these choices. The authors use a burn-in period of $50,000$ loops

and follow up with an additional $250,000$ MCMC loops. We however use a much smaller number of $10,000$ loops, follow up with an additional $10,000$ loops and find that in fact all parameters have converged. We use the same true parameter values as both papers: $\alpha = -10$, $\beta = 0.8$, $\sigma^2 = 0.2$, $\rho = 0.8$ and $\nu = 5$ and report the results below in Table 5.3.

In essence, it is clear from Table 5.3 that our method dominates for all parameters. In all cases, the standard deviation of our parameter estimate, the Percentage Error (PE) and Mean Squared Error (MSE) are consistently smaller. The 95% Confidence Interval bands are also tighter in all cases. The authors were able to estimate α, β, σ^2 accurately, and so is our model. The estimation of the shape parameter ν is noticeably different and our estimate is closer to the true value. Most notably, the authors were unable to correctly estimate ρ, but our model is able to estimate ρ to be very close to the true parameter value.

5.5 Empirical Data Analysis

In this section a set of empirical data is considered to illustrate the proposed models. We first provide a general broad scope on an investable basket of 114 cryptocurrencies by discussing model parameter estimates. Then we focus on five of the most popular cryptocurrencies, namely Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), NEM, and Dash, with model comparison, and follow up with a specific analysis on Bitcoin itself.

The data is sourced from the Brave New Coin (BNC) Digital Currency index which represents the cleanest and most comprehensive cryptocurrency database in the world. BNC surveys hundreds of trading platforms for crypto/fiat trading pairs and currently records $2,796$ cryptocurrency time-series. To date, there are many more, with thousands of cryptocurrencies being traded. However, some of these have market capitalizations which are small ($< \$1,000,000$) and traded very infrequently. Of the 2,796 data sets available on the BNC database, only 114 of these have been exchanged at least once per day since inception. Although cryptocurrencies were first introduced in 2008, BNC only reports price points when more formalized exchanges for each respective currency could be measured with reliability. As such, the number of observations recorded for each time series vary, but all end on or before the 30th of April,

Par.	Approach	True	Estimate	sd	95% CI	PE (%)	MSE	CP (%)
ν	Phillip et. al.'s	5	5.2508	1.9158	(4.2266,7.4366)	5.0168	0.5205	96
	Wange et. al.'s	5	5.7937	2.4708	(3.7092, 13.1320)	15.87	2.3335	98
	Choy and Chan's	5	5.7159	3.2069	(3.3416, 15.2972)	14.32	7.1341	92
α	Phillip et. al.'s	-10	-9.9921	0.0234	(-10.0744,-9.9097)	0.0807	0.0010	97
	Wang et. al.'s	-10	-9.9758	0.1133	(-10.1860, -9.7506)	0.24	0.0111	96
	Choy and Chan's	-10	-9.9999	0.1600	(-10.2836, -9.6827)	0.00	0.0684	96
β	Phillip et. al.'s	0.8	0.7970	0.0312	(0.7319,0.8479)	-0.3716	0.001	94
	Wang et. al.'s	0.8	0.8107	0.0747	(0.6205, 0.9057)	1.35	0.0039	99
	Choy and Chan's	0.8	0.8104	0.0831	(0.6064, 0.9209)	1.30	0.0035	98
σ^2	Phillip et. al.'s	0.2	0.2044	0.0460	(0.1423,0.2992)	2.2004	0.0021	94
	Wang et. al.'s	0.2	0.2085	0.0735	(0.1209, 0.3988)	4.24	0.0045	97
	Choy and Chan's	0.2	0.1913	0.0833	(0.0885, 0.4018)	-4.34	0.0048	96
ρ	Phillip et. al.'s	0.8	0.7993	0.0217	(0.7733,0.8194)	-0.0903	0.0034	96
	Wang et. al.'s	0.8	0.7136	0.1538	(0.3421, 0.9044)	-10.80	0.0169	95
	Choy and Chan's	0.8	0.7146	0.1810	(0.2977, 0.9279)	-10.67	0.0554	83

TABLE 5.3: Table comparing the parameter estimates of the modelling procedures of Wang, Chan, and Choy (2011) and Choy and Chan (2000) to our proposed methodology. The 95% credible intervals (CI) are presented, as well as the percentage error (PE), the mean squared error (MSE) and the coverage percentage (CP).

2018. The time series y_t is defined as the global weighted daily price percentage change $y_t = (P_t - P_{t-1})/P_{t-1}$, where P_t is the index price at time t.

5.5.1 In-sample fitting

All 114 Cryptocurrencies

We first fit our proposed model in (5.1-6.3) to the 114 cryptocurrencies discussed above. This extensive preliminary study aims to extract the basic properties of cryptocurrencies in general. In order to do so, each return series is estimated using the GMA-SV-LVG-HC model and parameter estimates are presented in Figure 5.1.

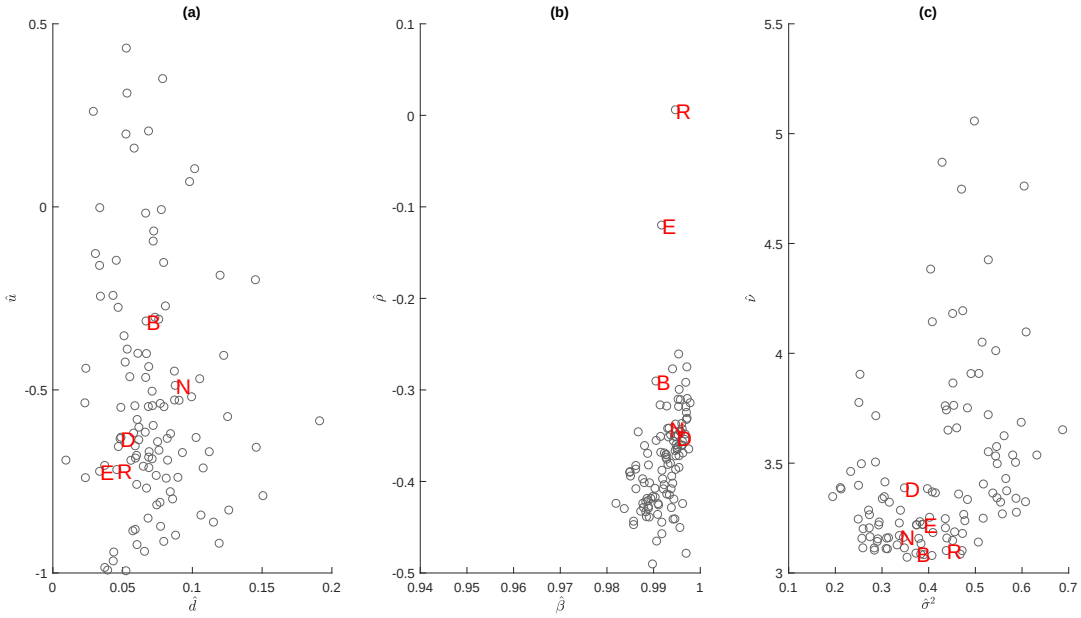

FIGURE 5.1: Scatter plots of parameter estimates: (a) $[\hat{u}, \hat{d}]$; (b) $[\hat{\rho}, \hat{\beta}]$; (c) $[\hat{\nu}, \hat{\sigma}^2]$, of 114 different cryptocurrency data sets under the GLM-SV-LVG-HC model. B: Bitcoin, E: Ethereum, R: Ripple, N: NEM, D: Dash.

It is clear from Figure 5.1 (a) that cryptocurrencies in general show signs of generalised long memory, and not the plain long memory case of Hosking (1981). The price persistence estimates $\hat{d}$ are generally spread across from 0.02 to 0.15 with the five popular cryptocurrencies exhibiting a low levels of long memory. This indicates these five cryptocurrencies have higher market efficiency in general. The periodicity estimates $\hat{u}$ are generally negative, which means their autocorrelation functions are typically instantaneously oscillating.

Conventionally, the leverage effect, if in existence, is known to be negative for most financial time series; which indicates that one day ahead volatility and returns are typically negatively correlated. The leverage effects $\hat{\rho}$ of our cryptocurrency universe are also negative and tend to cluster around -0.35. Also, when stochastic volatility does exist, the volatility persistence parameter estimate $\hat{\beta}$ for most financial time series is close to one (Kim, Shephard, and Chib, 1998). As evidenced, this too is the case with cryptocurrencies with most estimates of $\hat{\beta}$ clustered closely around one, including the popular five. Therefore, cryptocurrencies also commonly share these familiar and widely accepted behaviors.

Of most interest to those vested in cryptocurrencies are their variability characteristics which are summarized in Figure 5.1 (c). Interestingly, we see two main regimes which depict the variability characteristics of cryptocurrencies. We see that typically, cryptocurrencies have heavy tails as evidenced by the clustering of $\hat{\nu} < 3.5$. Doubly so, they also display unconventionally large estimates of $\hat{\sigma}^2$. As widely speculated in the media, cryptocurrencies in fact show evidence of heavier tails, and larger volatility of volatility estimates than fiat assets. This is a testimony to the ability of the GMA-SV-LVG-HC model to conditionally measure volatility of volatility compared to the degrees of freedom parameter and is able to separate the two effects.

Five of the most popular Cryptocurrencies

In this subsection, the focus is on five of the most popular cryptocurrencies as per Table 5.4, which are plotted in Figure 5.2.

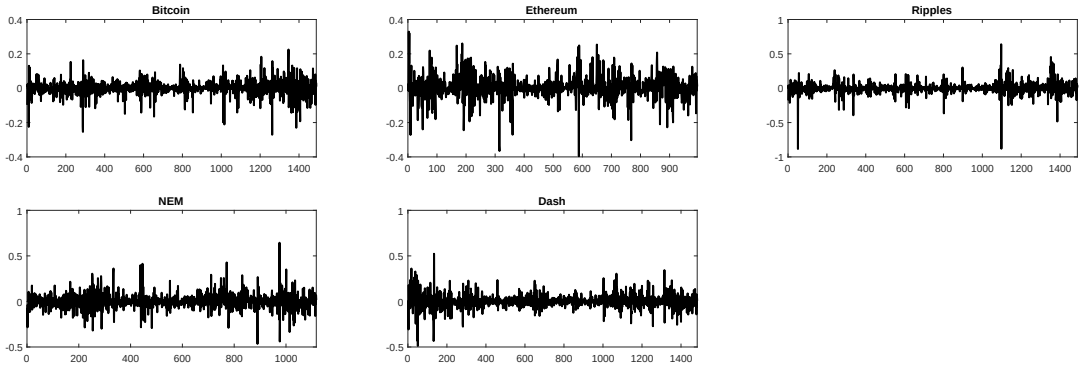

FIGURE 5.2: Plots of daily returns for the top five cryptocurrencies by market capitalization.

The descriptive statistics of each time-series transformed by taking the daily price percentage change are shown in Table 5.4. Similar to their fiat counterparts, we see that currencies with lower market capitalizations exhibit larger volatility characteristics. Remarkably, the volatility of these currencies are as large as 8.8% which is extremely different to fiat currencies. The Ljung-Box tests on $|y_t|$ and y_t^2 show strong evidence of long memory and error dependence respectively. The Kolomogorov-Smirnoff test for normality is also rejected.

	Rank	Market Cap. ($B)	No. of obs	Mean	Std.	Skewness	Kurtosis	Min.	Max.	L-B($\|y_t\|$)	L-B(y_t^2)	Normality test
Bitcoin	1	137.0078	1489	0.0012	0.0414	-0.8543	10.2330	-0.2709	0.2250	762.3674	255.8224	3426.8910
										($< .0001$)	($< .0001$)	($< .0001$)
Ethereum	2	67.0341	995	0.0043	0.0715	-0.2145	7.3622	-0.3959	0.3293	281.7172	141.7939	796.5499
										($< .0001$)	($< .0001$)	($< .0001$)
Ripple	3	26.0533	1489	0.0000	0.0781	-1.3992	31.8162	-0.8844	0.6393	559.8652	103.5672	52003.6272
										($< .0001$)	($< .0001$)	($< .0001$)
NEM	12	3.0601	1117	0.0026	0.0877	0.4502	9.3332	-0.4656	0.6443	291.9821	136.7968	1904.4886
										($< .0001$)	($< .0001$)	($< .0001$)
Dash	14	2.6920	1483	0.0018	0.0727	0.0821	10.6803	-0.4881	0.5232	851.7943	591.9198	3646.5302
										($< .0001$)	($< .0001$)	($< .0001$)

TABLE 5.4: Summary statistics of the global weighted average indices for each relevant Cryptocurrency. P-values of the relevant columns are reported in parantheses. L-B: Ljung-Box Q-test for residual autocorrelation. We note they all end on the 30^{th} of April, 2018 but the start date varies, as per the number of observations listed.

These five cryptocurrencies are also good examples to illustrate the properties of cryptocurrencies from a technological aspect. One of the most debated topics circulating the Cryptocurrency community is whether or not there are added benefits of one cryptocurrency over the other. One of these particular controversies is the so called *longer confirmation time problem*. Briefly, the biggest criticism of Bitcoin is that transaction can be an extremely slow process, sometimes taking up to 48 hours to be sent from one user to another. Two particular cryptocurrencies which attempt to overcome this issue are ETH and Dash. There is a larger community based approach with computer programmers actively making both cryptocurrencies quicker to transact. ETH uses so called smart contract to use blockchains and Dash uses instant transactions via the technology *InstantSend*. InstantSend is a feature of Dash which allows for almost near-instant transactions, and hence solves the longer confirmation time problem of other cryptocurrencies such as Bitcoin. The smart contract technology is widely contested as being the best solution, and the fastest way to transact. As a result of these quicker transaction times, it can be interpreted that ETH and Dash should have lower liquidity risk than Bitcoin. These differences have interesting statistical implication

as will be discussed.

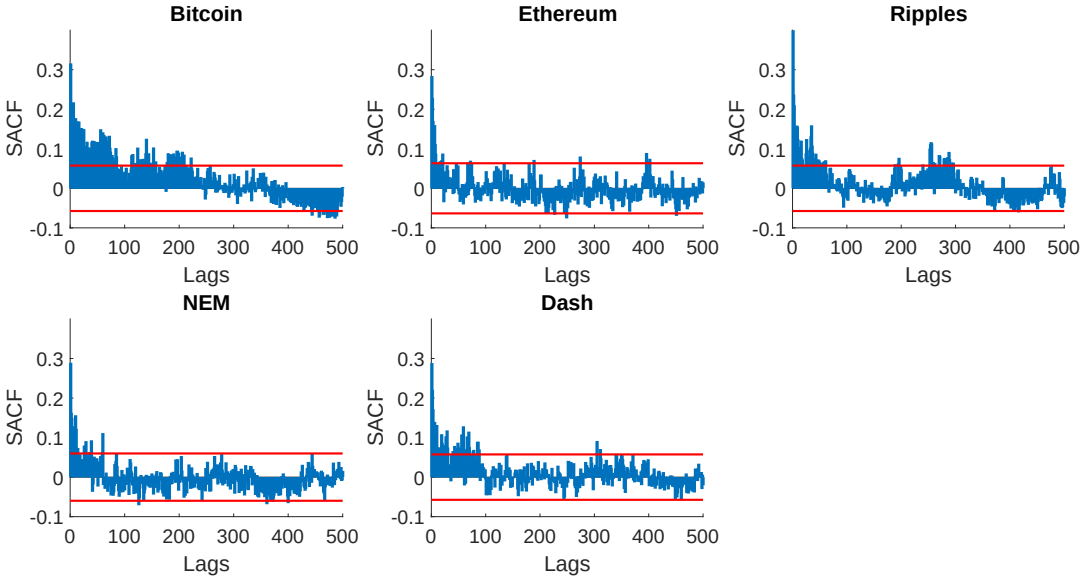

FIGURE 5.3: ACFs of absolute returns of the top five cryptocurrencies by market capitalization as of the 31st of July, 2017.

Figure 5.3 shows the sample autocorrelation plots of the absolute returns for each respective cryptocurrency. It is clear that evidence of long memory behavior exists due to the persistent autocorrelations. Interestingly, we see that Bitcoin displays the most recurrent behavior as evidenced by its strong cyclicality.

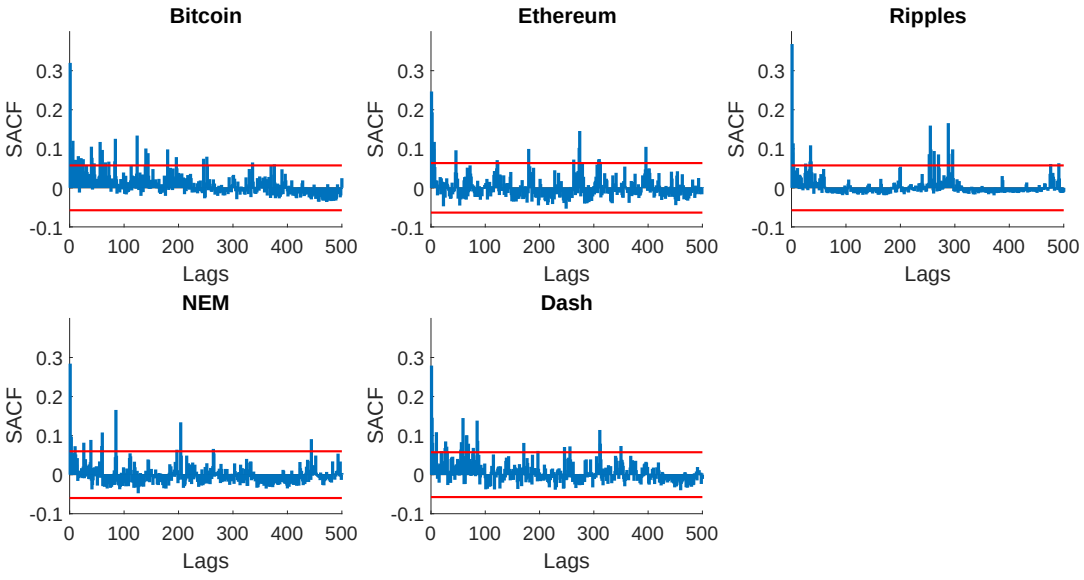

FIGURE 5.4: ACFs of squared returns of the top five cryptocurrencies by market capitalization as of the 31st of July, 2017.

The autocorrelations of the squared returns as depicted in Figure 5.4 also displays interesting behaviors of cryptocurrencies. We see that BTC, ETH and NEM display

properties of dependent errors as typically seen in the literature. Notably, we see that Ripple has the shortest dependence relative to the other four cryptocurrencies, whereas Dash has the most persistence. Interestingly, Ripple is not dependent on any third party for redemption, and as such, it is the only currency with no counter-party credit risk. In other words, there is virtually no risk between transactions performed today and that of any risks (volatility) observed the next day. This is indeed the case as per Figure 5.1 which shows the estimate of ρ of Ripple to be around zero. It has the advantage of users being able to store any fiat/cryptocurrency asset on the network, and as such is insulated from future exchange rate volatility. Due to this safety feature, Ripple has been increasingly used by banks and large corporations as their preferred settlement infrastructure technology. This is in contrast to Dash which is the only currency that uses instant transactions (*InstantSend)*. As it seems apparent due to this ability to transact faster, we see higher dependence amongst its squared returns.

In order to further inspect the properties of the cryptocurrencies listed in Table 5.4, each of the five most popular cryptocurrency data sets are fitted according to model (5.1-5.3) called the GMA-SV-LVG-HC, and five further nested sub-models including the Stochastic Voltility (SV) model of Taylor (1986) and the leverage (LVG) model. Details of these nested sub-models are outlined in the appendix as model 1-6. We particularly note the sub-models GMA-SV-LVG and GMA-SV-HC models are nested variations of our model which have not been studied in the literature previously. The deviance information criterion (DIC) of Spiegelhalter et al. (2002) is used as the Bayesian model selection criterion, which is defined as: DIC $= -2\log p(y|\hat{\theta}) + 2p_E$, where $p_E = 2[\log p(\boldsymbol{y}|\hat{\boldsymbol{\theta}}) - \mathbb{E}[\log p(\boldsymbol{y}|\boldsymbol{\theta})]]$ is the effective number of parameters estimated as the difference between the posterior mean of deviance and the deviance evaluated at the posterior means of each parameter. Alternative model criterion such as the predictive log score of Catania and Grassi, 2017 may also be considered.

Currency	SV	SV-LVG	GMA-SV	GMA-SV-LVG	GMA-SV-HC	GMA-SV-LVG-HC
Bitcoin	0.9432	0.8244	0.9251	0.8497	0.9277	1.0000
Ethereum	0.9496	0.6829	0.9527	0.7947	0.9710	1.0000
Ripples	0.5826	0.6681	0.5835	0.7670	0.8249	1.0000
NEM	0.5497	0.5801	0.5350	0.5106	0.8414	1.0000
Dash	0.8657	0.7894	0.8592	0.8237	0.9837	1.0000

TABLE 5.5: Ratio of DICs. Since they are all negative values, larger is better.

For each model, the DIC is calculated and reported relative to the full GMA-SV-LVG-HC model in Table 5.5. Since all of the DICs are negative, the model with larger ratio is better. This table can be interpreted in conjunction with Figure 5.1 to provide a richer understanding on the behavior of cryptocurrencies. These results show the GMA-SV-LVG-HC is the superior choice for all five cryptocurrencies.

Looking specifically at each cryptocurrency, Bitcoin was the first to be circulated and although the most popular, it suffers the most criticism for its design. One of these issues that is the most contended is the slow confirmation time problem as discussed above. Figure 5.1(c) shows Bitcoin has one of the lowest values of $\hat{\nu}$ out of all cryptocurrencies, and as such, one of the highest levels of liquidity risk. This is highly contrastive to their fiat counterparts, given that even though Bitcoin is the most transacted cryptocurrency, it still has one of the largest liquidity risks due to its older embedded technology. Although NEM is considered extremely safe relative to other cryptocurrencies and security is at the forefront of its design, it is only marginally faster than Bitcoin to transact, and suffers from the slow confirmation time issue. As a result, it behaves similar to Bitcoin as seen from Figure 5.1(c) and there is not much added benefit in terms of risk-reduction relative to ETH and Dash.

On the other hand, Ripple has been increasingly used by banks and large corporations as their preferred settlement infrastructure technology due to its safety feature that minimizes future exchange rate volatility risk. Ripple is not dependent on any third party for redemption, and as such, it is the only currency with no counter-party credit risk. In other words, there is virtually no risk between transactions performed today and that of any risks (volatility) observed the next day. This characteristic is again consistent with the results that Ripple has the lowest near zero $|\rho|$ indicating its weakest leverage effect amongst all cryptocurrencies. Although banks prefer to

use Ripple as it has the lowest overnight (leverage) risk, it still shows extremely high non-Gaussian errors ($\nu \approx 3$) and relatively large volatility of volatility $\hat{\sigma}^2$ estimates.

To the best of our knowledge, we confirm for the first time in the literature, that these assertions are indeed consistent demonstrating the capability of our models to explain many subtle facts of cryptocurrencies. In summary, cryptocurrencies which mainly focus on security are still considered just as risky without fixing the slow confirmation time issue. Therefore, currencies which focus more on reducing transaction time issues have less risk than those which do not, even with more robust security measures.

Bitcoin

We conclude this section by reporting in details parameter estimates of the models for Bitcoin in Table 5.6 and providing an in-depth discussion of the results.

Model	$\boldsymbol{h}$	u	d	α	β	σ^2	ν	ρ	DIC
SV									
$\hat{\theta}$				-7.2522	0.9313	0.2851			-9128
Std.				0.2368	0.0160	0.0593			
AR(%)	1.0000				0.9691				
GR				0.9999	1.0001	1.0005			
SV-LVG									
$\hat{\theta}$				-7.0132	0.9011	0.4628		-0.2913	-7978
Std.				0.2051	0.0144	0.0489		0.0333	
AR(%)	0.1055				0.9745			0.4115	
GR				1.0069	1.0777	1.7547		1.0035	
GMA-SV									
$\hat{\theta}$		-0.1783	0.0333	-7.2675	0.9274	0.3058			-8953
Std.		0.4275	0.0205	0.2343	0.0178	0.0706			
AR(%)	1.0000	0.2801	0.2575		0.9686	0.0000			
GR		1.0003	1.0004	1.0001	1.0033	1.0047			
GMA-SV-LVG									
$\hat{\theta}$		0.9289	0.0137	-6.9794	0.9110	0.3961		-0.2615	-8224
Std.		0.3010	0.0085	0.2040	0.0122	0.0234		0.0299	
AR(%)	0.1545	0.4819	0.2633		0.9759	0.3702		0.2079	
GR		1.0092	1.0190	1.0026	1.0211	1.3545		1.0212	
GMA-SV-HC									
$\hat{\theta}$		-0.2842	0.0724	-0.0028	0.9929	0.4448	3.0654		-8978
Std.		0.0191	0.0153	0.0472	0.0031	0.0325	0.0660		
AR(%)	0.2915	0.2800	0.2980		0.9063		0.0660		
GR		1.0034	1.5224	1.0004	1.0025	1.0718	1.0228		
GMA-SV-LVG-HC									
$\hat{\theta}$		-0.3130	0.0671	-0.0051	0.9906	0.3737	3.0882	-0.2908	-9678
Std.		0.0526	0.0137	0.0469	0.0032	0.0353	0.0914	0.0366	
AR(%)	0.0927	0.3251	0.3529		0.9290		0.0914	0.0366	
GR		1.0718	1.0010	1.0001	1.0000	1.2281	1.0018	1.1640	

TABLE 5.6: Analysis of BTC data.

In looking at the model results, firstly, the Gelman-Rubin convergence test statistic (GR) shows that all parameters have converged. The plain SV model with or without long memory effect shows an acceptance rate of $\boldsymbol{h}$ equal to 100%. This shows the target density proposal variance is too low and unable to search the entire space properly. Evidently, it is clear that leverage effects and/or heavy tailed error distributions are important model features to sample the latent volatility vector properly for Bitcoin. The estimates $\hat{u}$ are generally negative indicating instantaneously oscillating ACF for most models, but is not statistically significant for the GMA-SV model. Also,

$\hat{d}$ is low and significant. The introduction of heavy tailed effects reduces the level of $\hat{\alpha}$ to a value close to zero, and can be ignored in the future as it is not statistically significant. The value of $\hat{\beta}$ is close to one in the heavy tailed cases, which is commonly observed in most financial time series. Contrastingly, without including heavy tails, the value of $\hat{\beta}$ tends to be lower, and is an atypical observation. Possibly, this lowering of volatility persistence is due to the distorting effect of outliers which can be allowed for in a heavy tailed error distribution.

The volatility of volatility estimate $\hat{\sigma}^2$ is unconventionally large for all models. The degrees of freedom estimate $\hat{\nu}$ is approximately 3 for both models with and without leverage effect, and are both statistically significant. This result also confirms the necessity of adopting a heavy tailed error distribution to downweigh the effect of outliers and hence protect inference. Indeed, the level of kurtosis is significant in this data. The leverage parameter $\hat{\rho}$ indicates a consistent level of negative relationship between volatility and the previous day return rate, remains fairly consistent and is statistically significant in the three models where it is included.

5.5.2 Out-of-sample forecast with Bitcoin

In this section, we conduct a forecasting exercise using Bitcoin for demonstrative purposes. We only report on Bitcoin due to its popularity and space constraints. Out of sample forecasts are measured using Value at Risk (VaR). In essence, VaR is the worst expected loss forecast under certain model assumptions at a given level of confidence χ. Denote the VaR forecast at time t, conditional on the natural filtration $\mathcal{F}_{t-1}$ as $\text{VaR}_{t|t-1}(\chi)$. The VaR is defined as

$$\chi = \Pr(y_t \leq \text{VaR}_{t|t-1}(\chi))$$

where $\chi \in (0,1)$ is the probability level. Forecasting VaR is a straightforward and intuitive exercise in a Bayesian setting. Parameter vectors drawn from the posterior distributions are used to generate a new data set under the model. This new data set is then used to make inferences after averaging out. Define the predictive density to be the forecasted density at some future time point t as $p(y_t|\mathcal{F}_{t-1}, \boldsymbol{\theta})$. Thus, for each MCMC iterate of $\boldsymbol{\theta}^i$, we are able to easily generate one forecast estimate, $\hat{y}_t$, from the predictive density $p(y_t|\mathcal{F}_{t-1}, \boldsymbol{\theta}^i)$. The violation rate (VR), which is the average

number of violations across all forecasted time periods is the most widely accepted measure for comparing model performance based on VaR and is defined as

$$\text{VR} = \frac{1}{m}\sum_{t=T-m}^{T} I(y_t < \widehat{\text{V}}\text{aR}_{t|t-1}(\chi))$$

where m is the forecast window and $\widehat{\text{V}}\text{aR}_{t|t-1}(\chi)$ is the sample estimate. Under the Basel Accord it is preferable to have models which are too conservative (VR $< \alpha$) as opposed to models which are too risky (VR $> \alpha$). We use the most popular regulator's loss functions which are surveyed in Abad, Muela, and Martin (2015), and defined as

1. Lopez's quadratic Loss (D1) $= \begin{cases} 1 + (\text{VaR}_{t|t-1} - y_t)^2 & \text{if } y_t < \text{VaR}_{t|t-1}, \\ 0 & \text{if } y_t \geq \text{VaR}_{t|t-1}, \end{cases}$

2. Lineal Loss (D2)$= \begin{cases} (\text{VaR}_{t|t-1} - y_t) & \text{if } y_t < \text{VaR}_{t|t-1}, \\ 0 & \text{if } y_t \geq \text{VaR}_{t|t-1}, \end{cases}$

3. Quadratic Loss (D3) $= \begin{cases} (\text{VaR}_{t|t-1} - y_t)^2 & \text{if } y_t < \text{VaR}_{t|t-1}, \\ 0 & \text{if } y_t \geq \text{VaR}_{t|t-1}, \end{cases}$

4. Caporin's Loss 1 (D4) $= \begin{cases} |1 - |\frac{y_t}{\text{VaR}_{t|t-1}}|| & \text{if } y_t < \text{VaR}_{t|t-1}, \\ 0 & \text{if } r_t \geq \text{VaR}_{t|t-1}, \end{cases}$

5. Caporin's Loss 2 (D5) $= \begin{cases} \frac{(|\text{VaR}_{t|t-1}| - |y_t|)^2}{|\text{VaR}_{t|t-1}|} & \text{if } y_t < \text{VaR}_{t|t-1}, \\ 0 & \text{if } y_t \geq \text{VaR}_{t|t-1}, \end{cases}$

6. Caporin's Loss 3 (D6) $= \begin{cases} |\text{VaR}_{t|t-1} - y_t| & \text{if } y_t < \text{VaR}_{t|t-1}, \\ 0 & \text{if } y_t \geq \text{VaR}_{t|t-1}. \end{cases}$

The Bitcoin data consists of $1,489$ data points and our training dataset is 95% of the available data which contains $1,415$ data points and these are used to slide one day ahead to forecast VaR for the remaining 74 days.

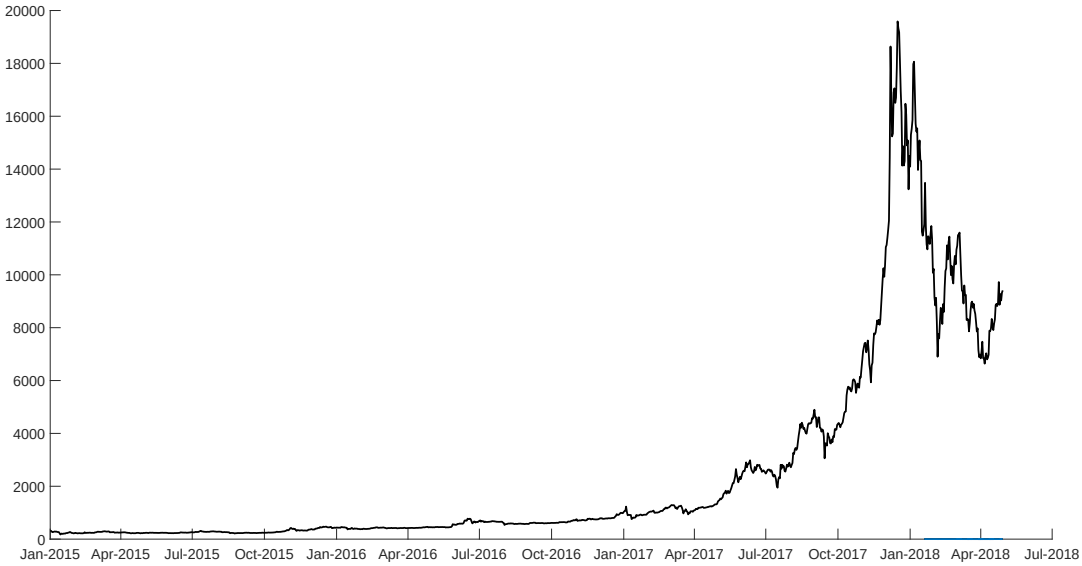

FIGURE 5.5: Plot of Bitcoin price data from the 1st of January, 2015 to the 30th of April, 2018

		Historical	R-Metrics	SV	SV-LVG	SV-GMA	SV-GMA-LVG	SV-GMA-HC	SV-GMA-LVG-HC
	DIC			1.07	1.15	1.07	1.16	1.00	1.00
	LL			0.95	0.85	0.95	0.84	1.01	1.00
1% VaR	VR	0.0%	3.4%	1.7%	5.1%	1.7%	0.0%	1.7%	0.0%
	D1	0.000	0.034	0.017	0.051	0.017	0.000	0.017	0.000
	D2	0.000	-0.000	-0.000	-0.001	-0.000	0.000	-0.000	0.000
	D3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	D4	0.000	0.003	0.001	0.015	0.001	0.000	0.005	0.000
	D5	0.000	0.013	0.007	0.015	0.006	0.000	0.003	0.000
	D6	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
5% VaR	VR	10.2%	10.2%	8.5%	11.9%	8.5%	11.9%	10.2%	6.8%
	D1	0.102	0.102	0.085	0.119	0.085	0.119	0.102	0.068
	D2	-0.002	-0.002	-0.002	-0.003	-0.002	-0.002	-0.003	-0.001
	D3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	D4	0.026	0.027	0.020	0.040	0.022	0.019	0.046	0.017
	D5	0.027	0.026	0.024	0.022	0.023	0.033	0.016	0.017
	D6	0.002	0.002	0.002	0.003	0.002	0.002	0.003	0.001
10% VaR	VR	22.0%	15.3%	15.3%	16.9%	15.3%	13.6%	18.6%	15.3%
	D1	0.221	0.153	0.153	0.170	0.153	0.136	0.187	0.153
	D2	-0.006	-0.004	-0.004	-0.005	-0.004	-0.004	-0.006	-0.004
	D3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	D4	0.079	0.048	0.045	0.069	0.048	0.047	0.081	0.050
	D5	0.035	0.031	0.031	0.027	0.030	0.029	0.029	0.031
	D6	0.006	0.004	0.004	0.005	0.004	0.004	0.006	0.004

TABLE 5.7: Each parameter is the average across the forecast horizon period: LL: Log-likelihood. DIC: Deviance Information Criterion. VR: Violation rate. D1: Lopez distance. D2: Lineal distance. D3: Quadratic distance. D4: Caporin1 distance. D5: Caporin2 distance. D6: Caporin3 distance.

The first two rows of Table 5.7 report the average DIC and LL across the forecast horizon. It is clear from these two measures, the inclusion of Gegenbauer long memory, stochastic volatility and heavy common tail are favoured. All violation rates and all deviations are the smallest in general for the two smallest percentiles under the SV-GMA-LVG-HC specification.

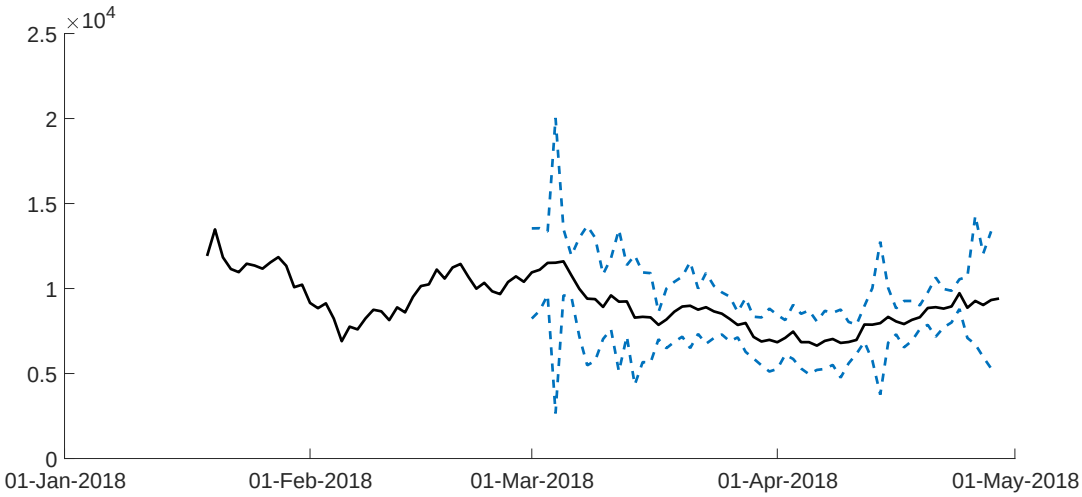

FIGURE 5.6: Black line: Bitcoin price. Blue dotted lines: 1% and 99% one-step ahead VaR forecasts using the GMA-SV-LVG-HC model.

Figure 5.6 shows the Bitcoin price across the forecast horizon, alongside the 1% and 99% one-step ahead forecast using the GMA-SV-LVG-HC model, which had a violation rate of 0%. Interestingly, it should be noted the large spike of the VaR bounds in early March of 2018 correspond to the beginning of the $\$10$ billion USD lawsuit against the apparent Bitcoin founder, Satoshi Nakamoto.

5.6 Conclusion and future research

The statistical field has yet to pay attention to the highly debated digital currency world. There is currently a global heated debate on the extreme volatility characteristics of Cryptocurrencies, and we aim to start this discussion within the financial time series community by shedding light on their unique statistical properties. As stylized facts were postulated on fiat currencies, these digital counterparts also require the same treatment. The standard models which are readily available to most statistics

seem to be inadequate to properly capture the variability of these extremely wild behaving currencies. There seem to be a myriad of statistical properties which together, in unison, are able to properly explain such wild behaviours.

In our work, we explain and detail a realistic investable basket of cryptocurrencies and explain their main properties. The work is carried out through the GMA-SV-LVG-HC model, which attempts to capture the main stylized facts of cryptocurrencies. Arguably, some of the highly debated topics surrounding cryptocurrencies manifest themselves in our results, and we are able to provide a statistical handle on this matter. The most cited and known Cryptocurrency, Bitcoin, is a prime example of this. We show its behaviours are best suited to be estimated by the GMA-SV-LVG-HC model via a VaR analysis.

To date, and for the foreseeable future, there is heated debate on whether or not the associated infrastructure and hardware differences between Cryptocurrencies lead to a reduction in risk. Much of this debate is carried out on internet forums in a speculative nature within programming circles. We prove, to the best of our knoweldge, for the first time, the differences in risks amongst the most popular and debated Cryptocurrencies. We measure, and scientifically prove one such important example of this, especially important to global banks, is that Ripple indeed does provide the lowest over-night risk compared to all other Cryptocurrencies. This is an important finding for financial institutions, such as American Express who convert overnight debt into Ripple for liquidity purposes now.

Interestingly, the entire Bitcoin ledger since its inception is available online. This data is extremely difficult to obtain for fiat currencies, and is an extremely exciting venture for future research.

Chapter 6

A new look at Cryptocurrencies

"The future of money is digital currency."

Bill Gates

6.1 Introduction

Academic interests in anonymous communications research date back to the early 1980s (Chaum, 1981), and the first digital currency, *DigiCash*, was launched in 1990 which offered anonymity through cryptographic protocols. Nakamoto (2008) resurrected philosophies of (Chaum, 1981) with the addition of crowd sourcing and peer-to-peer networking which both avoid centralized control. Today, this has manifested itself into a growing Cryptocurrency community which now includes banks, hedgefunds and even Government. The most popular Cryptocurrency and largest by market capitalization is Bitcoin. A $1,000 USD investment in Bitcoin in July of 2010 would have returned $81,000,000 just 7 years later. Bitcoin, or Cryptocurrencies in general face scrutiny as being speculative (Cheah and Fry, 2015). Conversely, there is evidence to suggest the Cryptocurrency market is still in its infancy and is inefficient (Urquhart, 2016), with properties such as price clustering (Urquhart, 2017). There is however a strong growing network of Bitcoin users and academics who are shedding light on this new technology. In this work, we discuss a large investable sample of Cryptocurrencies, and conditionally measure some important stylized facts.

The remainder of this note is organised as follows: in Section 2, we discuss our data source and the model; in Section 3 we discuss our empirical findings and conclude with Section 4.

6.2 Data and methodology

The long memory effect of Hosking (1981) was identified in Bitcoin by Bariviera (2017). We extend these findings to model and conditionally measure the generalised long memory effect of Gray, Zhang, and Woodward (1989b). Another important feature found in financial time series is the leverage effect which has its roots in the asymmetric return-volatility relationship attributed to financial leverage or debt-to-equity ratios. The leverage effect is the notion of a negative correlation between one-day ahead volatility and returns. Generalised autoregressive conditional hetereoscedastic (GARCH) models have been successfully used to measure time-varying volatility in Bitcoin data (Katsiampa, 2017). We however plan to do this using the stochastic volatility model of Taylor (1986) to describe the time varying nature of volatility typically found in financial returns. See Shephard (2005) for a detailed comparison of the two approaches.

An additional stylized fact of financial returns of assets such as stocks and currencies is they are not normally distributed. The usual treatment to measure this feature is to modify the observation and/or the latent equation to include a heavy-tailed distribution (Chib, Nardari, and Shephard, 2002; Omori and Watanabe, 2008). Incorporating all of these features commonly found in financial time series, we construct a model which describes all of these properties.

The data for this analysis is sourced from the Brave New Coin (BNC) Digital Currency indices (**BNC**). BNC surveys hundreds of trading platforms and currently records $2,796$ Cryptocurrency time-series indices. However, some of these have market capitalizations which are small ($< \$1,000,000$ USD) and traded very little. Of the $2,796$ data sets available on the BNC database, only 224 of these have been exchanged at least once per day since inception. The time series y_t is defined as the daily index price percentage change $y_t = (P_t - P_{t-1})/P_{t-1}$, where P_t is the daily index value at time t. It should be noted that alternative transformations to de-trend the data can be used, such as $y_t = \log(P_t/P_{t-1})$. Although Cryptocurrencies were first introduced in 2008, BNC only reports price points when more formalized exchanges for each respective currency could be measured with reliability. As such, the number of observations recorded for each Cryptocurrency vary, but all end on the 31st of July, 2017.

The time series model fitted in this note measures generalised long memory (GLM), stochastic volatility (SV), leverage (LVG) and heavy tails (HT). Let $y_t, t = 1, 2, ..., T$ be a stochastic process satisfying the equations

$$\text{GLM} : (1 - 2uB + B^2)^d y_t = \varepsilon_t, \tag{6.1}$$

$$\text{SV} : h_{t+1} = \alpha + \beta(h_t - \alpha) + \eta_{t+1}, \tag{6.2}$$

$$\text{LVG-HT} : \begin{pmatrix} \varepsilon_t \\ \eta_{t+1} \end{pmatrix} \sim t_\nu \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} e^{h_t} & \sigma\rho e^{h_t/2} \\ \sigma\rho e^{h_t/2} & \sigma^2 \end{pmatrix} \right). \tag{6.3}$$

It is known that y_t has long memory effects when $(\{|u|< 1, 0 < d < 0.5\} \cup \{|u|= 1, 0 < d < 0.25\})$. There is assumed to be a leverage effect between the errors of the observation equation (6.1) and the latent equation (6.2) such that $\mathbb{E}[\varepsilon_t \eta_{t+1}] = \rho$. Further, these components are assumed to follow a bivariate Student-t distribution. Clearly, h_t is the log-volatility, which evolves according to the state equation (6.2) for $t = 1, \ldots, T$, α is the constant level of the volatility, β is the persistence of the volatility process and σ^2 is the volatility of volatility. We assume $|\beta|< 1$ so h_{t+1} is stationary.

6.3 Empirical results

Firstly, we focus on the 5 largest Cryptocurrencies measured by market capitalization on the 31st of July, 2017 (**BNC**) (see Table 6.1). As expected, we see that currencies with lower market capitalizations exhibit larger variability. The Ljung-Box (L-B) tests of $|y_t|$ and y_t^2 show strong evidence of long memory and time-dependant volatility respectively. The Kolomogorov-Smirnov test for normality is also rejected. The L-B test, the normality test, the high level of kurtosis and the volatility clustering in Figure 1 all confirm the need for model 6.1-6.3.

	Rank	Market Cap. ($B)	No. of obs	Mean	Std.	Skewness	Kurtosis	Min.	Max.	L-B($\lvert y_t \rvert$)	L-B(y_t^2)	Normality test
Bitcoin	1	137.0078	1489	0.0012	0.0414	-0.8543	10.2330	-0.2709	0.2250	762.3674	255.8224	3426.8910
										($< .0001$)	($< .0001$)	($< .0001$)
Ethereum	2	67.0341	995	0.0043	0.0715	-0.2145	7.3622	-0.3959	0.3293	281.7172	141.7939	796.5499
										($< .0001$)	($< .0001$)	($< .0001$)
Ripple	3	26.0533	1489	0.0000	0.0781	-1.3992	31.8162	-0.8844	0.6393	559.8652	103.5672	52003.6272
										($< .0001$)	($< .0001$)	($< .0001$)
NEM	12	3.0601	1117	0.0026	0.0877	0.4502	9.3332	-0.4656	0.6443	291.9821	136.7968	1904.4886
										($< .0001$)	($< .0001$)	($< .0001$)
Dash	14	2.6920	1483	0.0018	0.0727	0.0821	10.6803	-0.4881	0.5232	851.7943	591.9198	3646.5302
										($< .0001$)	($< .0001$)	($< .0001$)

TABLE 6.1: Summary statistics of the global weighted average indices for each relevant Cryptocurrency. P-values of the relevant columns are reported in parantheses. L-B: Ljung-Box Q-test for residual autocorrelation. We note they all end on the 30^{th} of April, 2018 but the start date varies, as per the number of observations listed.

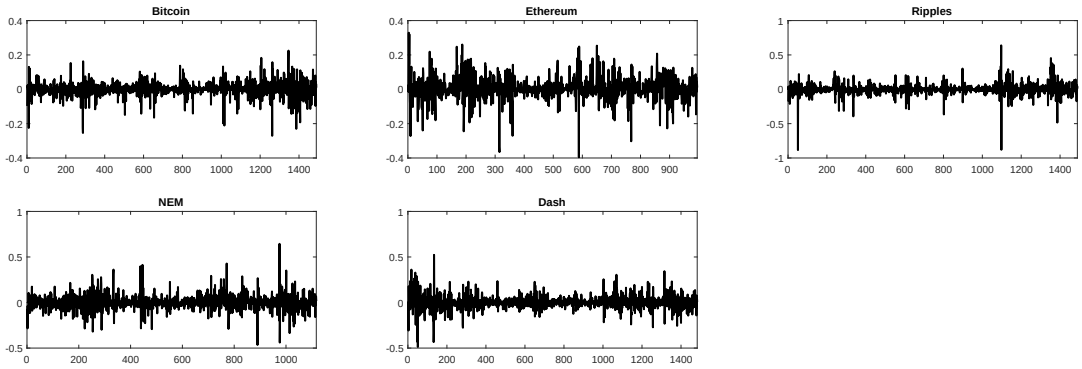

FIGURE 6.1: Time series plots of the price percentage change for the five largest Cryptocurrencies measured by market capitalization.

Model 6.1-6.3 is estimated using the filtered investable universe of 224 different Cryptocurrency indices[1]. We also plot the names of the top 5 Cryptocurrencies to show where they stand relative to their counterparts.

[1]The list of names is in the appendix attached to this letter.

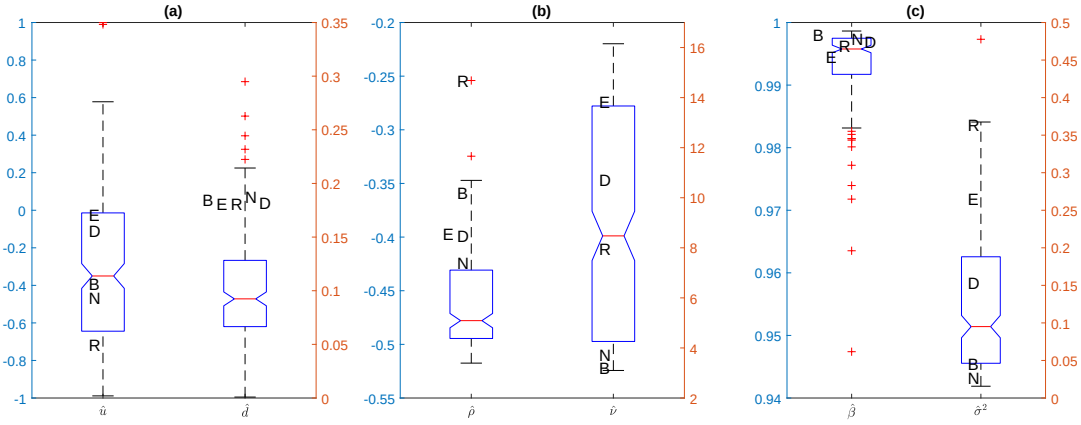

FIGURE 6.2: Notched boxplots of parameter estimates of 224 different cryptocurrency data sets under the GLM-SV-LVG-HT model. B: Bitcoin, E: Ethereum, R: Ripples, N: NEM, D: Dash. (a) $[\hat{u}, \hat{d}]$. (b) $[\hat{\rho}, \hat{\nu}]$. (c) $[\hat{\beta}, \hat{\sigma}^2]$.

As evidenced in Figure 6.2(a), most estimates of $\hat{u}$ are negative. As $\hat{u}$ approaches -1 from the right, the sample autocorrelation function becomes instantaneously oscillating. Twenty five percent of our investable universe have a positive $\hat{u}$, among which the largest is 0.6 corresponding to a period of around 7 days. Remarkably, the top 5 Cryptocurrencies by market capitalization have a value of $\hat{d}$ which is clustered around 0.18. This is suggestive that as Cryptocurrency markets mature, they tend to have similar long memory persistence characteristics. All estimates of $\hat{\rho}$ are negative and tend to cluster between -0.4 and -0.5, which implies that one day ahead volatility and returns are negativley correlated. This too is the assumed case in most financial time series to have a negative ρ, and therefore shows that Cryptocurrencies also share this behaviour. The *volatility of volatility* estimate $\hat{\sigma}^2$ shows the existence of a stochastic volatility process. Some commonly traded Cryptocurrencies, such as Ripples, show extreme volatility characterstics. The estimated parameter $\hat{\beta}$ reflects volatility persistence over time and is highly suggestive that all Cryptocurrencies in our investable universe show evidence of volatility clustering. This further validates the volatility clustering shown in Figure 6.1. After allowing for these various effects, the errors show a diverse level of kurtosis with $\hat{\nu}$ ranging from 3 for Bitcoin showing extreme kurtosis to 16 showing moderate level of kurtosis.

Interestingly, Ripples is not dependant on any third party for redemption, and as such, it is the only currency with no counter-party credit risk. Due to this safety feature, Ripples has been increasingly used by banks and large corporations as their preferred settlement infrastructure technology due to minimized future exchange rate volatility

risk - this is indeed in line with our findings as it has the lowest $|\rho|$ indicating it has the weakest leverage effect amongst all Cryptocurrencies.

The main features of Ethereum (ETH) and Dash compared to all other Cryptocurrencies is they are more user-friendly. There is a larger community based approach with computer programmers actively making both Cryptocurrencies easier, safer and quicker to use. The biggest criticism of BTC is that transacting money can be an extremely slow process, sometimes taking up to 48 hours for Bitcoins to be sent from one user to another. ETH uses smart contracts to use blockchains in comparison to BTC which does not. Also, Dash is the only currency that uses instant transactions ("InstantSend"). InstantSend is a feature of Dash which allows for almost near-instant transactions, which solves the longer confirmation time problem of Bitcoin. This can be perceived that ETH and Dash have lower liquidity risk than BTC. This is indeed consistent with our findings since both ETH and Dash have a higher value of ν, which implies their error distributions behave closer to a Gaussian distribution with smaller kurtosis than BTC which nearly has the lowest value of ν. While BTC has a relatively low value of σ^2 which is similar to other financial returns, it is clear that most of the variability of BTC can be attributed to a heavy tailed distribution.

6.4 Conclusion

This work is deeply motivated by the unique characteristics found in Cryptocurrency data, which are drawing media and academic attention. The empirical data analysis shows Cryptocurrencies exhibit long memory, leverage, stochastic volatility and heavy tailedness. We further shed light on a larger scope of the Cryptocurrency universe by expanding our analysis to cover 224 Cryptocurrency indices. Although still in its infancy, we contribute a deeper understanding surrounding Cryptocurrencies for the upcoming regulators, investors and governments to explore further on the topic.

Chapter 7

Further extensions to realised volatility, buffer threshold and jumps for Cryptocurrency modelling

"A statistician confidently tried to cross a river that was 1 metre deep on average. He drowned."

Unknown

Chapters 5 and 6 show that Cryptocurrencies are inherently different to traditional fiat currencies. The upcoming chapter extends the model discussed in the previous two chapters in order to further address the characteristics of Cryptocurrencies. As such, we now review some time series models for such extensions and discuss in more detail the methodologies of estimating these extensions.

As previously discussed, Cryptocurrencies are well known for their wild volatility. One approach that we consider in these two chapters is to allow for the negative association between returns and future volatilities via a correlation coefficient. However, volatilities in the SV model are estimated as a latent process in the return series which are constructed using closing prices, neglecting all intra-day price movements. To supplement this, one alternative approach is to measure volatility directly through realised volatility measures. realised measures are an important metric since they

provide a real-world handle on statistical models. The incorporation of realised measures has proven to greatly increase model accuracy, and there is no exception in the Cryptocurrency case.

Apart from the leverage effect on volatility, long memory effects should also be considered. Previous chapters have focused on modelling the long memory feature of returns. However, due to the presence of occasional jumps contributing to wild volatility in the return series, the long memory feature in the return series may not be detected as efficiently as in the volatility process. Upon further inspection of the ACF of the squared returns for a basket of Cryptocurrencies, we also find the presence of oscillatory long memory. Hence, the incorporation of long memory into the realised volatility measure is another direction we pursue.

We lastly address the particularly versatile features of Cryptocurrency returns by introducing non-linear attributes. We consider two main types of non-linear models, the threshold model and the jump model. Adopting the idea of Chan, Choy, and Lam (2014), we also consider threshold jump models and therefore combine both concepts. These models are reviewed in the next section.

7.1 A review of time series models for extensions

Although theoretically and practically pleasing, the model in (6.1) to (6.3) can still be improved upon in different directions. The non-linearity of Cryptocurrencies deserves further specialized attention in order to fully measure all of their unique features. Before further addressing these, we first briefly review the realised volatility model.

7.1.1 realised volatility models

SV Model with realised Volatility

The standard SV model with realised volatilities corrected for bias due to market microstructure noise and non-trading hours was first proposed by Takahashi, Omori,

and Watanabe (2009) and is denoted as

$$y_t = \exp(h_t/2)\varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0,1) \tag{7.1}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad \eta_t \sim \mathrm{N}(0,\sigma^2) \tag{7.2}$$

$$v_t = \gamma + h_t + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0,\sigma_v^2) \tag{7.3}$$

where v_t are the log realised volatilities at time t. realised volatility is typically an exogenous measure which is able to better filter $\boldsymbol{h}$ in order to improve estimation of the entire model. Further, γ has the conventional interpretation of being driven by market microstructure noise and non-trading hour effects. When γ is positive, realised volatility has an upward bias and when γ is negative, it has a downward bias. Therefore, we may observe the strength of the effect of the microstructure noise and non-trading hours from the sign of γ. Note that all currencies, including Cryptocurrencies, are trader driven markets, therefore the effects of γ are mainly due to market microstructure noise.

Heterogeneous Autoregressive model for the realised Volatility

One notable extension of the realised volatility model is the Heterogeneous Autoregressive with realised Volatility (HAR-RV) model of of Corsi (2004), which is

$$RV_{t+1,d}^{(d)} = \alpha + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \eta_{t+1,d}$$

where $RV^{(x)}, x = d, w, m$ are the realised volatilities of daily, weekly and monthly observations respectively. The purpose of this model is to capture more features of the data including long memory. McAleer and Medeiros (2008) further extended this model to include multiple smooth regime transitions.

The realised volatility model is an insightful extension to the previous modelling efforts which we have considered given the volatility characteristics of Cryptocurrencies.

7.1.2 Threshold model

In addition, the inclusion of threshold effects is also paramount to estimating some complicated time series such as Cryptocurrency returns, which may contain regime switching. We now survey some of the most popular threshold models.

Threshold AR model

One of the most notable examples of a threshold time series model is the Threshold Autoregressive (TAR) model of Tong (1990). Although simple, it provides an intuitive handle on the versatility of some return series, and is especially relevant to the context of Cryptocurrencies. The standard TAR model is given by

$$y_t = \begin{cases} \phi_1 y_{t-1} + \varepsilon_t, & \text{if } R_{t-1} = 1, \\ \phi_2 y_{t-1} + \varepsilon_t, & \text{if } R_{t-1} = 0, \end{cases}$$

where the regime indicator is

$$R_t \quad = \quad \begin{cases} 1, & \text{if } y_t > r, \\ 0, & \text{if } y_t \leq r. \end{cases} \tag{7.4}$$

The TAR model allows separate regimes to co-exist according to the level of the returns.

Threshold stochastic volatility

The Threshold SV (TSV) model was proposed by So, Lam, and Li (2002) and is given by

$$y_t = a + by_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, e^{h_t}),$$
$$h_{t+1} = (\alpha + \beta h_t)R_{t+1} + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \sigma^2)$$

where R_t is defined in (7.4). The purpose of this specification is to capture the variance asymmetry as an extension to the TAR model. Essentially, it is an alternative

method to measure the leverage effect. So and Choi (2009) extended this to the multivariate case.

Heavy Tailed TSV model

Chen, Liu, and So (2008a) generalised the TSV model to include exogenous effects, threshold effects and heavy tails given by

$$\begin{aligned} y_t &= a + by_{t-1} + \boldsymbol{\phi}\boldsymbol{X} + \varepsilon_t, \quad \varepsilon_t \sim t_\nu(0, e^{h_t}), \\ h_{t+1} &= (\alpha + \beta h_t)R_{t+1} + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \sigma^2) \end{aligned}$$

where R_t is again defined in (7.4), $\boldsymbol{\phi}$ is a vector of coefficients $(\phi_1, \ldots)$ and $\boldsymbol{X}$ is a matrix of exogenous variables in the typical set up. They noted the threshold variable could be based on local market information, such as lagged values of y_t or other exogenous variables.

Threshold of Error term

Wirjanto, Kolkiewicz, and Men (2016) proposed another variant of the TSV model

$$\begin{aligned} y_t &= \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, e^{h_t}\lambda_t^2), \\ h_{t+1} &= \alpha + \beta(h_t - \alpha) + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \sigma^2) \end{aligned}$$

where the error ε_t has a scaled variance which follows a threshold scheme with regime indicator

$$\lambda_t^2 = \begin{cases} \lambda_1^2, & \text{if } y_{t-1} > r, \\ \lambda_2^2, & \text{if } y_{t-1} \leq r. \end{cases}$$

Standard Buffered AR model

Another extension is the generalization of the regime switching scheme. This gives rise to a more sophisticated version of the TAR model of Tong (1990) called the Buffered Autoregressive (BAR) model, which was initially proposed by Zhu, Yu, and

Li (2014). The first order BAR model is given by

$$y_{t+1} = \begin{cases} \phi_1 y_t + \varepsilon_{t+1}, & \text{if } R_t = 1, \\ \phi_2 y_t + \varepsilon_{t+1}, & \text{if } R_t = 0, \end{cases} \tag{7.5}$$

with the regime indicator

$$R_t \quad = \quad \begin{cases} 1, & \text{if } y_{t-1} \leq r_L, \\ R_{t-1}, & \text{if } r_L < Y_{t-1} \leq r_U, \\ 0, & \text{if } y_{t-1} > r_L. \end{cases} \tag{7.6}$$

This results in two threshold points, r_L and r_U, to ensure the regime switching in both directions will only occur after passing a buffer.

Double Hysteric Heteroskedastic model

As the BAR model is relatively new, little extensions have been considered. The most notable one is that of Chen and Truong (2016) who proposed a double buffer AR model with Student's t-errors called the double AR(p)-GARCH(q, m) model

$$y_t = \begin{cases} \phi_1 + \sum_{i=1}^{p_1} \phi_i^{(1)} y_{t-i} + a_t, & \text{if } R_t = 1, \\ \phi_2 + \sum_{i=1}^{p_2} \phi_i^{(2)} y_{t-i} + a_t, & \text{if } R_t = 0 \end{cases}$$

where $a_t = \sqrt{h_t}\varepsilon_t$, $\varepsilon_t \sim t(0, 1)$ and

$$h_t = \begin{cases} \alpha_1 + \sum_{i=1}^{q_1} \alpha_i^{(1)} a_{t-i}^2 + \sum_{i=1}^{m_1} \beta_i^{(1)} h_{t-i}, & \text{if } R_t = 1, \\ \alpha_2 + \sum_{i=1}^{q_2} \alpha_i^{(2)} a_{t-i}^2 + \sum_{i=1}^{m_2} \beta_i^{(2)} h_{t-i}, & \text{if } R_t = 0 \end{cases}$$

with the regime indicator (7.6) in which the threshold variable y_{t-1} is replaced by any exogenous variable z_t, such as the d-lagged variable y_{t-d}.

7.1.3 Jump model

A simple jump SV model is defined as

$$y_t = k_t q_t + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, \exp(h_t)), \tag{7.7}$$

$$h_t = \alpha + \beta(h_{t-1} - \alpha) + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \sigma^2) \tag{7.8}$$

where $q_t \in \{0, 1\}$ is the jump indicator variable (1 = there is a jump at time t, 0 = no jump at time t), $\mathbb{P}(q_t = 1) = \kappa$, k_t is the jump size such that $k_t \sim \mathrm{N}(\mu_k, \sigma_k^2)$ and μ_k and σ_k^2 are the mean and variance of k_t respectively. We denote $\boldsymbol{q} = (q_1, \ldots, q_T)$ and $\boldsymbol{k} = (k_1, \ldots, k_T)$.

7.2 Potential model features

As discussed in the previous section, there are an abundance of potential model features to explore. Each feature provides its usefulness in particular situations, and so the choice of which feature to consider must be reflective of the features of the data under question. The most notable feature of Cryptocurrencies is undoubtedly their unconventionally large volatility. This volatility is often sensationalized in the media, with claims that Cryptocurrency prices "jump overnight" and "crash". The reason for these claims is because it is not uncommon for a well-established Cryptocurrency, such as Bitcoin, to experience moves of $\pm 10\%$ within the space of one day. Through a statistical lens, this would suggest the inclusion of jump type behavior in order to measure this effect. The presence of jumps is empirically evidenced in Figure 6.1 where it is clear that the return series' exhibit jumps.

Further, to address the volatility characteristics, we see the importance of including realised measures as they have been shown to greatly improve volatility measurement (Andersen et al., 2003; McAleer and Medeiros, 2008; Goldman et al., 2013; Shirota, Hizu, and Omori, 2014). To be specific, we model the log daily range, which is discussed in detail in Chapter 8.

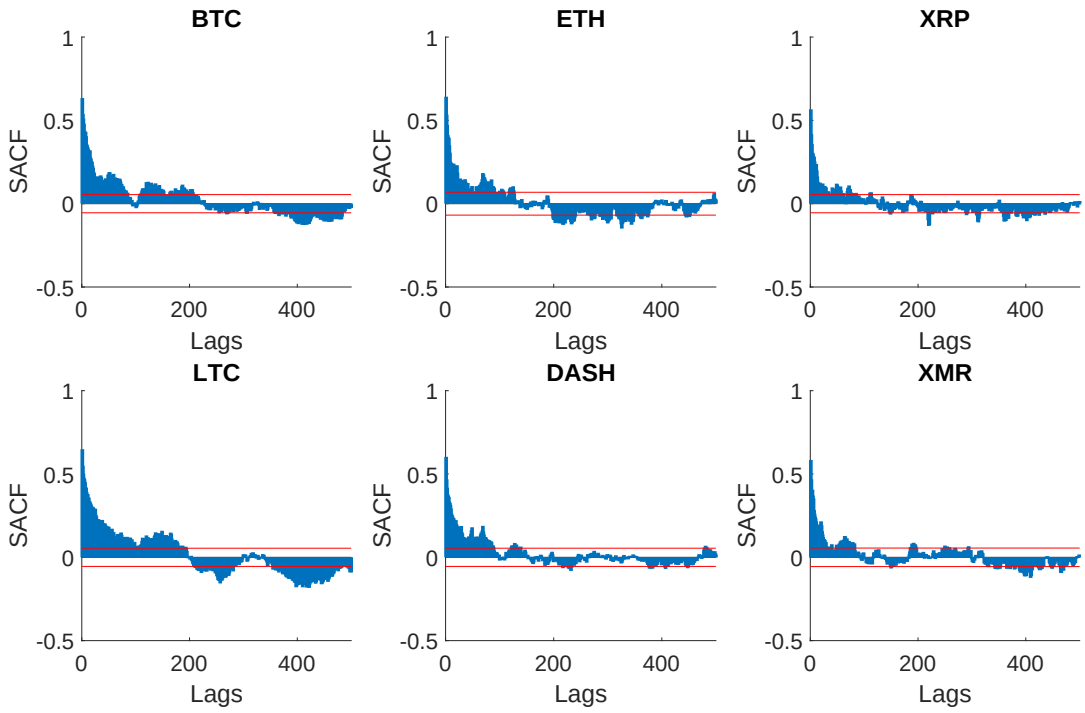

FIGURE 7.1: Sample ACF plots of the log daily return range of the 6 largest Cryptocurrencies measured by market capitalization on 31/12/2017. BTC: Bitcoin. ETH: Ethereum. XRP: Ripple. LTC: Litecoin. DASH: Dash. XMR: Monero.

There is clear evidence in Figure 7.1 of long memory, specifically, Gegenbauer long memory in the log realised volatility measure. As such, we see the need to include the log daily return range to hold important information about the volatility of Cryptocurrencies.

Finally, as will be discussed later, critics have argued that long memory effects may be confused with non-linear regime changes in the observations. As such, we include BAR effects into the return series y_t in order to allow for such effects.

Given all of these attributes, we aim to pursue a model which includes all of these features. This model is given in Section 7.3.3. However, before exploring such model, we first iteratively build up some Bayesian sampling schemes by first deriving samplers for each constituent component.

7.3 Bayesian inference

7.3.1 Bayesian inference for the jump model with SV errors

We begin with the jump model listed in (7.7) with SV errors

$$\begin{aligned} y_t &= k_t q_t + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, e^{h_t}), \\ h_t &= \alpha + \beta(h_{t-1} - \alpha) + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \sigma^2) \end{aligned}$$

where $\mathbb{P}(q_t = 1) = \kappa$ and $k_t \sim N(\mu_k, \sigma_k^2)$.

Sampling the jump indicator q_t

Since each jump indicator q_t is independently distributed as Bernoulli(κ), it can be sampled separately. The posterior distribution of q_t is

$$\begin{aligned} p(q_t|y_t, h_t, k_t, \tau^2, \kappa) &\propto f(y_t|h_t, k_t, q_t, \kappa) \times p(q_t) \\ &= [f_N(y_t|k_t, \exp(h_t))]^{q_t}[f_N(y_t|0, \exp(h_t))]^{1-q_t}\kappa^{q_t}(1-\kappa)^{1-q_t} \\ &= [\kappa f_N(y_t|k_t, \exp(h_t))]^{q_t}[(1-\kappa)f_N(y_t|0, \exp(h_t))]^{1-q_t} \\ &\propto \left[\frac{\kappa f_N(y_t|k_t, \exp(h_t))}{\kappa f_N(y_t|k_t, \exp(h_t)) + (1-\kappa)f_N(y_t|0, \exp(h_t))}\right]^{q_t} \\ &\quad \left[1 - \frac{\kappa f_N(y_t|k_t, \exp(h_t))}{\kappa f_N(y_t|k_t, \exp(h_t)) + (1-\kappa)f_N(y_t|0, \exp(h_t))}\right]^{1-q_t}, \end{aligned}$$

where $f_N(y_t|\mu, \tau^2)$ is the normal density evaluated at y_t with mean μ and variance τ^2. Hence, we have

$$q_t|y_t, h_t, \tau^2, \kappa \sim \text{Bernoulli}(\mathfrak{p}_q) \text{ where } \mathfrak{p}_q = \frac{\kappa f_N(y_t|k_t, \exp(h_t))}{\kappa f_N(y_t|k_t, \exp(h_t)) + (1-\kappa)f_N(y_t|0, \exp(h_t))}.$$

Sampling the jump probability κ

Assuming a prior of $\kappa \sim \mathrm{U}(0, 0.1)$, the posterior distribution of κ is

$$\begin{aligned} p(\kappa|\boldsymbol{Y}, \boldsymbol{h}, \boldsymbol{k}, \boldsymbol{q}) &\propto f(\boldsymbol{Y}|\boldsymbol{h}, \boldsymbol{k}, \boldsymbol{q}, \kappa) \times p(\kappa) \\ &= \prod_{t=1}^{T}[\kappa f_N(y_t|k_t, \exp(h_t))]^{q_t}[(1-\kappa)f_N(y_t|0, \exp(h_t))]^{1-q_t}\mathbb{1}_{(0,0.1)} \\ &\propto \kappa^{n_q}(1-\kappa)^{T-n_q}\mathbb{1}_{(0,0.1)} \end{aligned}$$

where $n_q = \sum_{t=1}^{T} q_t$ and $\mathbb{1}_{(0,0.1)}$ indicates the sampled κ^* to be within the range (0,0.1). Hence

$$\kappa|\boldsymbol{q} \sim \text{Beta}\left(1 + \sum_{t=1}^{T} q_t, 1 + T - \sum_{t=1}^{T} q_t\right) I_{(0,0.1)},$$

where Beta(a, b) represents a beta distribution with parameters a and b, whereby κ is sampled directly from this Beta distribution, and rejected if it falls outside of the prior domain (0,0.1).

Sampling the average jump size μ_k

As mentioned previously, it is assumed that $k_t \sim \text{N}(\mu_k, \sigma_k^2)$, and we therefore derive the posterior distributions of μ_k and σ_k^2 respectively. First, the posterior distribution of μ_k can be expressed as

$$p(\mu_k|\boldsymbol{Y}, \boldsymbol{h}, \boldsymbol{q}, \sigma_k^2) \propto p(\mu_k) \times f(\boldsymbol{Y}|\boldsymbol{h}, \mu, \boldsymbol{q}, \kappa, \sigma_k^2).$$

Assuming a Gaussian prior $\text{N}(0, 10)$ for μ_k, the log posterior density is

$$\begin{aligned} \log p(\mu_k|\boldsymbol{Y}, \boldsymbol{h}, \sigma_k^2, \boldsymbol{q}) &= \log p(\mu_k) + \log f(\boldsymbol{Y}|\boldsymbol{h}, \mu_k, \sigma_k^2, \boldsymbol{q}) + \text{terms independent of } \mu_k \\ &= -\frac{\mu_k^2}{2\sqrt{10}} - \frac{1}{2}\sum_{t=1}^{T} \log\left(\sigma_k^2 q_t + e^{h_t}\right) - \frac{1}{2}\sum_{t=1}^{T} \frac{(y_t - \mu_k q_t)^2}{\sigma_k^2 q_t + e^{h_t}} \\ &\quad + \text{terms independent of } \mu_k. \end{aligned} \tag{7.9}$$

We note that parameter k_t in (7.9) is replaced by its expected value δ_μ. The MAP sampler can then be used to sample μ_k^* and accepted/rejected with the MH algorithm.

Sampling the jump size variance σ_k^2

In the typical set up, the jump size variance is assumed to be log-normal, and jointly sampled with μ_k (Chan and Grant, 2016a). Although this may be a suitable assumption under regular stock returns, we find this to be an inadequate assumption in the case of extremely volatile returns, such as that of Cryptocurrencies. The variance parameter is typically unstable as it is often sampled with extremely large values, and it fails to explore the parameter space fully. Alternatively, we find an efficient estimator by dropping any prior distribution, and using only the likelihood function as

the posterior distribution. In addition, we use a Griddy Gibbs sampler as described in Section 4.2.9 on the data likelihood only, and accept/reject σ_k^{2*} using a random walk MH algorithm. The log posterior density therefore is

$$\begin{aligned}\log p(\sigma_k^2|\boldsymbol{Y},\boldsymbol{h},\boldsymbol{q},\mu_k) &= \log f(\boldsymbol{Y}|\boldsymbol{h},\mu_k,\sigma_k^2,\boldsymbol{q}) + \text{terms independent of } \sigma_k^2,\\ &= -\frac{1}{2}\sum_{t=1}^{T}\log\left(\sigma_k^2 q_t + e^{h_t}\right) - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t-\mu_k q_t)^2}{\sigma_k^2 q_t + e^{h_t}} + \text{terms independent of } \sigma_k^2.\end{aligned}$$

Sampling the jump size k_t

The posterior of the jump size k_t is

$$\begin{aligned}p(k_t|y_t,h_t,q_t,\kappa) &\propto f(y_t|h_t,k_t,q_t,\kappa)\times p(k_t)\\ &\propto [f_N(y_t|k_t,\exp(h_t))]^{q_t}[f_N(y_t|0,\exp(h_t))]^{1-q_t}\times f_N(k_t|\mu_k,\sigma_k^2),\end{aligned} \tag{7.10}$$

since $k_t \sim N(\mu_k,\sigma_k^2)$. We consider the following two cases for q_t:

Case 1: When $q_t = 0$, the posterior distribution in (7.10) becomes

$$p(k_t|y_t,h_t,q_t) \;\propto\; f_N(y_t|0,\exp(h_t))\times f_N(k_t|\mu_k,\sigma_k^2) \propto f_N(k_t|\mu_k,\sigma_k^2)$$

such that given $q_t = 0$,

$$k_t|y_t,h_t = \mu_k + \sqrt{\delta_{\sigma^2}}\,e_t, \quad e_t \sim \mathrm{N}(0,1).$$

Case 2: When $q_t = 1$, the posterior distribution in (7.10) becomes

$$p(k_t|y_t,h_t,q_t) \propto f_N(y_t|k_t,\exp(h_t))\times f_N(k_t|\mu_k,\sigma_k^2) \propto \mathrm{N}(V_{k,t}M_{k,t},V_{k,t}),$$

where

$$V_{k,t} \;=\; \left(\frac{1}{\sigma_k^2}+\frac{1}{e^{h_t}}\right)^{-1} \quad\text{and}\quad M_{k,t} = \left(\frac{\mu_k}{\delta_{\sigma^2}}+\frac{y_t}{e^{h_t}}\right),$$

since the posterior of μ in a normal conjugate is

$$p(\boldsymbol{x}|\mu,\sigma^2) \propto \prod_{t=1}^{n} f_N(x_t|\mu,\sigma^2)\times f_N(\mu|m,\tau^2) \propto f_N(V_\mu M_\mu, V_\mu),$$

where

$$V_\mu = \left(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\right)^{-1} \quad \text{and} \quad M_\mu = \left(\frac{m}{\tau^2} + \frac{\sum_{i=1}^n x_i}{\sigma^2}\right),$$

as given in (2.1). In summary, the posterior distribution of k_t given $q_t = 1$ is

$$k_t|y_t, h_t = M_{k,t} + \sqrt{V_{k,t}}e_t, \quad e_t \sim \text{N}(0,1).$$

Sampling the latent volatility vector h and the other volatility parameters

The sampling of the latent volatilities $\boldsymbol{h}$ is carried out as usual and can be found in Appendix F. The sampling of the volatility parameters α, β and σ^2 can be found in Section 2.2.

7.3.2 Bayesian inference for the realised volatility model

We discuss here the estimation of the model listed in (7.1) to (7.3) which is expressed as

$$\begin{aligned} y_t &= \exp(h_t/2)\varepsilon_t, \quad \varepsilon_t \sim \text{N}(0,1), \\ v_t &= \gamma + h_t + \epsilon_t, \quad \epsilon_t \sim \text{N}(0, \sigma_v^2), \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t, \quad \eta_t \sim \text{N}(0, \sigma^2). \end{aligned}$$

For ease of notation, let $\boldsymbol{Y} = (y_1, \ldots, y_T)$ and $\boldsymbol{\mathcal{V}} = (v_1, \ldots, v_T)$.

Sampling the latent volatility vector h

The estimation process of $\boldsymbol{h}$ requires some slight modification since there is information in the log realised volatility component in (7.2) which also involves h_t. The estimation is carried out in the same way as explained in Appendix F, except the log density (and therefore its first two partial derivatives) are different. To show this, we consider the conditional density of the realised volatility component which is

$$f(y_t|v_t, h_t, \gamma, \sigma_v^2) \propto \frac{1}{\sqrt{2\pi}e^{h_t/2}} \exp\left\{-\frac{y_t^2}{2e^{h_t}}\right\} \times \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left\{-\frac{(v_t - \gamma - h_t)^2}{2\sigma_v^2}\right\},$$

with log density

$$\log p(y_t|v_t, h_t, \gamma, \sigma_v^2) = -\frac{h_t}{2} - \frac{y_t^2}{2e^{h_t}} + \frac{2(v_t - \gamma)h_t - h_t^2}{2\sigma_v^2} + \text{other terms independent of } h_t,$$

such that the first and second derivatives with respect to h_t are

$$f'(h_t) = -\frac{1}{2} + \frac{y_t^2}{2e^{h_t}} + \frac{v_t - \gamma - h_t}{\sigma_v^2}, \quad \text{and} \quad f''(h_t) = -\frac{y_t^2}{2e^{h_t}} - \frac{1}{\sigma_v^2}.$$

This log density and the partial derivatives can be easily substituted in the process outlined in Appendix F and the estimation process is carried out as usual. Further, once $\boldsymbol{h}$ has been estimated, the volatility parameters α, β and σ^2 are also estimated in exactly the same fashion as described in Appendix F.

Sampling the constant term of the realised volatility model

We assume a normal prior for γ such that $\gamma \sim \text{N}(\mu_\gamma, \sigma_\gamma^2)$. Therefore, the posterior distribution of γ is a standard normal conjugate given by

$$\gamma|\boldsymbol{\mathcal{V}}, \boldsymbol{h}, \sigma_v^2 \sim \text{N}\left(\frac{\sigma_\gamma^2 \sum_{i=1}^{T}(v_t - h_t)}{T\sigma_\gamma^2 + \sigma_v^2}, \frac{\sigma_\gamma^2 \sigma_v^2}{T\sigma_\gamma^2 + \sigma_v^2}\right),$$

where it is assumed that $\mu_\gamma = 0$ and $\sigma_\gamma^2 = 0.1$ throughout this thesis.

Sampling the volatility of volatility parameter of the realised volatility model

Finally, we assume an Inverse-Gamma prior for σ_v^2 such that $\sigma_v^2 \sim \text{IG}\left(\frac{a_v}{2}, \frac{b_v}{2}\right)$. Therefore, the posterior distribution of σ_v^2 is the standard inverse gamma conjugate given by

$$\sigma_v^2|\boldsymbol{\mathcal{V}}, \boldsymbol{h}, \gamma \ \sim \text{IG}\left(\frac{T + a_v}{2}, \frac{b_v + \sum_{i=1}^{T}(v_t - \gamma - h_t)^2}{2}\right).$$

7.3.3 Bayesian inference for the BAR-SV model

In this section, we develop a new Bayesian simulator of the BAR model in order to arrive at the model which will be proposed later. The model is

$$y_t = \begin{cases} \phi^U y_{t-1} + \varepsilon_t, & \text{if } R_t = 1, \\ \phi^L y_{t-1} + \varepsilon_t, & \text{if } R_t = 0, \end{cases}$$

$$h_t = \alpha + \beta(h_{t-1} - \alpha) + \eta_t,$$

where $\varepsilon_t \sim \text{N}(0, e^{h_t})$, with the regime indicator

$$R_t = \begin{cases} 1, & \text{if } y_{t-1} \leq r_L, \\ R_{t-1}, & \text{if } r_L < y_{t-1} \leq r_U, \\ 0, & \text{if } y_{t-1} > r_L, \end{cases} \tag{7.11}$$

where ϕ^U and ϕ^L are the AR terms of the upper and lower regime respectively.

Sampling the regime indicator R_0 of the BAR-SV model

In a typical time series model set-up with no threshold effects, the likelihood function of y_t is clearly defined. However, the introduction of AR terms and buffered threshold adds extra difficulty to parameter estimation as the likelihood function depends on the regime indicators R_t which are unobserved. It is clear from the traditional threshold model of Tong (1990) that R_t is determined for $t > 1$ given the data y_{t-1} and the threshold/regime switching variable r but not for R_1 which depends on the initial state variable y_0. Although this y_0 can be estimated from some proposed methodologies or found from records, this problem becomes more complicated in the buffered threshold case when R_1 depends also on R_0 as shown in (7.11). This means that R_1 depends also on whether y_0 crosses a buffer defined by r^U and r^L.

To overcome this problem, we assume $y_0 = 0$ for simplicity and sample R_0 using $R_0 \sim$ Bernoulli($\mathfrak{p}_R$), where

$$\mathfrak{p}_R = \frac{p_1}{p_0 + p_1} \quad \text{with} \quad p_1 = f_N(y_1 | \phi^U, e^{h_1}) \text{ and } p_0 = f_N(y_1 | \phi^L, e^{h_1}),$$

is the density weighted probability of being in the upper regime.

Sampling the threshold levels, r^U and r^L, of the BAR-SV model

In the typical set-up of Tong (1990), there is only one threshold parameter which is estimated using quasi-maximum likelihood via a search of the likelihood space. In order to sample r^U and r^L for the BAR model, we use the MAP sampler on the observational likelihood, which is

$$\begin{aligned}\boldsymbol{Y}|\boldsymbol{Y}_{-T},\boldsymbol{h},\phi^U,\phi^L &= -\frac{T}{2}\log 2\pi - \frac{1}{2}h_1 + \frac{1}{2}\log(1-\phi_1^{*2}) - \frac{1}{2}\frac{(y_1-\phi_1^*)^2}{e^{h_1}} \\ &\quad -\frac{1}{2}\sum_{t=2}^{T} h_t - \frac{1}{2}\sum_{t=2}^{T}\frac{(y_t-\phi_t^* y_{t-1})^2}{e^{h_t}}, \end{aligned} \tag{7.12}$$

where $\boldsymbol{Y} = (y_1,\ldots,y_T), \boldsymbol{Y}_{-T} = (y_1,\ldots,y_{T-1}), \phi_t^* = \phi^U R_t + \phi^L \times (1-R_t)$. The sampling of r^U and r^L are both performed using the MAP sampler, with a normal prior. Specifically, the posteriors of both parameters are

$$\log p_{r^U}(r^U|\boldsymbol{Y},\boldsymbol{h},\phi^U,\phi^L,r^L) = \log f(\boldsymbol{Y}|\boldsymbol{Y}_{-T},\boldsymbol{h},\phi^U,\phi^L,r^U,r^L) + \log N(\mu_{r^U},\sigma^2_{r^U}) + \text{constants}, \tag{7.13}$$

$$\log p_{r^L}(r^L|\boldsymbol{Y},\boldsymbol{h},\phi^U,\phi^L,r^U) = \log f(\boldsymbol{Y}|\boldsymbol{Y}_{-T},\boldsymbol{h},\phi^U,\phi^L,r^U,r^L) + \log N(\mu_{r^L},\sigma^2_{r^L}) + \text{constants}. \tag{7.14}$$

The same MAP principle detailed in Section 3.3.1 applies here also such that:

1. Sample r^{U*} from the proposal distribution $\text{N}(\tilde{r}^U, c^2_{r^U}V_{r^U})$ denoted by $q_{r^U}(\cdot)$ in which $\tilde{r}^U$ is the mode of (7.13), $V_{r^U} = 2$ (similar to V_u in Section 3.3.1) and c_{r^U} is the scaling parameter and Appendix A provides details for the tuning of c_{r^U}.

2. Reject r^{U*} unless $(r^L + \varsigma < r^{U*} < \max\{y_t\})$.[1] Otherwise, accept r^{U*} with acceptance probability ϱ, where

$$\varrho = \min\left\{1, \frac{p_{r^U}(r^{U(m)*}|\boldsymbol{Y},\boldsymbol{h},\phi^U,\phi^L,r^L)q_{r^U}(r^{U(m)})}{p_{r^U}(r^{U(m)}|\boldsymbol{Y},\boldsymbol{h},\phi^U,\phi^L,r^L)q_{r^U}(r^{U(m)*})}\right\}.$$

We then apply a similar procedure to r^{L*}, except reject r^{L*} unless $(\min\{y_t\} < r^L < r^{U*} - \varsigma)$.

[1] ς is a small number that is chosen to avoid classification issues when $|r^U - r^L| \to 0$. We find $\text{sd}(\boldsymbol{Y}) \times 0.05$ is a good choice, where $\text{sd}(\cdot)$ is the standard deviation operator.

Sampling the autoregressive terms ϕ^U and ϕ^L of the BAR-SV model

The sampling of the autoregressive terms of the threshold component is standard. We consider the prior $\phi^U \sim N(\mu_{\phi^U}, \sigma^2_{\phi^U})$ and define $y^*_t = y_t I(R_t = 1)$ and $x^*_t = y_{t-1} I(R_t = 1)$ where $I(E)$ is an indicator function for the event E and zero otherwise. Hence, it can be shown the posterior distribution of ϕ^U is the normal conjugate

$$\phi^U | \boldsymbol{Y}, \boldsymbol{h}, \phi^L, r^U, r^L \sim \mathrm{N}\left(\frac{\sum_{t=1}^{T} x^*_t y^*_t e^{-h_t}}{\sum_{t=1}^{T} x_i^{2*} e^{-h_i} + {\sigma^2_{\phi^U}}^{-1}}, \left(\sum_{t=1}^{T} x_t^{2*} e^{-h_t} + {\sigma^2_{\phi^U}}^{-1} \right)^{-1} \right) \mathbb{1}(|\phi_U| < 1), \tag{7.15}$$

which is also the result of (3.9) in Section 3.3.1 for a single mean function parameter. The case of ϕ^L is also performed in the same way, except when considering the case $R_t = 0$. In all applications throughout this thesis, we assume $\mu_{\phi^U} = 0$ and $\sigma^2_{\phi^U} = 0.2$.

7.3.4 Bayesian inference for the JBAR-SV-GLR model

Finally, we now combine buffer effects, long memory, SV, jumps and realised volatility into one single model called the jump buffered autoregressive stochastic volatility with Gegenbauer log range (JBAR-SV-GLR) model, which forms the basis of Chapter 8.

Model specification

The JBAR-SV-GLR model is given by

$$y_t = \begin{cases} \phi^U y_{t-1} + k_t q_t + \varepsilon_t, & \text{if } R_t = 1, \\ \phi^L y_{t-1} + k_t q_t + \varepsilon_t, & \text{if } R_t = 0, \end{cases} \tag{7.16}$$

$$h_t = \alpha + \beta(h_{t-1} - \alpha) + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \sigma^2), \tag{7.17}$$

$$v_t = (1 - 2uB + B^2)^{-d}(\gamma + h_t + \epsilon_t), \quad \epsilon_t \sim \mathrm{N}(0, \sigma_v^2), \tag{7.18}$$

where $\varepsilon_t \sim \mathrm{N}(0, e^{h_t})$, and the regime indicator R_t is given in (7.11). This is an important model which addresses the issues raised in Section 7.2. The general procedure of estimating this model is to first estimate a certain effect and then consider a detrended series with this estimated effect removed. Subsequently, the remaining parameters can be estimated based on this de-trended series, and so on.

Sampling the regime indicator R_0 of the JBAR-SV-GLR model

We apply the same procedure as Section 7.3.3 to sample the regime indicator R_0 and define the likelihood function.

Sampling the BAR parameters of the JBAR-SV-GLR model

We start off by first defining the jump de-trended returns as $y_t^\star = (y_t - k_t q_t)$. We then estimate the BAR parameters ϕ^U, ϕ^L, r^U, r^L using the same procedures from the BAR-SV model as outlined in Sections 7.3.3 and 7.3.3 but instead replacing y_t with $y_t^\star$.

Sampling the jump parameters of the JBAR-SV-GLR model

Once the parameters in the BAR terms have been estimated, we let $y_t^{(1)} = y_t - \left[\phi^U y_{t-1} R_t + \phi^L y_{t-1}(1 - R_t)\right]$, which is the de-trended returns using parameters in the BAR model. The usual derivations from section 7.3.1 are now applied to estimate the jump parameters k_t, q_t and κ.

Sampling the stochastic volatility parameters of the JBAR-SV-GLR model

In a similar fashion, let $y_t^{(2)} = y_t^{(1)} - k_t q_t = \varepsilon_t$. The process is carried out exactly in Section 7.3.2 with $y_t^{(2)}$ replacing y_t.

Sampling the Gegenbauer long memory parameters

We reconsider the realised volatility component in (7.18). This can be written out in matrix notation as

$$\boldsymbol{\mathcal{V}} = \boldsymbol{G}_J \boldsymbol{\mu}_v + \boldsymbol{G}_J \boldsymbol{\epsilon}$$

where

$$\boldsymbol{\epsilon} \sim \mathrm{N}(0, \boldsymbol{\Sigma}_v), \quad \boldsymbol{\Sigma}_v = \mathrm{diag}\left(\sigma_v^2, \ldots, \sigma_v^2\right), \quad \boldsymbol{\mu}_v = \gamma \mathbf{1} + \boldsymbol{h},$$

and $\mathbf{1}$ is a vector of 1s so that

$$\boldsymbol{\mathcal{V}}|\boldsymbol{G}_J, \boldsymbol{\mu}_v, \boldsymbol{\Sigma}_v \sim \mathrm{N}(\tilde{\boldsymbol{\mu}}_v, \boldsymbol{G}_J\boldsymbol{\Sigma}_v\boldsymbol{G}_J'),$$

where $\tilde{\boldsymbol{\mu}}_v = \boldsymbol{G}_J\boldsymbol{\mu}_v$. Hence

$$f(\boldsymbol{\mathcal{V}}|\boldsymbol{h}, u, d, \gamma, \sigma_v^2) = (2\pi)^{-\frac{T}{2}}|\boldsymbol{G}_J\boldsymbol{\Sigma}_v\boldsymbol{G}_J|^{-1}\exp\left\{-\frac{1}{2}(\boldsymbol{\mathcal{V}} - \tilde{\boldsymbol{\mu}}_v)'(\boldsymbol{G}_J\boldsymbol{\Sigma}_v\boldsymbol{G}_J')^{-1}(\boldsymbol{\mathcal{V}} - \tilde{\boldsymbol{\mu}}_v)\right\}. \quad (7.19)$$

The Gegenbauer parameters (u, d) are now estimated using the MAP sampler once again, based on the realised volatility likelihood given in (7.19). The complete posteriors for both u and d are

$$\log p_u(u|\boldsymbol{h}, d, \gamma, \sigma_v^2) = \log f(\boldsymbol{\mathcal{V}}|\boldsymbol{h}, u, d, \gamma, \sigma_v^2) + \log \mathrm{N}(\mu_u, \sigma_u^2)\mathbb{1}_{ud},$$
$$\log p_d(d|\boldsymbol{h}, u, \gamma, \sigma_v^2) = \log f(\boldsymbol{\mathcal{V}}|\boldsymbol{h}, u, d, \gamma, \sigma_v^2) + \log \mathrm{N}(\mu_d, \sigma_d^2)\mathbb{1}_{ud}.$$

where $\mathbb{1}_{ud}$ is defined in Section 3.3.1. The same MAP principle detailed in Section 3.3.1 also applies here.

Sampling the realised volatility parameters

Again, we consider the transformation $\boldsymbol{\mathcal{V}}^* = \boldsymbol{G}_J^{-1}\boldsymbol{\mathcal{V}} = \boldsymbol{\mu}_v + \boldsymbol{\epsilon}$. It then becomes straightforward to estimate γ and σ_v^2 by using the same methodology of estimating respectively α and σ^2 as outlined in Section 7.3.2 with $\boldsymbol{\mathcal{V}}^*$ replacing $\boldsymbol{Y}^*$.

7.4 Conclusion

Following the model development explored in this chapter and subsequently deriving the Bayesian MCMC sampler of the JBAR-SV-GLR model, the next chapter will see the application of this model to Cryptocurrency returns. This is paramount to overcome the shortcomings of the models proposed in Chapters 5 and 6. The whole chapter is our third publication which is to appear in *Finance Research Letters*.

Chapter 8

On long memory effects in the volatility measure of Cryptocurrencies

"If you torture the data enough, nature will always confess."

Ronald Coase

Cryptocurrencies as of late have commanded global attention on a number of fronts. Most notably, their variance properties are known for being notoriously wild, unlike their fiat counterparts. We highlight some stylized facts about the variance measures of Cryptocurrencies using the logarithm of daily return range and relate these results to their respective cryptographic designs such as intended transaction speed. The results favor oscillatory long run autocorrelations over standard long run autocorrelation filters to model the log daily return range. The overarching implication of this result is the volatility of Cryptocurrencies can be better understood and measured via the use of fast moving autocorrelation functions, as opposed to smoothly decaying functions for fiat currencies.

8.1 Introduction

Financial controllers globally are now at a cross road of accepting Cryptocurrencies as a medium of exchange, or purely as a speculative alternative asset class. In this note, a time series model is used to further address such issues, by providing a novel approach to better understand their unique volatility properties.

As of late, there has been an emergence of methods attempting to explain the long run autocorrelation properties of Cryptocurrencies, particularly Bitcoin. For instance, Jiang, Nie, and Ruan, 2017 finds evidence of a standard long run autocorrelation in Bitcoin returns only, but does not consider coupling this finding with the possibility of time-varying daily volatility. Time-varying daily volatility models are appealing as they are intuitive and capture more empirical properties compared to non-time varying volatility models, but are often avoided because they are difficult to estimate. Lahmiri, Bekiros, and Salvi, 2018 models long run autocorrelations in the daily time varying volatility component itself, but they do not consider the unique long run trend behaviors of Cryptocurrencies such as jumps.

In this work a model is proposed extending the work of Phillip, Chan, and Peiris (2018) by the inclusion of daily time varying volatility measures which have been shown to greatly improve model performance (Koopman, Jungbacker, and Hol, 2005). This is especially true when market nuances such as time-of-day effects are present. Incorporating volatility measures into financial time series models, such as the CBOE Volatility Index (VIX), have gained traction as of late since they are efficient measures of the true volatility and reduce model error. Given the extreme volatility of Cryptocurrencies, such measures are extremely valid and a worthwhile pursuit. Hence, including such measures in the time series model is an important step in the volatility estimation process and by not doing so, a weaker signal about the current level of volatility is obtained. This is the case for the Stochastic Volatility (SV) model of Taylor (1986), one of the most commonly used models to measure daily time-varying volatility, as the model assumes the volatility process is latent based on the information of returns. Imposing the limitations of a weaker volatility signal within the context of Cryptocurrencies is extremely debilitating since they are notorious for wild volatility characteristics. Takahashi, Omori, and Watanabe (2009) first suggested incorporating additionally realised volatility into the SV model to supplement such a limitation in estimating time varying volatility and the model is referred to as the realised Stochastic Volatility (RSV) model.

Traditionally, realised volatility is defined as the sum of squared intraday returns over a specific time interval (Andersen and Bollerslev, 1998; Barndorff-Nielsen and others, 2001). The purpose of including such a volatility measure is to provide a robust estimator to filter the volatility component of the time series. Although it

is a popular choice, the use of realised volatility as an additional measure suffers a significant drawback of being dependent on the intraday sampling interval which can bias the results. Alizadeh, Brandt, and Diebold (2002) find that daily range based volatility measures are highly efficient extracts of structural volatility components and are robust to market microstructure noise.

The presence of long run autocorrelation in range based volatility measures has been known to exist within some financial assets such as stocks and currencies. A routine solution to modelling this is to include a standard long run autocorrelation filter, which assumes exponential decay over time (Raggi and Bordignon, 2012; Corsi, 2009; Koopman, Jungbacker, and Hol, 2005). This is due to the fact that long run autocorrelation persistence in fiat assets is slowly decaying, with no oscillatory behavior. We however find a completely different case for the range based volatility measure of Cryptocurrencies which show oscillatory long run autocorrelation behavior in general. These sporadic long run autocorrelations can be measured using suitable Gegenbauer long run autocorrelation filters, which are able to capture oscillatory behaviors.

However, one of the main criticisms of long run autocorrelation estimation in general is that such effects may indeed be confused for regime changes in the long run trend component; see Guégan (2005) for a detailed review. We respond to such criticisms by incorporating for the first time, the so-called Buffered Autoregressive (BAR) model of Zhu, Yu, and Li, 2014 in conjunction with the time varying SV model of Taylor, 1986. By doing so, we justify the use of the long-run autocorrelations together with structural changes. In addition to including BAR effects, we also simultaneously allow for occasional jumps, as are often reported about Cryptocurrencies. These jumps are assumed to occur in the long-run trend component.

The aim of this note is to advance Cryptocurrency models by addressing the above mentioned issues with respect to their wild volatilities. Our proposed model is novel in the Crypotocurrency literature in three aspects: an additional daily range volatility model within the SV model structure, Gegenbauer long run autocorrelation filters for the volatility measure and the BAR model with jumps in the trend model. We assume persistence in the volatility measures rather than returns as the persistence in returns can be distorted by jumps - especially Bitcoin. Our model can capture any jump features in the returns so that the persistence in volatility can be more

easily detected. This advanced model not only provides substantial improvement in model performance but also offers new implications about the volatility features of Cryptocurrencies.

The remainder of this note is organised as follows: in Section 2, we discuss the data source and the model; Section 3 discusses empirical findings and concludes with Section 4.

8.2 Data and Methodology

The Cryptocurrency data is sourced from the Brave New Coin (BNC) Digital Currency indices database. Currently, there are more than 2, 800 Cryptocurrency index series available on the BNC database. However, some of these have market capitalizations which are small ($< \$1,000,000$ USD) and traded very little. After filtering out for a meaningful investable basket, this leaves a total of 149 Cryptocurrencies and the inception date of these time series vary but all end on the 31st of December, 2017.

We assume in this note that Cryptocurrency behaviour can be decomposed into two components. The first being a long run trend (or mean) of the time series y_t defined as the daily index price percentage change $y_t = (P_t - P_{t-1})/P_{t-1}$ (conceptually similar to return) where P_t is the daily index value at time (day) t. The second component is the volatility measure, the log daily return range, which is defined as

$$v_t = \log(R_{h,t} - R_{l,t}), \tag{8.1}$$

where the high and low daily return on day t are $R_{k,t} = (P_{k,t} - P_{c,t-1})/P_{c,t-1}, k = h, l$ respectively and $P_{k,t}, k = h, l, c$ represents the high, low and closing price of day t. We use this particular definition because it is guaranteed to have support that agrees with the normal distribution. Our model attempts to explain the behavior of this daily super imposed volatility component of Cryptocurrencies.

Cryptocurrencies are also plagued with a host of other competing issues due to their infrastructure set-up. To address the issues, the long run trend component in y_t is assigned to a buffered threshold model with different jump features in each structural component. Moreover, the autocorrelation functions (ACFs) of the log daily

return range volatility for the top six Cryptocurrencies displayed in Figure 8.1 confirm the presence of oscillating long run autocorrelations. This oscillatory behavior strongly suggests the use of a Gegenbauer long run autocorrelation filter to properly estimate such effects. This extended model makes it a natural contender by correctly measuring these oscillatory effects in the presence of a large investable universe of Cryptocurrencies.

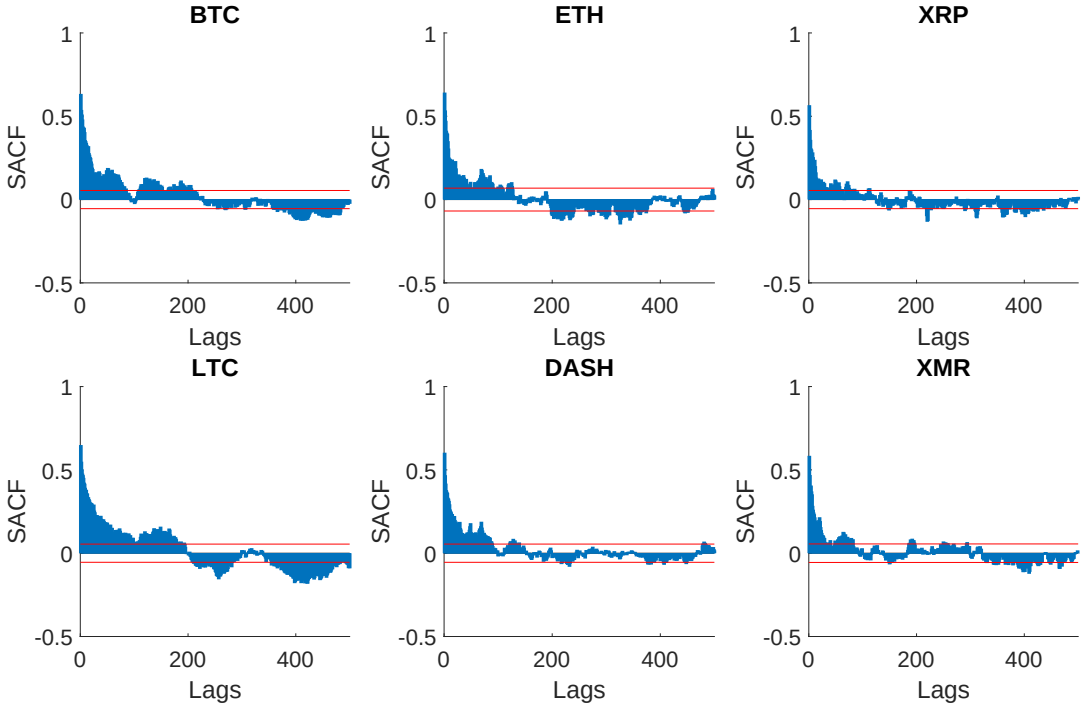

FIGURE 8.1: Sample ACF plots of the log daily return range of the 6 largest Cryptocurrencies measured by market capitalization on 31/12/2017. BTC: Bitcoin. ETH: Ethereum. XRP: Ripple. LTC: Litecoin. DASH: Dash. XMR: Monero.

We describe our proposed Jump BAR SV Gegenbauer Log Range (JBAR-SV-GLR) model below. Let the return $y_t,\ t = 1, 2, \ldots, T$ and its volatility measure $v_t,\ t = 1, 2, \ldots, T$ satisfy the equations

$$\text{JBAR:} \quad y_t = \begin{cases} \phi^U y_{t-1} + k_t q_t + \varepsilon_t, & \text{if } R_t = 1, \\ \phi^L y_{t-1} + k_t q_t + \varepsilon_t, & \text{if } R_t = 0, \end{cases} \tag{8.2}$$

$$\text{SV:} \quad h_t = \alpha + \beta(h_{t-1} - \alpha) + \eta_t, \quad \eta_t \sim \text{N}(0, \sigma^2), \tag{8.3}$$

$$\text{GLR:} \quad (1 - 2uB + B^2)^d v_t = \gamma + h_t + \epsilon_t, \quad \epsilon_t \sim \text{N}(0, \sigma_v^2), \tag{8.4}$$

where $\varepsilon_t \sim \text{N}(0, e^{h_t})$, and the buffer regime indicator and initial model at $t = 1$ are respectively

$$R_t = \begin{cases} 1, & \text{if } y_{t-d} \leq r_L, \\ R_{t-1}, & \text{if } r_L < y_{t-d} \leq r_U, \\ 0, & \text{if } y_{t-d} > r_L, \end{cases} \quad y_1 \sim \begin{cases} \mathrm{N}\left(\frac{k_1 q_1}{1-\phi_1}, \frac{e^{h_1}}{1-\phi_1^2}\right) & \text{if } R_1 = 0, \\ \mathrm{N}\left(\frac{k_1 q_1}{1-\phi_2}, \frac{e^{h_1}}{1-\phi_2^2}\right) & \text{if } R_1 = 1. \end{cases}$$

The jump indicator $q_t \in \{0, 1\}$ has probability of jumping equal to $\mathbb{P}(q_t = 1) = \kappa$ and the jump size $k_t \sim \mathrm{N}(\mu_k, \sigma_k^2)$. It is known the volatility measure v_t has long run autocorrelation effects when $(\{|u| < 1, 0 < d < 0.5\} \cup \{|u| = 1, 0 < d < 0.25\})$. Additionally, γ is the level of the volatility measure, and σ_v^2 is the volatility of the volatility measure. When $u = 1$, v_t has standard long run autocorrelation effects such that equation 8.4 becomes $(1 - B)^{2d} v_t = \gamma + h_t + \epsilon_t$. The volatility component h_t in the SV model evolves according to the state equation 8.3 for $t = 1, \ldots, T$, α is the constant level of the volatility, β is the persistence of the volatility process and σ^2 is the volatility of the volatility process. We assume $|\beta| < 1$ so h_{t+1} is not explosive.

This model allows the return y_t to experience buffered regime changes (equation 8.2) and the logarithm of the volatility component to have an autoregressive structure (equation 8.3). It also relates the volatility measure v_t via another linear model with Gegenbauer long run autocorrelation filter (equation 8.4).

8.3 Empirical results

In order to contest the current literature of utilizing a standard long run autocorrelation specification to model volatility, we estimate model 8.2-8.4 and also its special case, which is in fact the standard case by setting $u = 1$ in equation 8.4. Results are reported in Table 1 for the six largest Cryptocurrencies measured by market capitalization.

Data	Model	r^U	r^L	ϕ^U	ϕ^L	u	d	α	β	σ^2	γ	σ_v^2	κ	μ_k	σ_k^2	DIC
BTC	Standard															
	$\hat{\theta}$	0.0535	-0.2036	0.0222	-0.6498		0.0005	-7.5249	0.6539	0.4106	4.0328	0.0049	0.0010	0.0241	2.0379	-8810
	Std.	0.0829	0.0059	0.0227	0.1706		0.0003	0.0640	0.0208	0.0163	0.0372	0.0019	0.0011	0.0917	1.5542	
	AR(%)	0.1435	0.0322				0.0068		0.3988					0.1270	0.3648	
	GR	1.0001	1.0319	1.0000	1.0011		1.1945	1.0005	1.0013	1.0004	1.0201	1.0120	1.0006	1.0000	1.0001	
BTC	Gegenbauer															
	$\hat{\theta}$	0.0032	-0.1900	0.0209	-0.2077	-0.7717	0.2106	-7.5493	0.7933	0.4003	2.9470	0.0048	0.0009	0.0025	2.2259	-8874
	Std.	0.0853	0.0767	0.0850	0.2319	0.0157	0.0080	0.0906	0.0171	0.0159	0.0630	0.0016	0.0010	0.0475	1.5020	
	AR(%)	0.1233	0.1498			0.1308	0.0759		0.3776					0.1239	0.3812	
	GR	1.0054	1.0009	1.0030	1.0002	2.0505	1.0054	1.0037	1.0011	1.0036	1.6666	1.0001	1.0008	1.0009	1.0000	
ETH	Standard															
	$\hat{\theta}$	0.0121	-0.1252	0.0384	-0.3803		0.0004	-6.0038	0.6765	0.3707	3.2992	0.0070	0.0016	0.0232	2.1043	-4304
	Std.	0.0447	0.0189	0.0310	0.1116		0.0004	0.0824	0.0259	0.0191	0.0484	0.0031	0.0019	0.1219	1.5275	
	AR(%)	0.1378	0.1237				0.0122		0.3982					0.1414	0.3729	
	GR	1.0005	0.9999	0.9999	1.0001		1.0140	1.0714	1.0005	1.0002	1.2345	1.0176	1.0002	0.9999	1.0001	
ETH	Gegenbauer															
	$\hat{\theta}$	-0.0012	-0.1198	0.0338	-0.3570	-0.7920	0.2397	-6.0833	0.8157	0.3707	2.3768	0.0057	0.0017	0.0204	2.0976	-4365
	Std.	0.0468	0.0285	0.0347	0.1393	0.0453	0.0068	0.1265	0.0205	0.0188	0.0658	0.0021	0.0022	0.0752	1.5533	
	AR(%)	0.1192	0.0896			0.1135	0.1245		0.3760					0.1511	0.3733	
	GR	0.9999	0.9999	0.9999	0.9999	3.1808	1.0415	1.0029	0.9999	1.0198	1.0027	1.0413	1.0000	0.9999	1.0000	
XRP	Standard															
	$\hat{\theta}$	0.0194	-0.1325	-0.0301	-0.1791		0.0008	-6.4377	0.5953	0.4701	3.6175	0.0066	0.0039	0.8756	0.9691	-6762
	Std.	0.0571	0.0409	0.0264	0.0599		0.0005	0.0603	0.0223	0.0188	0.0329	0.0027	0.0032	0.5179	1.0500	
	AR(%)	0.1436	0.1049				0.0115		0.3915					0.1705	0.2863	
	GR	0.9999	1.0005	0.9999	1.0002		1.2884	1.0464	1.0030	0.9999	1.0371	1.0054	1.0106	1.0100	1.0037	
XRP	Gegenbauer															
	$\hat{\theta}$	-0.0408	-0.1028	0.0057	-0.1833	-0.7630	0.2267	-6.5116	0.7547	0.4773	2.6783	0.0051	0.0023	1.1437	1.2943	-7349
	Std.	0.0482	0.0344	0.0280	0.0718	0.0096	0.0106	0.0834	0.0192	0.0189	0.0716	0.0016	0.0017	0.4868	1.1905	
	AR(%)	0.1417	0.1431			0.1313	0.0894		0.3655					0.1638	0.2941	
	GR	0.9999	1.0004	1.0026	0.9999	1.0000	1.8408	0.9999	1.0385	1.0074	1.4316	1.0036	1.0017	1.0019	1.0000	
LTC	Standard															
	$\hat{\theta}$	0.0720	-0.2166	-0.0208	-0.0911		0.0011	-7.1119	0.6720	0.4882	3.8409	0.0055	0.0122	0.2981	0.0137	-7394
	Std.	0.0957	0.1058	0.0327	0.1135		0.0006	0.0823	0.0204	0.0194	0.0609	0.0021	0.0047	0.0475	0.0230	
	AR(%)	0.1532	0.1474				0.0105		0.3402					0.1420	0.3431	
	GR	1.0003	1.0003	1.0000	0.9999		1.4556	1.0001	1.0035	1.0007	1.0368	1.0001	1.0008	1.0000	1.0000	
LTC	Gegenbauer															
	$\hat{\theta}$	0.0008	-0.0767	-0.0102	-0.0886	-0.8512	0.2819	-7.2094	0.8395	0.5014	2.3199	0.0045	0.0102	0.2594	0.0216	-7961
	Std.	0.0994	0.1077	0.0481	0.1338	0.0043	0.0067	0.1309	0.0154	0.0198	0.0423	0.0015	0.0041	0.0537	0.0163	
	AR(%)	0.1461	0.1673			0.0490	0.0577		0.2814					0.1274	0.3132	
	GR	0.9999	1.0000	0.9999	1.0011	2.2859	2.4969	1.0363	1.0096	1.0045	1.0004	1.0124	1.0304	1.0441	1.0025	
DASH	Standard															
	$\hat{\theta}$	0.0164	-0.1961	0.0143	-0.3646		0.0008	-6.1169	0.6278	0.3195	3.4832	0.0056	0.0038	0.5961	0.4318	-6528
	Std.	0.0833	0.0649	0.0236	0.2271		0.0005	0.0577	0.0216	0.0127	0.0425	0.0021	0.0029	0.2433	0.8365	
	AR(%)	0.1310	0.1468				0.0108		0.3948					0.1240	0.2558	
	GR	1.0004	1.0302	1.0002	1.0052		1.0356	1.0242	1.0015	1.0005	1.0686	1.0050	1.0030	1.0014	1.0003	
DASH	Gegenbauer															
	$\hat{\theta}$	-0.0045	-0.1960	0.0031	-0.2642	-0.7501	0.3284	-6.1748	0.8180	0.3397	2.0001	0.0042	0.0011	0.0839	2.0251	-6954
	Std.	0.0880	0.0862	0.0310	0.2624	0.0049	0.0097	0.0936	0.0163	0.0137	0.0696	0.0012	0.0013	0.2088	1.5468	
	AR(%)	0.1587	0.1425			0.0866	0.0637		0.3910					0.1698	0.3642	
	GR	1.0000	1.0001	1.0000	1.0003	1.0199	2.4144	1.0120	1.0353	1.0256	1.7588	1.0010	1.0001	1.0003	0.9999	
XMR	Standard															
	$\hat{\theta}$	-0.0229	-0.1885	-0.0188	-0.4546		0.0012	-5.7350	0.6119	0.2543	3.3578	0.0052	0.0022	0.4239	0.9184	-6033
	Std.	0.0816	0.0480	0.0250	0.1888		0.0008	0.0570	0.0229	0.0104	0.0443	0.0019	0.0018	0.2858	1.3065	
	AR(%)	0.1423	0.1632				0.0122		0.3656					0.1474	0.2673	
	GR	1.0002	1.0024	1.0001	1.0013		1.6475	1.0001	1.0058	0.9999	1.0546	1.0013	0.9999	1.0041	1.0001	
XMR	Gegenbauer															
	$\hat{\theta}$	0.0033	-0.1873	-0.0171	-0.3320	-0.8601	0.1727	-5.7598	0.7453	0.2546	2.7331	0.0045	0.0012	0.0974	1.9898	-6192
	Std.	0.0798	0.0680	0.0294	0.2009	0.0240	0.0168	0.0672	0.0204	0.0106	0.0750	0.0016	0.0014	0.2333	1.5438	
	AR(%)	0.1369	0.1438			0.1383	0.1133		0.3332					0.1401	0.3635	
	GR	1.0013	1.0000	1.0002	1.0003	1.1686	1.0039	1.0333	1.0003	0.9999	1.0754	1.0018	1.0005	1.0000	1.0000	

TABLE 8.1: Parameter estimates for each dataset under both model specifications.

The main points of interest from Table 8.1 are the long run autocorrelation parameters, u and d. Both models share a d parameter, which measures the strength of

long run autocorrelations present in the volatility measures. As shown, the standard model does not show any significance of long run autocorrelations, as most estimates of d are close to 0. This is in stark contrast, however, with the Gegenbauer model, as most estimates of d are around 0.25. Since d is limited to the range $[0, 0.5]$, these estimates of d are economically significant. The parameter u is limited between the range $[-1, 1]$ and measures the level of oscillation in the long run ACF of the volatility measure. The closer u is to -1, the more oscillatory the ACF is, and $u = 1$ means there is no long run autocorrelation oscillation. All values of u are statistically and economically significant for the Gegenbauer filter, with most values being estimated around -0.7. After allowing for the long run autocorrelation structure, most trend components for y_t do not possess jump behavior, even for BTC which displays sporadic y_t. The only exceptions are LTC and XRP. This interesting finding also confirms the necessity of modelling persistence in the volatility measures rather than returns as in most Cryptocurrency models.

We next expand the analysis to a large and practically investable universe of 149 Cryptocurrencies, and provide an intuitive handle on the results. We note the DICs reported in Table 8.1 measure the model misfit and hence a lower DIC (more negative) indicates better model fit. In order to gauge a broad overview of the data, the DIC ratios of the standard model ($u = 1$) to the Gegenbauer model for all 149 Cryptcurrencies are measured. These DIC ratios which we call 'volatility oscillation memory ratios' (VOMRs) are powerful metrics that provide a deeper understanding on the properties of Cryptocurrencies: a VOMR greater than one indicates that higher model misfit for the standard model relative to the Gegenbauer model and hence a preference for the Gegenbauer model over the standard model. Figure 8.2 depicts the density plot of the VOMRs for all 149 Cryptocurrencies. In total, 118 (79%) of the Cryptocurrencies have a VOMR which is considered high (greater than one), compared to only 31 (21%) which are low.

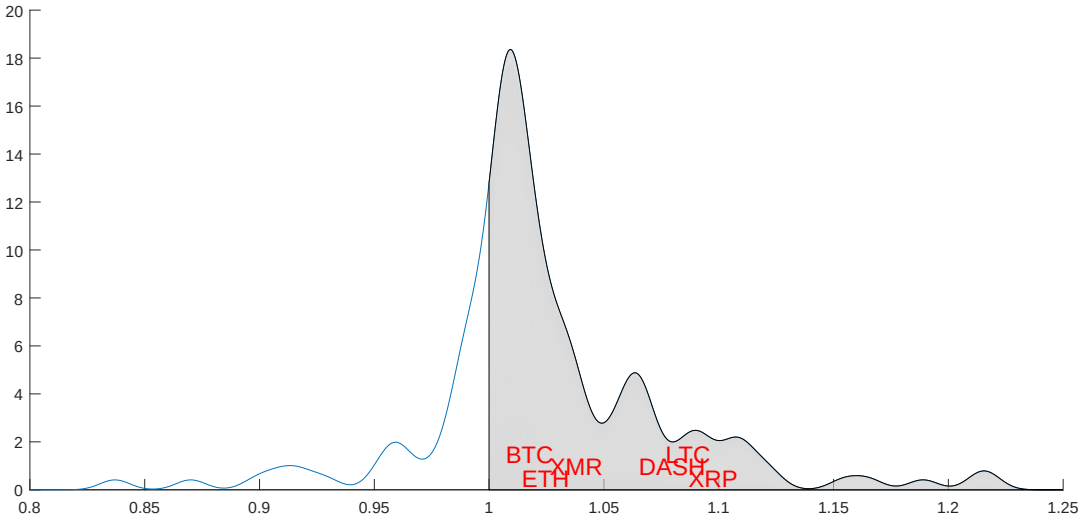

FIGURE 8.2: Density plot of VOMRs of Gegenbauer to standard long run autocorrelation filter for log daily return range. The top six Cryptocurrencies by market capitalization are overlaid.

It is clear from Figure 8.2 that all of the top six Cryptocurrencies have a VOMR greater than one. Upon closer inspection, it is evident the VOMR is closely related to completion time (transaction speed). This completion time issue presents new challenges which are not present in fiat currencies. One of the pioneering aspects of Cryptocurrencies is the use of Blockchain technology; which can be intuitively interpreted as a clearing house for transactions. Arguably, the most appealing aspect of such technology is that transactions are intended to be almost instantaneous and have a negligible bid-ask spread. This feature is very different from fiat currencies, which do have these market frictions. The most commonly discussed example where this would benefit the most is within the international money transfer services community (such as Western Union) in which there is a clear need to send cash overseas very cheaply and instantly.

The intuitive relationship of VOMR with completion time seems to explain how the day-to-day volatility correlation is dependent on completion times, and therefore liquidity. To illustrate the transaction speed (and hence liquidity issue), we look at the top six Cryptocurrencies. It is commonly known that BTC, ETH and XMR have long completion times, compared to LTC, DASH and XRP which have shorter completion times. For example, BTC, the largest and most widely traded Cryptocurrency today, can take up to two days to transact, whereas XRP takes only seconds. This intuitive understanding helps to settle important speculative debates over Cryptocurrencies as it reveals their transaction speeds, an important factor to the role of currency, are related to the oscillatory long run autocorrelation structure of their volatility measures.

Regarding transaction speed, BTC receives the biggest criticism as its infrastructure set-up was not designed to handle such a large volume of trades that it currently experiences. As such, critics argue that it is not a sustainable Cryptocurrency, since it now has extremely slow transaction speeds and is therefore not a long-term viable solution. This can shed light from its close-to-one VOMR showing weak preference for Gegenbauer specification. Another interesting finding is the VOMR of ETH being again close to one. ETH claims to have embedded "Smart Contracts" to circumvent the slow transaction fallacy of BTC. However in reality, the transaction time of ETH has also increased considerably due to a lack of infrastructure upgrades to deal with growing pains. XMR is a coin which mainly focuses itself on security and privacy, but not on speed. This finding too is evidenced on the chart, since it has a VOMR close to one. These three cryptocurrencies as a group are in sharp contrast to LTC, DASH and XRP who pride themselves on having faster transactions (almost instant), and this is indeed the case depicted in Figure 8.2 as they are clustered on the right of the chart.

An extremely important example which further illustrates the relationship between transaction speed and preference for Gegenbauer specification is XRP which is one of the most popular Cryptocurrencies and has one of the highest VOMRs (1.09). XRP is the most commonly used Cryptocurrency by financial institutions since there is virtually no overnight risk. By design of the cryptographic integrity of XRP, there is no positive dampening correlation across time for overnight risk, and is therefore very liquid. XRP is now the preferred Cryptocurrency used by large banks and is the main Cryptocurrency used by banks to connect with other banks, with an emphasis on almost instantaneous transaction speeds of up to apparently 17 seconds, compared to traditional transaction times. As such, financial institutions now routinely convert fiat to XRP for liquidity.

8.4 Conclusions and future research

This note addresses several important issues. Digital assets present challenges which are unlike their fiat counterparts, and require specific treatment. Previous debates on the role of Cryptocurrencies mainly focus on measuring their volatility, but do not provide a practical handle on the broader financial implications of this.

We label a trend that stronger oscillating long run volatility autocorrelations are associated with shorter transaction times. Upon closer observation of the top six Cryptocurrencies, it is found that slower transacted Cryptocurrencies, such as Bitcoin, have less oscillatory features (VOMR$\approx$1) whereas faster transacted coins, such as Ripple (VOMR$>$1), display oscillatory features. As faster transacted Cryptocurrencies have lower liquidity risk during transactions, these are more preferable purely as a medium of exchange. This trend of oscillatory long run autocorrelations and transaction time is important and has broader practical implications to investors and policy regulators as it provides an alternative tool to explain the speculative nature of Cryptocurrencies based on their volatility measures. Finally, it is confirmed the long run autocorrelation patterns found in Cryptocurrency time series are not regime changes; and their investigation should be orientated through their realised volatility measures instead of returns, contrary to current methodologies. Future avenues of research include conditioning on further stylized facts, such as the leverage effect, or fat-tails.

Chapter 9

Conclusion

“All good things must come to an end, but all bad things can continue forever."

Thornton Wilder

9.1 Contributions: addressing the motivations

In this thesis, we have addressed our motivations detailed in Chapter 1 by providing three key contributions.

Our first contribution is to extend the current literature by creating a new model class which couples long memory and stochastic volatility. The literature that currently exists is rich in long memory, or stochastic volatility errors but rarely both in one model. Our modelling contributions are first corner-stoned with Chapter 3 in which we describe the basic GMA-SV model. This chapter supplements the current literature by providing robust parameter estimates of long memory in the observation equation and also in the SV equation. This was evidenced by applying this model to the US CPI and the US ERP and it was found to outperform relative to competing models. The US CPI has always been synonymous within the long memory literature since it is a summation of first order Markov processes, yet its residuals also display SV features. The GMA-SV model combines the best of both of these worlds to arrive at an efficient solution.

Additionally, with the advent of digital currencies, the econometric literature must move quickly in order to adapt to this new phenomenon which has commanded global attention. It is found that Cryptocurrencies show a whole host of effects which require further advanced modelling techniques. Specifically, we consider two further

extensions of the GMA-SV model in order to adequately measure some unique properties of Cryptocurrencies - most notably their wild volatilities. The first extension proposes the GMA-SV model with leverage and heavy tails. The GMA-SV model is further refined with the second extension, the JBAR-SV-GLR model, and it is found that estimation improves with greater modelling efforts.

Our second contribution to the literature is the creation of Bayesian techniques in order to estimate these new models. It should be duly noted there is a distinct difference from deriving a posterior distribution and generating posterior samples. In the traditional sense, most Bayesian models are typically programmed, such that the need to derive the complete posterior distribution and also to design a sampler is made redundant. Although this has helped academics to sample complex hierarchical models quickly and efficiently, there are far more efficient sampling methods available that require individual assessment. For example, the leverage model in Yu (2005) is programmed using the OpenBUGS Bayesian package, so there is no need to manually derive the posterior distribution or design a suitable sampler. However, the estimation of the correlation coefficient, ρ, can be further improved, and estimated with greater accuracy by tuning, and applying the MAP sampler. This particular example is evidenced in section 5.4.2 where our estimator of ρ is much more accurate in comparison to the OpenBUGS sampler used in Wang, Chan, and Choy (2011) and Choy and Chan (2000). Also, by deriving the posterior distributions and discussing how they are derived in a systematic manner, our approach opens a workable and straightforward method as a gateway for future research.

Our estimation techniques are rigorously tested throughout this thesis via simulation studies, and proven to be successful. By conducting these simulation studies we are able to pin-point potential discrepancies within our simulator, then improve and recheck the newly improved simulator. Specifically, we are able to reduce the MSE of parameter estimates in a synthetic simulation study, improve acceptance rates, and lower computational time. For example, to the best of our knowledge, the Bayesian estimation literature of the BAR model is non-existent and the estimation is conducted using classical approaches. By borrowing the ideas of the MAP estimator, which was rigorously tested in Chapter 3 on the estimation of GMA-SV model, we are able to quickly and efficiently apply this MAP estimator to estimate the parameters of the BAR model. In general, the BAR model can now be conditionally implemented

into a host of other models by using our framework.

Lastly, our third contribution is the facilitation of rich and interesting applications which provide discussion points for a host of stakeholders. The US ERP is an important time series, as it is considered a receptacle of investor risk appetite. Although rarely discussed within the econometric literature, it is an important time series which deserves further attention. It is shown in Figure 3.5 that by using a standard ARFIMA model to detect the presence of long memory, the persistence parameter is estimated to be close to 0. However, when applied using the GARFIMA filter, the persistence parameter is found to be strong, persistent and almost constant over time. This observation alone is a powerful finding since it reveals the US ERP has long-run persistent behavior and therefore strongly supports the idea of a mean reverting market structure. This shows the superiority of using the GARFIMA-SV model over the standard ARFIMA-SV model, and when also applied to forecasting shows superior results as measured by the log Bayes factor (Figure 3.6).

Finally, the most significant applications discussed throughout this thesis are on digital currencies. Digital currencies operate in a completely different fashion to what is considered the norm, and demand further inspection. Their main unique features are their unparalleled volatility characteristics, which are very different to traditional financial currencies. The current attitude towards Cryptocurrencies is that they are extremely speculative in nature. This is due to the fact that it is not unusual for them to move, for example, $\pm 10\%$ in any given day. These unique and peculiar features demand the use of more sophisticated econometric techniques which are seldom used in unison. For example, the combined use of the bivariate Student's t-distribution in unison with leverage effects, long memory and SV rarely occurs in the literature, since it is not common to find such data. However, these features are all too common within digital assets and are therefore suitable for their estimation. Further, Cryptocurrencies also display clear oscillatory ACFs and jumps which make them valuable archetypes for our second extended model. For example, we find that Ripple which has no overnight risk has a near zero leverage effect. Moreover, although Cryptocurrency returns seem to behave wildly without restraint, jumps are mostly non-significant after allowing for other effects. In summary, these findings on jumps and long memory, whether in returns or volatilities, indicate that Cryptocurrencies are in fact more predictable than expected.

9.2 Future potential research

The findings of this thesis can be extended in several ways; accordingly, we first discuss some limitations of our work that can be addressed, and conclude with potential avenues for further research.

Undoubtedly, our work has close parallels with the high frequency data literature, and a natural extension is the consideration of high frequency Cryptocurrency data. In some situations however, the JBAR-SV-GLR model can take up to 12 hours to run for one dataset at $20,000$ loops where the number of observations, $T = 2,000$. In the high frequency case, the number of observations could potentially be much larger. The reason for this long computation time is due to the evaluation of the likelihood, in particular, the size of the $\boldsymbol{G}_J$ matrix. Although this matrix is sparse, it forms a part of the calculation of the quadratic term in the likelihood and is a very computationally expensive exercise. As there are techniques to reduce the computation for sparse matrices (Tropp and Wright, 2010), it would be a worthwhile pursuit to implement these to significantly reduce computational time. Another potential avenue to address some limitations in our research is the exploration of alternative sampling techniques. The individual sampling procedures for each parameter can be compared to other sampling methodologies in order to ascertain their relative efficiency.

9.2.1 Alternative mean structures

The findings in this thesis are rich and iteratively built up from basic concepts. Hence, it is relatively straightforward for future researchers to supplement (or remove) any model features whenever deemed suitable for a particular data set. The first way is to potentially investigate alternative mean structures. There are an abundance of mean structures that are outside the scope of this thesis and are applicable to the observation equation, latent volatility equation or even realised volatility equation. For modelling trend movements, ARMA(p, q) models, mean smoothers (such as splines), or even advanced AI techniques such as neural networks are potential choices. For capturing long range persistence, the long memory component can be assigned to the observation equation, latent equation and/or the realised volatility equation, depending on their ACF structures. To allow for nonlinearity, the threshold or buffered threshold effects can be adopted to either of these three equations. In fact, these

non-linear effects can also occur in the long memory parameters (such as in (u, d)) if the long memory feature itself displays regime switching. Another potential avenue is the comparision of the Smooth Transition Autoregressive (STAR) model of Terasvirta and Anderson (1992) to the BAR model. The STAR model can be thought of as an extension to the classical autoregressive model by allowing for changes in the model parameters to transition according to the value of exogenous variables. As such, the STAR model can be represented in its most simplest form as:

$$y_t = \pi + \pi_1' w_t + (\pi_{20} + \pi_2' w_t) F(y_{t-d}) + \varepsilon_t$$

where $\varepsilon_t \sim N(0, \sigma^2), \pi_j = (\pi_{j1}, ..., \pi_{jp})', j = 1, 2, w_t = (y_{t-1}, ..., y_{t-p})'$ and d is some lag $d = 1, 2, \ldots$. F is a transition function which is bounded by zero and one, and serves the purpose of “smoothly transitioning" between one regime to the next. As such, this differs from the BAR or TAR model since there is no “jump" type behaviour between transitions.

The leverage effect can also be considered between the latent and realised volatility equations; and leverage itself can be measured in alternative ways via ρ or asymmetric terms. The exploration of covariates, such as the hash rate or completion times for Cryptocurrencies in particular, in either of the three equations could also further supplement the analysis. Lastly, it would also be interesting to investigate the significance or insignificance of these equations in order to gain more insightful knowledge on the data.

9.2.2 Distributional assumptions

The main two distributional assumptions which were used in this thesis are the Gaussian and Student's t-distribution. The Student's t-distribution was found to be a superior choice when compared to the Gaussian in most cases. Alternative heavy tailed distributions can also be explored and one such notable example is the Variance Gamma (VG) distribution, which is known to be particularly useful for high frequency data. The VG model structure is nearly identical to the Student's t-distribution, except the mixing distribution is Gamma instead of Inverse-Gamma. Additionally, the

Skew t-distribution can easily be adopted with minor modification. Another interesting avenue to consider is the dependency structure between the three main equations jointly. More specifically, a trivariate distribution considering all three equations jointly. The most notable application would be the modification of the degrees of freedom parameter, ν, which is common in the observation and latent volatility equation. Choy, Chen, and Lin (2014) considered a similar model to our GMA-SV-LVG-HC model (without GMA effects), but with different degrees of freedom for the observation and latent volatility equations as ν_1 and ν_2 respectively. A similar proposal can also be trialed under our models. Further, the symmetric distribution assumptions can be challenged throughout this thesis with an asymmetric proposal such as the Skew-t distribution. For a given normally distributed random variable $y \sim N(\mu, \sigma^2)$, the Skew t-distribution representation can be easily derived with the mean-scale mixture representation $y \sim N(\mu + \gamma U, U\sigma^2), \quad U \sim \text{IG}(\nu, \nu)$, where γ is the skewness parameter.

9.2.3 Multivariate extensions

Although we consider bivariate effects between y_t and h_t in this thesis, our proposed models are all univariate in nature. A potential avenue for further research is a multivariate approach by considering $I > 1$ observation series for $y_{i,t}$ and $h_{i,t}$, where $i = 1, \ldots, I$. This is especially relevant to Cryptocurrencies, since it has been empirically shown that that Cryptocurrency returns are generally correlated to other financial assets such as stock and bond prices (Bianchi, 2018). As such, the inclusion of exogenous variables to model their returns is a sensible suggestion. Further, it is noted that Cryptocurrencies are also generally correlated and move together in response to bad news. As such, a sensible multivariate model to propose is the Multivariate SV model with RV and pairwise realised correlations of Yamauchi, Omori, and Others (2016). This model can be extended to include exogenous variables (such as stock and bond prices) and would prove useful given the observed empirical properties of Cryptocurrencies.

A notable application of multivariate modelling of Cryptocurrencies is that of Catania, Grassi, and Ravazzolo (2018) who use Vector Autoregressive models, Bayesian VAR, time-varying parameters and stochastic volatility VAR models to jointly predict

4 Cryptocurrency series. They find that density prediction is more accurate in the multivariate case compared to the univariate case, but not point predicition.

9.2.4 Other minor extensions

Finally, there are potentially other minor aspects for future investigation. Firstly, the complexity of the long memory features can be extended to include the so-called k-factor Gegenbauer filter of Woodward, Cheng, and Gray (1998). In essence, this is a product of K Gegenbauer filters so that there exists $(u_1, \ldots, u_K)$ and $(d_1, \ldots, d_K)$ to describe these K factors. This filter can be applied to either one of the three equations and compared to our case of $K = 1$ using a penalty function.

Secondly, there are a host of more efficient realised volatility measures using intraday price movement information (Parkinson, 1980; Garman and Klass, 1980; Rogers and Satchell, 1991). Indeed, there exists the potential to further investigate the performance of our choice of the log daily realised range, relative to these realised measures. Another starting point for such an investigation is to derive an equivalent of the VIX for Cryptocurrencies using the Black-Scholes model, and use this as a realised measure.

Also, the exploration of different prior assumptions can be made. In particular, the truncated normal assumption of u, d, and ρ can be tested with other popular distributions such as the Beta distribution. However, long memory models require a large number of observations to correctly estimate the long memory parameters, and as such, the choice of prior does not make a difference from our experience.

The loss functions explored in Section 5.5.2 can be extended to include the tick-loss function of Giacomini and Komunjer (2005). The tick-loss at the q^{th} quantile is denoted as $\mathcal{T}_q$ and defined as

$$\mathcal{T}_q = (q - \mathbb{1}(e_{t+1} < 0))e_{t+1},$$

where e_t is the forecast error and defined as $e_t = y_t - \text{VaR}_{t|t-1}$. We note the tick-loss function, is the same as the loss function in quantile regression, except in a different form. The tick-loss is advantageous over other methods because it is said to be 'encompassing'. An encompassing loss function is one that:

1. Involves the computation of the expected loss; and
2. involves the weights of the forecast combinations.

9.3 Concluding remarks

In conclusion, it is found that superior model choices lead to superior outcomes. Popular financial time series such as U.S. CPI and the more modern, Cryptocurrencies, are better estimated using the Gegenbauer long memory model with latent stochastic volatility effects. This thesis deals with a myriad of issues, which at its core stem from improving financial time series estimation. Although the SV model is robust, it lacks a diligent framework to properly address the unique features which are commonly found in financial time series. The Gegenbauer long memory filter addresses longreaching persistent autocorrelations commonly found in return series. A typical issue often faced by the time series analyst is the problem of lingering residual autocorrelations at higher lags once the analysis is complete. A common quick-fix is to model the errors themselves with a large lagged moving average parameter.

However, we believe a more reliable solution is to consider the class of time series models proposed in this thesis. The GMA-SV model and its extensions should also be considered during the initial empirical data analysis stage. We conclude with a relevant quote:

"To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of."

Ronald Fisher (1938)

Appendix A

Tuning the Proposal Distribution of $[u, d]$

In order to achieve high efficiency when sampling $[u, d]$, we tune the precision parameter of the proposal distribution(s). An acceptance rate that is too high could mean the proposal variance is too low and always accepting values around the current value. An acceptance rate that is too low could mean the proposal variance is too high and always rejecting, and therefore, the chain is not moving.

Gelman et al. (2013) notes that care must be taken when tuning to avoid convergence to the wrong distribution. Since the updating rule is dependant on our previous simulation steps the transition probabilities are now more complicated than before. The chain may move more quickly through flat parts of the distribution and slower through non-smooth parts of the distribution. This of course would result in the incorrect sampling of the entire proposal distribution. The general advice here to rectify such a situation is to tune in one phase of the sampling, and make the relevant inferences from a second phase where no tuning is performed. We follow this advice and tune only in the burn-in period.

We calculate the acceptance rate for every 250 MCMC iterates. If this acceptance rate is below 15% or above 50%, then we update our tuning parameter c_u and c_d according to

$$c = \max\left\{c \times \frac{\Phi^{-1}\left(p_{optimal}/2\right)}{\Phi^{-1}\left(p_{current}/2\right)}, 0.01\right\}$$

where $\Phi^{-1}(\cdot)$ is the inverse Normal CDF, $p_{optimal}$ is the optimal acceptance rate and $p_{current}$ is the current acceptance rate. Roberts and Rosenthal (2001) prove an acceptance rate between 15% and 50% is at least 80% efficient. We choose an optimal

acceptance rate of 23.4% due to the seminal work of Roberts, Gelman, and Gilks (1997).

This procedure is repeated for every 250 loops, and $p_{current}$ resets after each 250 MCMC set (i.e. from 1 to 250, from 251 to 500,...). After the burn-in has completed, we record only one acceptance rate, which is what is reported in all of our inferences.

Appendix B

Estimating h

We discuss here the estimation of the latent variable $\boldsymbol{h} = [h_1, \ldots, h_T]$. Clearly, modifying this to cater for alternative means in the observation equation is trivial, so we discuss the estimation of the GARMA-SV model only to focus on the relevant derivations. In essence, we modify the precision sampler of Chan (2013) to to exploit the banded structure of $p(\boldsymbol{h}|\alpha, \beta, \sigma^2)$. First, we seek a linear expression for $\boldsymbol{h}$

$$\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{G}_J \boldsymbol{\varepsilon} \\
\boldsymbol{G}_J^{-1}\boldsymbol{Y} &= \boldsymbol{\varepsilon} \\
(\boldsymbol{G}_J^{-1}\boldsymbol{Y}) \circ (\boldsymbol{G}_J^{-1}\boldsymbol{Y}) &= \boldsymbol{\varepsilon} \circ \boldsymbol{\varepsilon} \\
\log\left[(\boldsymbol{G}_J^{-1}\boldsymbol{Y}) \circ (\boldsymbol{G}_J^{-1}\boldsymbol{Y})\right] &= \log\left[\boldsymbol{\varepsilon} \circ \boldsymbol{\varepsilon}\right].
\end{aligned}$$

where A∘B refers to the Hadamard product of A and B. Let $\boldsymbol{Y}^* = \log \boldsymbol{G}_J^{-2}\boldsymbol{Y}^2 = [y_1^*, \ldots, y_T^*]$ and $\boldsymbol{\varepsilon}^* = [\log \varepsilon_1^2, \ldots, \log \varepsilon_T^2]'$ for notational convenience. The sampling of $\boldsymbol{\varepsilon}^*$ is highly non-standard as this is now a log-χ_1^2 distribution. Kim, Shephard, and Chib (1998) suggest to approximate this using an offset Gaussain mixture representation. Essentially, the probability density function is approximated as

$$p(\varepsilon_t^*) \approx \sum_{i=1}^{K} p_i \mathrm{N}(\varepsilon_t^*; \mu_i, \sigma_i^2).$$

Each p_i is the probability of the i^{th} mixture component. The authors estimate $K, p_i, \mu_i, \sigma_i^2$ by matching the first four moments of the true theoretical distributions. This is performed using non-linear least squares optimisation techniques until the approximating densities are within a small distance to the true density.

i	p_i	μ_i	σ_i^2
1	0.00609	1.92677	0.11265
2	0.04775	1.34744	0.17788
3	0.13057	0.73504	0.26768
4	0.20674	0.02266	0.40611
5	0.22715	-0.85173	0.62699
6	0.18842	-1.97278	0.98583
7	0.12047	-3.46788	1.57469
8	0.05591	-5.55246	2.54498
9	0.01575	-8.68384	4.16591
10	0.00115	-14.65000	7.33342

TABLE B.1: $K = 10$ mixture components as found in Omori et al. (2007)

Kim, Shephard, and Chib (1998) find satisfactory results with $K = 7$ mixture components, however, Omori et al. (2007) remark that $K = 10$ is a more reliable fit when leverage effects are considered. Although we do not consider leverage effects in our work, we favour this more conservative approach and use the following parameters as shown in Table B.1. Evidently, these parameters do not need to be estimated during each MCMC sweep since they are independant of all other parameters in the sampler. It should be noted that the mixture density can be written in terms of a component indicator variable s_t such that $P(s_t = i) = p_i$. Therefore, it is computationally cheap to sample the mixture components, which are denoted as $\boldsymbol{s}$.

It is worthwhile to reinforce here that $\boldsymbol{s}$ is a $T \times 1$ vector, and we sample s_t for each time point. Each s_t is independant so that $p(\boldsymbol{s}|\boldsymbol{Y}^*, \boldsymbol{h}) = \prod_{t=1}^{T} p(s_t|y_t^*, h_t)$. Since s_t is discrete, it is easy to sample using the slice sampler.

Once s_t has been sampled, we are able to sample $\boldsymbol{\varepsilon}^*$ as

$$\boldsymbol{\varepsilon}^*|\boldsymbol{s} \ \sim \mathrm{N}(\boldsymbol{\mu}_{\varepsilon^*}, \boldsymbol{\Sigma}_{\varepsilon^*})$$

where $\boldsymbol{\mu}_{\varepsilon^*} = (\mu_{s_1}, \ldots, \mu_{s_T})$, $\Sigma_{\varepsilon^*} = \mathrm{diag}(\sigma_{s_1}^2, \ldots, \sigma_{s_T}^2)$. Hence, it is clear to see that

$$p(\boldsymbol{Y}^*|\boldsymbol{s}, \boldsymbol{h},) \ \sim \mathrm{N}(\boldsymbol{h} + \boldsymbol{\mu}_{\varepsilon^*}, \boldsymbol{\Sigma}_{\varepsilon^*})$$

So the likelihood of $\boldsymbol{Y}^*$ is

$$p(\boldsymbol{Y}^*|\boldsymbol{s}, \boldsymbol{h}) = (2\pi)^{-\frac{T}{2}} |\boldsymbol{\Sigma}_{\varepsilon^*}| \exp\left\{ -\frac{1}{2} (\boldsymbol{Y}^* - \boldsymbol{h} - \boldsymbol{\mu}_{\varepsilon^*})' \boldsymbol{\Sigma}_{\varepsilon^*}^{-1} (\boldsymbol{Y}^* - \boldsymbol{h} - \boldsymbol{\mu}_{\varepsilon^*})' \right\}$$

$$\log p(\boldsymbol{Y}^*|\boldsymbol{s}, \boldsymbol{h}) \propto -\frac{1}{2} (\boldsymbol{Y}^* - \boldsymbol{h} - \boldsymbol{\mu}_{\varepsilon^*})' \boldsymbol{\Sigma}_{\varepsilon^*}^{-1} (\boldsymbol{Y}^* - \boldsymbol{h} - \boldsymbol{\mu}_{\varepsilon^*})'.$$

Recall that

$$h_{t+1}|h_t \sim \begin{cases} \mathrm{N}\left(\alpha, \frac{\sigma^2}{1-\beta^2}\right), & t = 0 \\ \mathrm{N}(\alpha + \beta(h_t - \alpha), \sigma^2), & t \neq 0 \end{cases}$$

which can be written out in matrix notation as

$$\boldsymbol{H}_\phi \boldsymbol{h} = \tilde{\boldsymbol{\alpha}} + \boldsymbol{\omega}$$

where

$$\begin{aligned} \boldsymbol{\omega} &\sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_h) \\ \boldsymbol{\Sigma}_h &= \text{diag}\left(\frac{\sigma^2}{1-\beta^2}, \sigma^2, \ldots, \sigma^2, \sigma^2\right) \\ \tilde{\boldsymbol{\alpha}} &= (\alpha, \alpha(1-\beta), \ldots, \alpha(1-\beta))' \\ \boldsymbol{H}_\phi &= \begin{pmatrix} 1 & 0 & \ldots & \ldots & \ldots \\ -\beta & 1 & \ldots & \ldots & \ldots \\ 0 & -\beta & 1 & \ldots & \ldots \\ \vdots & 0 & -\beta & \ldots & \ldots \\ 0 & \vdots & \vdots & \ldots & \ldots \end{pmatrix}. \end{aligned}$$

Now,

$$\begin{aligned} \boldsymbol{H}_\phi \boldsymbol{h} &= \tilde{\boldsymbol{\alpha}} + \boldsymbol{\omega} \\ \boldsymbol{h} &= \boldsymbol{H}_{\boldsymbol{\phi}}^{-1} \tilde{\boldsymbol{\alpha}} + \boldsymbol{H}_{\boldsymbol{\phi}}^{-1} \boldsymbol{\omega}. \end{aligned}$$

So that

$$\boldsymbol{h}|\alpha, \beta\, \sigma^2 \sim \mathbf{N}(\boldsymbol{H}_{\boldsymbol{\phi}}^{-1} \tilde{\boldsymbol{\alpha}}, (\boldsymbol{H}_{\boldsymbol{\phi}}' \boldsymbol{\Sigma}_{\boldsymbol{h}}^{-\mathbf{1}} \boldsymbol{H}_{\boldsymbol{\phi}})^{-1})$$

hence

$$\boldsymbol{h}|\alpha, \beta, \sigma^2 = (2\pi)^{-\frac{T}{2}} |(\boldsymbol{H}_\phi' \boldsymbol{\Sigma}_h^{-1} \boldsymbol{H}_\phi)^{-1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{h} - \boldsymbol{\mu}_h)'(\boldsymbol{H}_\phi' \boldsymbol{\Sigma}_h^{-1} \boldsymbol{H}_\phi)(\boldsymbol{h} - \boldsymbol{\mu}_h)\right\}$$

where $\boldsymbol{\mu}_h = \boldsymbol{H}_\phi^{-1}\tilde{\boldsymbol{\alpha}}$. However, it is clear to see that $|(\boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi)^{-1}| = \sigma^{2T}/(1-\beta^2)$. So the log-density of $\boldsymbol{h}$ can be expressed as

$$\log p(\boldsymbol{h}|\alpha,\beta,\sigma^2) \propto -\frac{1}{2}\log\frac{\sigma^{2T}}{(1-\beta^2)} - \frac{1}{2}(\boldsymbol{h}-\boldsymbol{\mu}_h)'(\boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi)(\boldsymbol{h}-\boldsymbol{\mu}_h).$$

Therefore, the full conditional distribution of $\log(\boldsymbol{h})$ is

$$\begin{aligned}
&\log p(\boldsymbol{h}|\boldsymbol{Y}^*,\alpha,\beta\,\sigma^2) \qquad\qquad (B.1)\\
&\propto \log p(\boldsymbol{Y}^*|\boldsymbol{s},\boldsymbol{h},\alpha,\beta\,\sigma^2) + \log p(\boldsymbol{h}|\alpha,\beta\,\sigma^2)\\
&= -\frac{1}{2}(\boldsymbol{Y}^*-\boldsymbol{h}-\boldsymbol{\mu}_{\varepsilon^*})'\boldsymbol{\Sigma}_{\varepsilon^*}^{-1}(\boldsymbol{Y}^*-\boldsymbol{h}-\boldsymbol{\mu}_{\varepsilon^*})' - \frac{1}{2}\log\frac{\sigma^{2T}}{(1-\beta^2)}\\
&- \frac{1}{2}(\boldsymbol{h}-\boldsymbol{\mu}_h)'(\boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi)(\boldsymbol{h}-\boldsymbol{\mu}_h)\\
&\propto (\boldsymbol{Y}^*-\boldsymbol{h}-\boldsymbol{\mu}_{\varepsilon^*})'\boldsymbol{\Sigma}_{\varepsilon^*}^{-1}(\boldsymbol{Y}^*-\boldsymbol{h}-\boldsymbol{\mu}_{\varepsilon^*})' + (\boldsymbol{h}-\boldsymbol{\mu}_h)'(\boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi)(\boldsymbol{h}-\boldsymbol{\mu}_h)\\
&= (\boldsymbol{Y}^*-\boldsymbol{\mu}_{\varepsilon^*})'\boldsymbol{\Sigma}_{\varepsilon^*}^{-1}(\boldsymbol{Y}^*-\boldsymbol{\mu}_{\varepsilon^*}) - (\boldsymbol{Y}^*-\boldsymbol{\mu}_{\varepsilon^*})'\boldsymbol{\Sigma}_{\varepsilon^*}^{-1}\boldsymbol{h} - \boldsymbol{h}'\boldsymbol{\Sigma}_{\varepsilon^*}^{-1}(\boldsymbol{Y}^*-\boldsymbol{\mu}_{\varepsilon^*}) + \boldsymbol{h}'\boldsymbol{\Sigma}_{\varepsilon^*}^{-1}\boldsymbol{h}\\
&+ \boldsymbol{h}'\boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi\boldsymbol{h} - \boldsymbol{h}'\boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi\boldsymbol{\mu}_h - \boldsymbol{\mu}_h'\boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi\boldsymbol{h} + \boldsymbol{\mu}_h'\boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi\boldsymbol{\mu}_h\\
&\propto -(\boldsymbol{Y}^*-\boldsymbol{\mu}_{\varepsilon^*})'\boldsymbol{\Sigma}_{\varepsilon^*}^{-1}\boldsymbol{h} - \boldsymbol{h}'\boldsymbol{\Sigma}_{\varepsilon^*}^{-1}(\boldsymbol{Y}^*-\boldsymbol{\mu}_{\varepsilon^*}) + \boldsymbol{h}'\boldsymbol{\Sigma}_{\varepsilon^*}^{-1}\boldsymbol{h} + \boldsymbol{h}'H_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi\boldsymbol{h}\\
&- \boldsymbol{h}'\boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\tilde{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi\boldsymbol{h}\\
&= \boldsymbol{h}'\left(\boldsymbol{\Sigma}_{\varepsilon^*}^{-1} + \boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi\right)\boldsymbol{h} - 2\boldsymbol{h}'[\boldsymbol{\Sigma}_{\varepsilon^*}^{-1}(\boldsymbol{Y}^*-\boldsymbol{\mu}_{\varepsilon^*}) + \boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\tilde{\boldsymbol{\alpha}}]. \qquad (B.2)
\end{aligned}$$

Now, consider some multivariate Gaussian distribution $\boldsymbol{\theta} \sim \mathbf{N}(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$ with log PDF

$$\log p(\theta) \propto \boldsymbol{\theta}'\boldsymbol{\Sigma}_\theta^{-1}\boldsymbol{\theta} - 2\boldsymbol{\theta}'\boldsymbol{\Sigma}_\theta^{-1}\boldsymbol{\mu}_\theta. \qquad (B.3)$$

If we compare B.2 with B.3, then it is clear to see that

$$\begin{aligned}
\boldsymbol{\mu}_\theta &= \boldsymbol{\Sigma}_\theta(\boldsymbol{\Sigma}_{\varepsilon^*}^{-1}(\boldsymbol{Y}^*-\boldsymbol{\mu}_{\varepsilon^*}) + \boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\tilde{\boldsymbol{\alpha}})\\
\boldsymbol{\Sigma}_\theta &= \left(\boldsymbol{\Sigma}_{\varepsilon^*}^{-1} + \boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi\right)^{-1}.
\end{aligned}$$

Finally, the posterior distribution of $\boldsymbol{h}$ can be sampled as a block from

$$p(\boldsymbol{h}|\boldsymbol{Y}^*,\alpha,\beta,\sigma^2) \sim N\left(\boldsymbol{\Sigma}_\theta[\boldsymbol{\Sigma}_{\varepsilon^*}^{-1}(\boldsymbol{Y}^*-\boldsymbol{\mu}_{\varepsilon^*}) + \boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\tilde{\boldsymbol{\alpha}}], (\boldsymbol{\Sigma}_{\varepsilon^*}^{-1} + \boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi)^{-1}\right). \qquad (B.4)$$

Appendix C

Estimating the Marginal Likelihood

Suppose we want to compare a set of models $\{M_1, \ldots, M_K\}$ in a Bayesian setting. The frequentist is able to use the classical log likelihood ratio test, which if of course distributed as a χ^2 with degrees of freedom equal to the difference in parameters between the two models. In Bayesian analysis, we use the Bayes Factor, which is given by

$$BF_{ij} = \frac{p(\boldsymbol{y}|M_i)}{p(\boldsymbol{y}|M_j)}$$

where

$$p(\boldsymbol{y}|M_k) = \int p(\boldsymbol{y}|\boldsymbol{\theta}, M_k)p(\boldsymbol{\theta}|M_k)d\boldsymbol{\theta}. \tag{C.1}$$

The quantity C.1 is called the *marginal likelihood* (ML). It can be shown to be asymptotically equivalent to the SIC. The estimation of the ML is typically nontrivial.

$$\begin{aligned} p(\boldsymbol{y}|M_k) &= \int p(\boldsymbol{y}|\boldsymbol{\theta}, M_k)p(\boldsymbol{\theta}|M_k)d\boldsymbol{\theta} \\ &= \frac{\int p(\boldsymbol{y}|\boldsymbol{\theta}, M_k)p(\boldsymbol{\theta}|M_k)}{g(\boldsymbol{\theta})}g(\boldsymbol{\theta})d\boldsymbol{\theta}. \end{aligned}$$

The ML is estimated as

$$\widehat{p(\boldsymbol{y})} = \frac{1}{R}\sum_{i=1}^{R}\frac{p(\boldsymbol{y}|\boldsymbol{\theta}^{(i)})p(\boldsymbol{\theta}^{(i)})}{g(\boldsymbol{\theta}^{(i)})} \tag{C.2}$$

where each $\boldsymbol{\theta}^{(i)}$ is a draw from the importance density, $g(\boldsymbol{\theta}^{(i)})$, and R is the total number of draws. The choice of g is important. Ideally, we would use the posterior as it carries all the information that we need, but the normalizing constant is not known. We instead use something as close as possible - $p(\hat{\boldsymbol{\theta}})$. It can be shown that this density minimizes the Kullback-Leibler distance to the posterior.

In the case of the GARFIMA-SV model, it is easy to see that $p(\boldsymbol{\theta}^{(i)})$ is

$$\begin{aligned}
p(u)p(d)p(\alpha)p(\beta)p(\sigma^2) &= \mathcal{N}_T(\mu_u, \sigma_u^2)\mathcal{N}_T(\mu_d, \sigma_d^2)\mathcal{N}(\mu_\alpha, \sigma_\alpha^2)\mathcal{N}_T(\mu_\beta, \sigma_\beta^2)\mathcal{IG}(\frac{a}{2}, \frac{b}{2}) \\
&= \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left\{-\frac{1}{2}\frac{(u-\mu_u)^2}{\sigma_u^2}\right\} \times \Delta_u \times \frac{1}{\sqrt{2\pi\sigma_d^2}} \exp\left\{-\frac{1}{2}\frac{(d-\mu_d)^2}{\sigma_d^2}\right\} \\
&\times \Delta_d \times \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left\{-\frac{1}{2}\frac{(\alpha-\mu_\alpha)^2}{\sigma_\alpha^2}\right\} \times \frac{1}{\sqrt{2\pi\sigma_\beta^2}} \exp\left\{-\frac{1}{2}\frac{(\beta-\mu_\beta)^2}{\sigma_\beta^2}\right\} \\
&\times \Delta_c \frac{\left(\frac{b}{2}\right)^{\frac{a}{2}}}{\Gamma(\frac{a}{2})} \left(\sigma^2\right)^{-\frac{a}{2}-1} \exp\left\{-\frac{\frac{b}{2}}{\sigma^2}\right\}
\end{aligned}$$

where

$$\begin{aligned}
\Delta_u &= (\Phi(1; \mu_u, \sigma_u^2) - \Phi(-1; \mu_u, \sigma_u^2))^{-1} \\
\Delta_d &= [(\Phi(0.5; \mu_d, \sigma_d^2) - \Phi(0; \mu_d, \sigma_d^2))^{-1}]\mathbb{1} + [(\Phi(0.25; \mu_d, \sigma_d^2) - \Phi(0; \mu_d, \sigma_d^2))^{-1}](1-\mathbb{1}) \\
\Delta_\beta &= (\Phi(1; \mu_\beta, \sigma_\beta^2) - \Phi(-1; \mu_\beta, \sigma_\beta^2))^{-1}
\end{aligned}$$

where $\Phi(x; m, s^2)$ is the normal CDF with mean m and variance s^2 evaluated at x, and $\mathbb{1}$ is an indicator variable that is equal to 1 when $|u| \leq 1$, and 0 otherwise.
We work with logarithms for ease of computation. Thus,

$$\begin{aligned}
&\log p(u)p(d)p(\alpha)p(\beta)p(\sigma^2) = \\
&-\frac{1}{2}\log(2\pi\sigma_u^2) - \frac{1}{2}\frac{(u-\mu_u)^2}{\sigma_u^2} + \log\Delta_u - \frac{1}{2}\log(2\pi\sigma_d^2) - \frac{1}{2}\frac{(d-\mu_d)^2}{\sigma_d^2} + \log\Delta_d \\
&-\frac{1}{2}\log(2\pi\sigma_\alpha^2) - \frac{1}{2}\frac{(\alpha-\mu_\alpha)^2}{\sigma_\alpha^2} - \frac{1}{2}\log(2\pi\sigma_\beta^2) - \frac{1}{2}\frac{(\beta-\mu_\beta)}{\sigma_\beta^2} + \log\Delta_\beta + \frac{a}{2}\log\frac{b}{2} \\
&-\log\Gamma\left(\frac{a}{2}\right) - \left(\frac{a}{2}+1\right)\log\sigma^2 - \frac{\frac{b}{2}}{\sigma^2}.
\end{aligned}$$

Now, the observed-likelihood $p(\boldsymbol{y}|\boldsymbol{\theta})$ is calculated as

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \int p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{h}) p(\boldsymbol{h}|\boldsymbol{\theta}) d\boldsymbol{h}.$$

This again, is a nontrivial quantity to estimate. We once again use an importance sampling scheme to estimate the term $p(\boldsymbol{y}|\boldsymbol{\theta}, M_k)$ in (C.2)

$$\widehat{p(\boldsymbol{y}|\boldsymbol{\theta})} = \frac{1}{R} \sum_{i=1}^{R} \frac{p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{h}^{(i)}) p(\boldsymbol{h}^{(i)}|\boldsymbol{\theta})}{p(\boldsymbol{h}^{(i)})}.$$

It can be shown that a good approximating density for $p(\boldsymbol{h}^{(i)})$ is $p(\boldsymbol{h}|\boldsymbol{y}, \boldsymbol{\theta})$. Hence,

$$\begin{aligned}
\log p(\boldsymbol{h}|\boldsymbol{\theta}) &= -\frac{T}{2}\log(2\pi) - \frac{1}{2}(T \log \sigma^2 - \log(1-\beta^2)) - \frac{1}{2}(\boldsymbol{h} - \boldsymbol{\mu}_h)' \boldsymbol{H}_\phi \boldsymbol{\Sigma}_h^{-1} \boldsymbol{H}_\phi' (\boldsymbol{h} - \boldsymbol{\mu}_h) \\
\log p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{h}) &= -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} h_t - \frac{1}{2}\boldsymbol{Y}' \boldsymbol{\Gamma}^{-1} \boldsymbol{Y} \\
-\log p(\boldsymbol{h}) &= \frac{T}{2}\log(2\pi) + \frac{1}{2}\log(|\boldsymbol{\Sigma}_\theta|) + \frac{1}{2}(\boldsymbol{h} - \tilde{\boldsymbol{h}})' \Sigma_\theta^{-1} (\boldsymbol{h} - \tilde{\boldsymbol{h}})
\end{aligned}$$

where $\tilde{\boldsymbol{h}}$ is the mode of $\boldsymbol{h}$, and $\boldsymbol{\Gamma} = \boldsymbol{G}_J \boldsymbol{V} \boldsymbol{G}_J'$. The mode can be found using a search method, such as a Newton-Raphson scheme.

Appendix D

Simulation Study Diagnostics

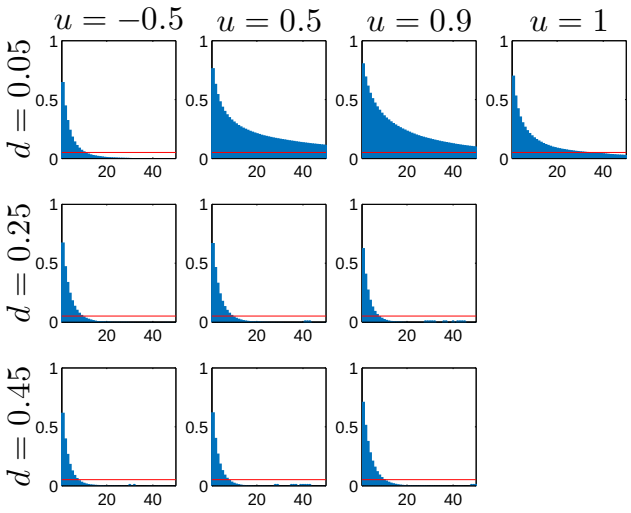

FIGURE D.1: Each graph depicts the mean Sample Autocorrelation of $1,000$ MCMC runs of $\hat{u}$ for various values of $[u, d]$. The most notable observation is that when d is low, and as $u \to 0.\dot{9}$ the convergence of u to its true value gets slower. This is an expected result, since as $d \to 0$, the process has less information, and becomes "less long-memory". Furthermore, this slow decay is not a result of boundary issues, since we do not see the same slow decay with $d = 0.45$, which too is 0.05 units away from the boundary.

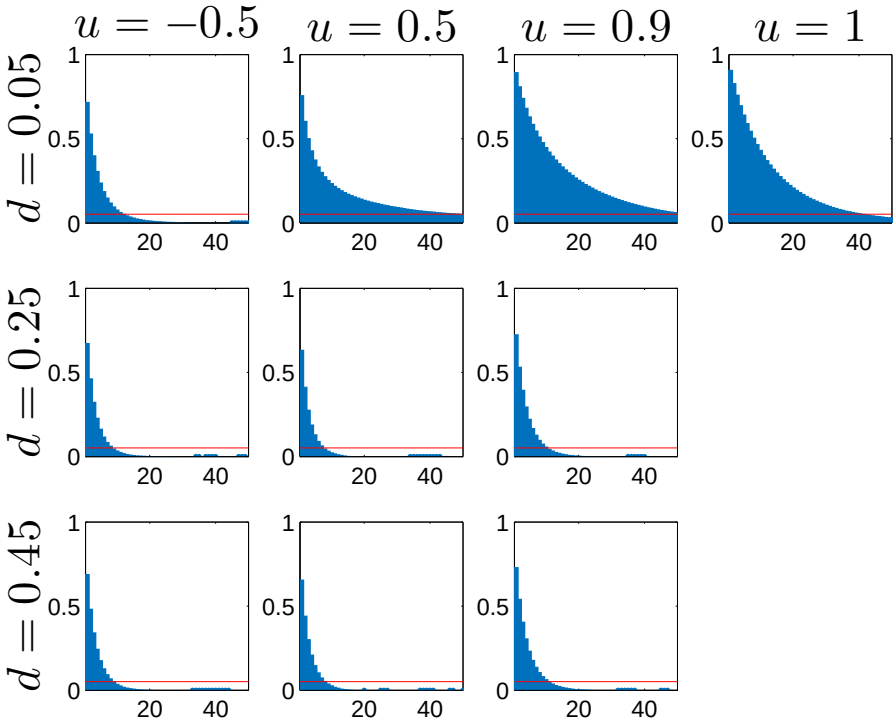

FIGURE D.2: Each graph depicts the mean Sample Autocorrelation of $1,000$ MCMC runs of $\hat{d}$ for various values of $[u, d]$. Similarly to u, d exhibits slow decay for low values of d.

u	d	GR_u	GR_d	GR_μ	GR_ϕ	GR_α	GR_β	GR_{σ^2}
-0.50	0.05	1.038	1.092	1.025	1.099	1.052	1.078	1.070
-0.50	0.25	1.052	1.102	1.015	1.092	1.046	1.097	1.077
-0.50	0.45	1.015	1.106	1.008	1.106	1.051	1.090	1.072
0.50	0.05	1.138	1.058	1.027	1.127	1.049	1.084	1.072
0.50	0.25	1.049	1.101	1.026	1.098	1.048	1.089	1.081
0.50	0.45	1.016	1.100	1.023	1.101	1.047	1.092	1.066
0.90	0.05	1.199	1.053	1.031	1.085	1.047	1.096	1.072
0.90	0.25	1.020	1.103	1.057	1.106	1.045	1.092	1.076
0.90	0.45	1.004	1.105	1.107	1.096	1.048	1.084	1.075
1.00	0.05	1.144	1.071	1.057	1.085	1.053	1.094	1.077

TABLE D.1: Gelman-Rubin statistics for the in-sample simulation study using T = 1,500. All parameters have a statistic close to 1, which is suggestive of convergence.

Appendix E

Model definitions

Let $y_t, t = 1, 2, ..., T$ be a stochastic process satisfying the equations

1. Stochastic volatility (SV):

$$\begin{aligned} y_t &= \varepsilon_t, \\ h_t &= \alpha + \beta(h_{t-1} - \alpha) + \eta_t, \end{aligned}$$

   where $\varepsilon_t \sim N(0, e^{h_t})$, $\eta_t \sim \mathrm{N}(0, \sigma^2)$ and $E[\varepsilon_t \eta_t] = 0$.

2. Stochastic volatility model with leverage (SV-LVG):

$$\begin{aligned} &y_t = \varepsilon_t, \\ &h_{t+1} = \alpha + \beta(h_t - \alpha) + \eta_{t+1}, \\ &\begin{pmatrix} \varepsilon_t \\ \eta_{t+1} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} e^{h_t} & \sigma\rho e^{h_t/2} \\ \sigma\rho e^{h_t/2} & \sigma^2 \end{pmatrix} \right). \end{aligned}$$

3. Gegenbauer long memory model with stochastic volatility (GMA-SV)

$$\begin{aligned} (1 - 2uB + B^2)^d y_t &= \varepsilon_t, \\ h_t &= \alpha + \beta(h_{t-1} - \alpha) + \eta_t, \end{aligned}$$

   where $\varepsilon_t \sim N(0, e^{h_t})$, $\eta_t \sim N(0, \sigma^2)$ and $E[\varepsilon_t \eta_t] = 0$.

4. Gegenbauer long memory model with stochastic volatility and leverage (GMA-SV-LVG):

$$
\begin{aligned}
&(1-2uB+B^2)^d y_t = \varepsilon_t,\\
&h_{t+1} = \alpha + \beta(h_t - \alpha) + \eta_{t+1},\\
&\begin{pmatrix} \varepsilon_t^* \\ \eta_{t+1}^* \end{pmatrix} \sim \boldsymbol{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} e^{h_t} & \sigma\rho e^{h_t/2} \\ \sigma\rho e^{h_t/2} & \sigma^2 \end{pmatrix}\right).
\end{aligned}
$$

5. Gegenbauer long memory model with stochastic volatility and heavy common tails (GMA-SV-HC)

$$
\begin{aligned}
&(1-2uB+B^2)^d y_t = \varepsilon_t,\\
&h_{t+1} = \alpha + \beta(h_t - \alpha) + \eta_{t+1},
\end{aligned}
$$

where $\varepsilon_t \sim t_\nu(0, e^{h_t})$, $\eta_t \sim t_\nu(0, \sigma^2)$ and $E[\varepsilon_t \eta_t] = 0$.

6. Gegenbauer long memory model with stochastic volatility, leverage and heavy common tails (GMA-SV-LVG-HC)

$$
\begin{aligned}
&(1-2uB+B^2)^d y_t = \varepsilon_t^*,\\
&h_{t+1} = \alpha + \beta(h_t - \alpha) + \eta_{t+1}^*,\\
&\begin{pmatrix} \varepsilon_t^* \\ \eta_{t+1}^* \end{pmatrix} \sim t_\nu\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} e^{h_t} & \sigma\rho e^{h_t/2} \\ \sigma\rho e^{h_t/2} & \sigma^2 \end{pmatrix}\right).
\end{aligned}
$$

Appendix F

Bayesian analysis of the GARMA-SV model with bivariate Student's-t errors and leverage

Sampling for u and d

In order to estimate u and d, we consider two independent truncated normal priors with support in the region where generalised long-memory holds such that $p(u) \sim \text{N}(\mu_u, \sigma_u^2)\mathbb{1}_{ud}$ and $p(d) \sim \text{N}(\mu_d, \sigma_d^2)\mathbb{1}_{ud}$ where $\mathbb{1}_{ud} = \mathbb{1}(\{-1 < u < 1, 0 < d < 0.5\} \cup \{|u| = 1, 0 < d < 0.25\})$ and $\mathbb{1}$ is an indicator function. Note that we impose Gegenbauer long-memory stationarity through the prior distributions of u and d.

The posterior of both u and d are complicated and do not have a tractable conjugate form. Subsequently samples from these distributions cannot be obtained directly. In order to sample u and d, we use an approximation based on posterior modes from Gelman et al. (2013), coupled with a proposal distribution precision tuning algorithm which we conduct only within the burn-in period. We briefly note that attempts to estimate $[u, d]$ using the Metropolis algorithm proved futile due to extremely slow convergence, and "boundary trap" issues. Consider the following independence chain Metropolis-Hastings algorithm:

1. Maximimize the log posterior of u and d to find the modes $\tilde{u}$ and $\tilde{d}$ respectively. The log posterior modes are found by maximising

$$\log p_u(u|d,\boldsymbol{h}) = \log f(\boldsymbol{Y}|d,\boldsymbol{h}) + \log \mathrm{N}(\mu_u,\sigma_u^2)\mathbb{1}_{ud},$$
$$\log p_d(d|u,\boldsymbol{h}) = \log f(\boldsymbol{Y}|u,\boldsymbol{h}) + \log \mathrm{N}(\mu_{d_1},\sigma_d^2)\mathbb{1}_{ud} + \log \mathrm{N}(\mu_{d_2},\sigma_d^2)(1-\mathbb{1}_{ud}).$$

   where the prior choices for d are designed to consider the event when $u=1$, such that $\mu_{d_1}=0.125$, $\mu_{d_d}=0.25$, $\sigma_d^2=0.05$. Also, $\mu_u=0$ and $\sigma_u^2=0.1$.

2. Sample u^* from the proposal distribution $\mathrm{N}(\tilde{u},c_u^2)$ denoted by q_u, where c_u is the scaling parameter.

3. Reject u^* unless $(\{-1<u^*<1, 0<d<0.5\}\cup\{|u^*|=1, 0<d<0.25\})$.[1] Otherwise, accept u^* with probability ζ, where

$$\zeta = \min\left\{1, \frac{p_u(u^*|d,\boldsymbol{h})q_u(u^{(m)})}{p_u(u^{(m)}|d,\boldsymbol{h})q_u(u^*)}\right\}.$$

4. Repeat steps 2-3 by replacing d with d^*, u^* with u and c_u with c_d.

If we accept u^* and d^*, then we update $u^{(m+1)}=u^*$ and $d^{(m+1)}=d^*$ respectively, and generate the updated $\boldsymbol{G}_J$ using the new Gegenbauer parameters.

Sampling for $\boldsymbol{h}$

We discuss here the estimation procedure of the latent variable $\boldsymbol{h}=[h_1,\ldots,h_T]$ which has its roots in Chan and Grant (2016b) and Chan and Strachan (2012). Let $\boldsymbol{Y}^*=\boldsymbol{G}_J^{-1}\boldsymbol{Y}$ where $\boldsymbol{Y}^*=[y_1^*,\ldots,y_T^*]$. First, consider the density of $h_{t+1}|h_t$

$$h_{t+1}|h_t \sim \begin{cases}\mathrm{N}\left(\alpha,\frac{\sigma^2\xi_1}{1-\beta^2}\right), & t=0\\ \mathrm{N}(\alpha+\beta(h_t-\alpha),\sigma^2\xi_{t+1}), & t\neq 0\end{cases}$$

which can be written in matrix notation as $\boldsymbol{H}_\phi\boldsymbol{h}=\tilde{\boldsymbol{\alpha}}+\boldsymbol{\omega}$ where

$$\boldsymbol{\omega}\sim\mathrm{N}(\boldsymbol{0},\boldsymbol{\Sigma}_h),\quad \boldsymbol{\Sigma}_h=\mathrm{diag}\left(\frac{\sigma^2\xi_1}{1-\beta^2},\sigma^2\xi_2,\ldots,\sigma^2\xi_{T+1}\right),\quad \tilde{\boldsymbol{\alpha}}=(\alpha,\alpha(1-\beta),\ldots,\alpha(1-\beta))',$$

[1] Practically, when $u^*\geq 0.99$, then we set $u^*=1$ in order to give the event $|u|=1$ non-zero probabilities

$$\boldsymbol{H}_{\phi} = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots \\ -\beta & 1 & \dots & \dots & \dots \\ 0 & -\beta & 1 & \dots & \dots \\ \vdots & 0 & -\beta & \dots & \dots \\ 0 & \vdots & \vdots & \dots & \dots \end{pmatrix}.$$

Therefore, the density of $\boldsymbol{h}$ is $\boldsymbol{h}|\xi_1, \boldsymbol{\xi}, \alpha, \beta, \sigma^2 \sim \mathbf{N}(\boldsymbol{H}_{\boldsymbol{\phi}}^{-1}\tilde{\boldsymbol{\alpha}}, (\boldsymbol{H}_{\boldsymbol{\phi}}'\boldsymbol{\Sigma}_{\boldsymbol{h}}^{-\mathbf{1}}\boldsymbol{H}_{\boldsymbol{\phi}})^{-1})$. However, it is clear to see that $|(\boldsymbol{H}_{\phi}'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_{\phi})^{-1}| = \sigma^{2T}/(1-\beta^2)$, hence the log-density is

$$\log f(\boldsymbol{h}|\xi_1, \boldsymbol{\xi}, \alpha, \beta, \sigma^2) \approx -\frac{1}{2}(\boldsymbol{h}-\boldsymbol{\mu}_h)'(\boldsymbol{H}_{\phi}'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_{\phi})(\boldsymbol{h}-\boldsymbol{\mu}_h)$$

where $\boldsymbol{\mu}_h = \boldsymbol{H}_{\phi}^{-1}\tilde{\boldsymbol{\alpha}}$.

In order to estimate $\log f(\boldsymbol{Y}^*|\boldsymbol{h}, \boldsymbol{\xi}, \alpha, \beta, \sigma^2, \rho)$, consider its Taylor series expansion around the neighborhood $\tilde{\boldsymbol{h}}$ of $\boldsymbol{h} \in \Re^T$, such that

$$\begin{aligned} \log f(\boldsymbol{Y}^*|\boldsymbol{h}, \boldsymbol{\xi}, \alpha, \beta, \sigma^2, \rho) &\approx \log f(\boldsymbol{Y}^*|\tilde{\boldsymbol{h}}, \boldsymbol{\xi}, \alpha, \beta, \sigma^2, \rho) + (\boldsymbol{h}-\tilde{\boldsymbol{h}})'\boldsymbol{f} - \frac{1}{2}(\boldsymbol{h}-\tilde{\boldsymbol{h}})'\boldsymbol{G}(\boldsymbol{h}-\tilde{\boldsymbol{h}}) \\ &\approx \boldsymbol{h}'\boldsymbol{f} - \frac{1}{2}\left[\boldsymbol{h}'\boldsymbol{G}\boldsymbol{h} - \boldsymbol{h}'\boldsymbol{G}\tilde{\boldsymbol{h}} - \tilde{\boldsymbol{h}}\boldsymbol{G}\boldsymbol{h}\right] \\ &\approx -\frac{1}{2}\left[\boldsymbol{h}'\boldsymbol{G}\boldsymbol{h} - 2\boldsymbol{h}'(\boldsymbol{f}+\boldsymbol{G}\tilde{\boldsymbol{h}})\right] \end{aligned}$$

where $\boldsymbol{f}$ and $\boldsymbol{G}$ are the gradient and Hessian respectively denoted as:

$$\boldsymbol{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_{T+1} \end{pmatrix}, \quad \boldsymbol{G} = \begin{pmatrix} G_{11} & G_{12} & 0 & \dots & 0 \\ G_{12} & G_{22} & G_{23} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & G_{T-1,T} & G_{TT} & G_{T,T+1} \\ 0 & \dots & 0 & G_{T,T+1} & G_{T+1,T+1} \end{pmatrix}$$

such that for $t = 2, \ldots, T+1$,

$$
\begin{aligned}
&f(y_t^*|h_{t+1}, h_t, \xi_{t+1}, \alpha, \beta, \sigma^2, \rho) = \\
f(y_t^*) &= \frac{1}{\sqrt{2\pi\xi_{t+1}(1-\rho^2)}e^{h_t/2}} \exp\left\{-\frac{\left\{y_t^* - \frac{\rho}{\sigma}e^{h_t/2}[h_{t+1} - \alpha - \beta(h_t - \alpha)]\right\}^2}{2(1-\rho^2)\xi_{t+1}e^{h_t}}\right\}, \\
f_1 &= \frac{\partial \log f(y_1^*)}{\partial h_1}, \quad f_t = \frac{\partial}{\partial h_t}(\log f(y_t^*) + \log f(y_{t-1}^*)), \\
G_{11} &= -\frac{\partial^2 \log f(y_1^*)}{\partial h_1^2}, \quad G_{tt} = -\frac{\partial^2}{\partial h_t^2}(\log f(y_t^*) + \log f(y_{t-1}^*)), \quad G_{t-1,t} = -\frac{\partial^2 \log f(y_t^*)}{\partial h_t \partial h_{t+1}}
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial \log f(y_t^*)}{\partial h_t} &= -\frac{1}{2} - \frac{1}{2(1-\rho^2)\xi_{t+1}}\left(-e^{-h_t}y_t^{2*} - \frac{2\beta\rho^2}{\sigma^2}[h_{t+1} - \beta h_t - \alpha(1-\beta)]\right. \\
&\quad \left. + y_t^*\frac{\rho}{\sigma}e^{-h_t/2}[h_{t+1} - \beta h_t - \alpha(1-\beta) + 2\beta]\right) \\
\frac{\partial^2 \log f(y_t^*)}{\partial h_t^2} &= -\frac{1}{2(1-\rho^2)\xi_{t+1}}\left(e^{-h_t}y_t^{*2} + \frac{2\beta^2\rho^2}{\sigma^2} - y_t^*\frac{\rho}{2\sigma}e^{-h_t/2}[h_{t+1} - \beta h_t - \alpha(1-\beta) + 4\beta]\right) \\
\frac{\partial \log f(y_t^*)}{\partial h_{t+1}} &= \frac{\rho}{\sigma(1-\rho^2)\xi_{t+1}}\left(y_t^* e^{-h_t/2} - \frac{\rho}{\sigma}[h_{t+1} - \beta h_t - \alpha(1-\beta)]\right) \\
\frac{\partial^2 \log f(y_t^*)}{\partial h_{t+1}^2} &= -\frac{\rho^2}{\sigma^2(1-\rho^2)\xi_{t+1}} \\
\frac{\partial^2 \log f(y_t^*)}{\partial h_t \partial h_{t+1}} &= \frac{\rho}{\sigma(1-\rho^2)\xi_{t+1}}\left(\frac{\beta\rho}{\sigma} - \frac{y_t^*}{2}e^{-h_t/2}\right)
\end{aligned}
$$

So the log-likelihood for the latent volatilities is

$$
\begin{aligned}
\log p(\boldsymbol{h}|\boldsymbol{Y}^*, \boldsymbol{\xi}, \alpha, \beta, \sigma^2) &\approx \log f(\boldsymbol{Y}^*|\boldsymbol{h}, \boldsymbol{\xi}, \alpha, \beta, \sigma^2, \rho) + \log f(\boldsymbol{h}|\boldsymbol{\xi}, \alpha, \beta\, \sigma^2) \\
&\approx -\frac{1}{2}\left[\boldsymbol{h}'\boldsymbol{G}\boldsymbol{h} - 2\boldsymbol{h}'(\boldsymbol{f} + \boldsymbol{G}\tilde{\boldsymbol{h}})\right] - \frac{1}{2}(\boldsymbol{h} - \boldsymbol{\mu}_h)'(\boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi)(\boldsymbol{h} - \boldsymbol{\mu}_h) \\
&\approx -\frac{1}{2}(\boldsymbol{h}'\boldsymbol{K}_h\boldsymbol{h} - 2\boldsymbol{h}'\boldsymbol{k}_h).
\end{aligned}
$$

Thus $\boldsymbol{h} \sim \mathrm{N}(\tilde{\boldsymbol{h}}, \boldsymbol{K}_h^{-1})$ where $\tilde{\boldsymbol{h}}$ is the mode of $\boldsymbol{h}$, $\boldsymbol{K}_h = \boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}_\phi + \boldsymbol{G}$ and $\boldsymbol{k}_h = \boldsymbol{H}_\phi'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{H}\boldsymbol{\mu}_h + \boldsymbol{f} + \boldsymbol{G}\tilde{\boldsymbol{h}}$. The sampling of $\boldsymbol{h}$ can be summarized with the following three steps:

1. Search for $\tilde{\boldsymbol{h}}$:

   We first find the mode $\tilde{\boldsymbol{h}}$ with a Newton-Raphson scheme. Noting that in higher dimensions, the Newton-Raphson method can be generalised to the iterative

scheme $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - [Hf(\boldsymbol{x}_n)]^{-1}\nabla f(\boldsymbol{x}_n)$ where $Hf(\boldsymbol{x}_n)$ is the Hessian evaluated at $\boldsymbol{x}_n$ and $\nabla f(\boldsymbol{x}_n)$ is the gradient evaluated at $\boldsymbol{x}_n$, then

$$-\frac{1}{2}\frac{d}{d\boldsymbol{h}}(\boldsymbol{h}'\boldsymbol{K}_h\boldsymbol{h} - 2\boldsymbol{h}'\boldsymbol{k}_h) = -\boldsymbol{h}'\boldsymbol{K}_h + \boldsymbol{k}_h$$
$$-\frac{1}{2}\frac{d^2}{d\boldsymbol{h}^2}(\boldsymbol{h}'\boldsymbol{K}_h\boldsymbol{h} - 2\boldsymbol{h}'\boldsymbol{k}_h) = -\boldsymbol{K}_h$$

Therefore:

$$\boldsymbol{h}_{n+1} = \boldsymbol{h}_n + \boldsymbol{K}_h^{-1}(-\boldsymbol{K}_h\boldsymbol{h}_n + \boldsymbol{k}_h) = \boldsymbol{K}_h^{-1}\boldsymbol{k}_h$$

is repeated until some condition is satisfied. [2],

2. Sample $\boldsymbol{h}^*$ using a modified Acceptance-Rejection method:
   A modified version of the Acceptance-Rejection Metropolis-Hastings (ARMH) of Chib and Greenberg (1995) is used. Denote the density of $\boldsymbol{h}$ as $p(\boldsymbol{h}|\boldsymbol{Y}^*, \boldsymbol{\theta}) \propto f(\boldsymbol{Y}^*|\boldsymbol{h}, \boldsymbol{\theta})f(\boldsymbol{h}|\boldsymbol{\theta})$. We first make a draw $\boldsymbol{h}^*$ and accept with probability α_{AR} as

$$\alpha_{AR} = \min\left[\frac{p(\boldsymbol{h}^*|\boldsymbol{Y}^*, \boldsymbol{\theta})}{p_N(\boldsymbol{h}^*|\tilde{\boldsymbol{h}}, \boldsymbol{K}_h^{-1})}, 1\right]$$

   where $p_N(\cdot|\boldsymbol{m}, \boldsymbol{C})$ is a Gaussian proposal with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{C}$. We keep repeating Step 2 until a suitable $\boldsymbol{h}^*$ is accepted.

3. Acceptance/Rejection using a modified Metropolis-Hastings step:
   Define the set $\mathcal{S}$ as:

$$\mathcal{S} = \left\{\boldsymbol{h} : p(\boldsymbol{h}|\boldsymbol{Y}^*, \boldsymbol{\theta}) - p_N(\boldsymbol{h}|\tilde{\boldsymbol{h}}, \boldsymbol{K}_h^{-1}) \leq 0\right\}$$

   (a) if $\boldsymbol{h} \in \mathcal{S}$, set $\alpha_{MH} = 1$

   (b) if $\boldsymbol{h} \in \mathcal{S}^c$ and $\boldsymbol{h}^* \in \mathcal{S}$, set

$$\alpha_{MH} = \frac{p_N(\boldsymbol{h}|\tilde{\boldsymbol{h}}, \boldsymbol{K}_h^{-1})}{p(\boldsymbol{h}|\boldsymbol{Y}^*, \boldsymbol{\theta})}$$

[2] We find good trade-off between computational time and accuracy with $\max\left\{|\tilde{\boldsymbol{h}}^{n+1} - \tilde{\boldsymbol{h}}^n|\right\} < 10^{-4}$.

(c) If $\boldsymbol{h} \in \mathcal{S}^c$ and $\boldsymbol{h}^* \in \mathcal{S}^c$, set

$$\alpha_{MH} = \min\left\{\frac{p(\boldsymbol{h}^*|\boldsymbol{Y}^*,\boldsymbol{\theta})}{p_N(\boldsymbol{h}^*|\tilde{\boldsymbol{h}},\boldsymbol{K}_h^{-1})}\frac{p_N(\boldsymbol{h}|\tilde{\boldsymbol{h}},\boldsymbol{K}_h^{-1})}{p(\boldsymbol{h}|\boldsymbol{Y}^*,\boldsymbol{\theta})},1\right\}$$

Finally, accept $\boldsymbol{h}^*$ with probability α_{MH}.

Sampling of α

We use $\alpha \sim \mathcal{N}(\mu_\alpha,\sigma_\alpha^2)$. Note that a vague prior is most commonly used in the literature. We assume the time series are percentage log returns so that it is assumed $\mu_\alpha = 0$, and use $\sigma_\alpha^2 = \sqrt{10}$ similar to Kim, Shephard, and Chib (1998). The posterior distribution of α is easily derived as

$$p(\alpha|h_1,\boldsymbol{h},\boldsymbol{\xi},\beta,\sigma^2,\rho,\boldsymbol{Y}^*) \propto f(h_1|\xi_1,\alpha,\beta,\sigma^2)\prod_{t=1}^{T} f(h_{t+1}|h_t,\xi_{t+1},\alpha,\beta,\sigma^2) \times \prod_{t=1}^{T} f(y_t^*|h_{t+1},h_t,\xi_{t+1},\alpha,\beta,\sigma,\rho)\times f(\alpha).$$

Thus $p(\alpha|h_1,\boldsymbol{h},\boldsymbol{\xi},\beta,\sigma^2,\rho,\boldsymbol{Y}^*) \sim \mathrm{N}(V_\alpha M_\alpha, V_\alpha)$ where

$$M_\alpha = \frac{(1-\beta^2)h_1}{\sigma^2\xi_1} + \frac{1-\beta}{\sigma^2}\sum_{t=1}^{T}\frac{h_{t+1}-\beta h_t}{\xi_{t+1}} + \frac{(1-\beta)\rho}{(1-\rho^2)\sigma}\sum_{t=1}^{T}\frac{1}{\xi_{t+1}}\left[\frac{\rho}{\sigma}(h_{t+1}-\beta h_t) - \frac{y_t^*}{e^{h_t/2}}\right] + \frac{\mu_\alpha}{\sigma_\alpha^2}$$

$$V_\alpha = \left(\frac{1-\beta^2}{\sigma^2\xi_1} + \frac{(1-\beta)^2}{\sigma^2}\sum_{t=1}^{T}\frac{1}{\xi_{t+1}} + \frac{(1-\beta)^2\rho^2}{(1-\rho^2)\sigma^2}\sum_{t=1}^{T}\frac{1}{\xi_{t+1}} + \frac{1}{\sigma_\alpha^2}\right)^{-1}.$$

We find the Gibbs sampling of α is inefficient as the sampler is unable to sample the posterior correctly typically in the case of low values of ν. We instead favor an adaptive Random-Walk Metropolis algorithm, which has efficient results.

Sampling of β

The unconditional likelihood of β is intractable due to the inclusion of the marginal likelihood of $f(h_1|\beta,\sigma^2,\xi_1)$. To resolve this issue, the candidate density is set to the conditional likelihood, and the target density equal to $f(h_1)$; see Chib and Greenberg (1994) for further details. Thus

$$p(\beta|h_1,\boldsymbol{h},\boldsymbol{\xi},\alpha,\sigma^2,\rho,\boldsymbol{Y}^*) \propto \prod_{t=1}^{T} f(h_{t+1}|h_t,\xi_{t+1},\alpha,\beta,\sigma^2)\prod_{t=1}^{T} f(y_t^*|h_{t+1},h_t,\xi_{t+1},\alpha,\beta,\sigma,\rho)\, f(\beta).$$

Hence, $p(\beta|h_1, \boldsymbol{h}, \boldsymbol{\xi}, \alpha, \sigma^2, \rho, \boldsymbol{Y}^*) \sim \text{N}(V_\beta M_\beta, V_\beta)$ where:

$$M_\beta = \frac{1}{\sigma^2}\sum_{t=1}^{T}\frac{(h_{t+1}-\alpha)(h_t-\alpha)}{\xi_{t+1}} + \frac{\rho}{\sigma(1-\rho^2)}\sum_{t=1}^{T}\frac{h_t-\alpha}{\xi_{t+1}}\left[\frac{\rho}{\sigma}(h_{t+1}-\alpha) - \frac{y_t^*}{e^{h_t/2}}\right] + \frac{\mu_\beta}{\sigma_\beta^2},$$

$$V_\beta = \left(\frac{1}{\sigma^2}\sum_{t=2}^{T}\frac{(h_t-\alpha)^2}{\xi_{t+1}} + \frac{\rho^2}{\sigma^2(1-\rho^2)}\sum_{t=1}^{T}\frac{(h_t-\alpha)^2}{\xi_{t+1}} + \frac{1}{\sigma_\beta^2}\right)^{-1}.$$

We accept β' with probability $\min\left\{\frac{q(\beta')}{q(\beta^{(i-1)})}, 1\right\}$, where $q(\cdot)$ is

$$q(x) = p(h_1|\xi_1, \alpha, x, \sigma^2) = \frac{(1-x^2)^{\frac{1}{2}}}{\sqrt{2\pi}\sqrt{\xi_1}\sigma}\exp\left\{-\frac{(h_1-\alpha)(1-x^2)}{2\xi_1\sigma^2}\right\} \propto (1-x^2)^{\frac{1}{2}}\exp\left\{-\frac{(h_1-\alpha)(1-x^2)}{2\xi_1\sigma^2}\right\}$$

The most typical scenario found in financial time series is β to be close to 1, so we set $\beta \sim N(0.99, 0.2)$.

Sampling of σ^2

To sample σ^2, a modified version of the maximisation at posterior (MAP) sampler is used, which is outlined below:

1. Derive the posterior log-density of σ^2 as

$$\begin{aligned}
&\log p(\sigma^2|h_1, \boldsymbol{h}, \boldsymbol{\xi}, \alpha, \beta, \rho, \boldsymbol{Y}^*) \\
\approx\ &\log f(h_1|\xi_1, \alpha, \beta, \sigma^2) + \sum_{t=1}^{T} \log f(h_{t+1}|h_t, \xi_{t+1}, \alpha, \beta, \sigma^2) + \\
&\sum_{t=1}^{T} \log f(y_t^*|h_{t+1}, h_t, \xi_{t+1}, \alpha, \beta, \sigma, \rho) + c_2 \\
\approx\ &-\frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2} - \frac{1}{2\sigma^2}\sum_{t=1}^{T}[h_{t+1}-\alpha-\beta(h_t-\alpha)]^2 - (T+1)\log\sigma \\
&-\sum_{t=1}^{T}\frac{\{\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}[h_{t+1}-\alpha-\beta(h_t-\alpha)]\}^2}{2(1-\rho^2)\xi_{t+1}}.
\end{aligned}$$

   Then, perform non-linear least squares optimisation to find the mode of the log-density, denoted as $\tilde{\sigma}^2$.

2. Find parameters a^* and b^* of the inverse gamma distribution IG(a^*, b^*) by matching the mode $\tilde{\sigma}^2$ and an assumed variance of 0.01 to that of the IG distribution and solving this system of linear equations.

3. Sample $\sigma^{2^*} \sim$ IG(a^*, b^*) and accept/reject using the Random walk Metropolis algorithm.

Sampling of ρ

To sample ρ, we use a Gaussian prior such that $\rho \sim \mathrm{N}(\mu_\rho, \sigma_\rho^2)$. It is simple to see that

$$
\begin{aligned}
& p(\rho|h_1, \boldsymbol{h}, \boldsymbol{\xi}, \alpha, \beta, \sigma^2, \boldsymbol{Y}^*) \\
\propto & f(h_1|\xi_1, \alpha, \beta, \sigma^2) \prod_{t=1}^{T} f(h_{t+1}|h_t, \xi_{t+1}, \alpha, \beta, \sigma^2) \prod_{t=1}^{T} f(y_t^*|h_{t+1}, h_t, \xi_{t+1}, \alpha, \beta, \sigma, \rho)\, f(\rho) \\
= & \frac{(1-\beta^2)^{-\frac{1}{2}} \xi_1^{-\frac{1}{2}}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2\xi_1}\right\} \times \frac{1}{(\sqrt{2\pi}\sigma)^T} \exp\left\{-\sum_{t=1}^{T} \frac{[h_{t+1}-\alpha-\beta(h_t-\alpha)]^2}{2\sigma^2}\right\} \times \\
& \frac{1}{[2\pi(1-\rho^2)]^{T/2} \prod_{t=1}^{T} \xi_{t+1}} \exp\left\{-\sum_{t=1}^{T} \frac{\{\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}[h_{t+1}-\alpha-\beta(h_t-\alpha)]\}^2}{2(1-\rho^2)\xi_{t+1}}\right\} \times \\
& \frac{1}{\sqrt{2\pi}\sigma_\rho} \exp\left\{-\frac{(\rho-\mu_\rho)^2}{2\sigma_\rho^2}\right\} \\
\propto & \frac{1}{(1-\rho^2)^{T/2}} \exp\left\{-\sum_{t=1}^{T} \frac{\{\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}[h_{t+1}-\alpha-\beta(h_t-\alpha)]\}^2}{2(1-\rho^2)} - \frac{\rho^2}{2\sigma_\rho^2} + \frac{2\mu_\rho\rho}{2\sigma_\rho^2}\right\}.
\end{aligned}
$$

Then ρ is estimated using the MAP sampler, and accepted/rejected using the Metropolis Hastings sampling scheme. We assume $\rho \sim \mathrm{N}(-0.1, 0.05)$.

Sampling of $\boldsymbol{\xi}$

To sample $(\xi_1, \boldsymbol{\xi})$ first note that each element is independent and $\xi_t \sim \text{IG}(\frac{\nu}{2}, \frac{\nu}{2})$. When $t = 0$,

$$p(\xi_1|h_1, \alpha, \beta, \sigma^2, \nu) \propto f(h_1|\xi_1, \alpha, \beta, \sigma^2)\, f(\xi_1)$$

so that:

$$\xi_1|h_1, \alpha, \beta, \sigma^2, \nu \ \sim \text{IG}\left(\frac{\nu+1}{2}, \frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2} + \frac{\nu}{2}\right).$$

Similarly, when $t > 0$, we have

$$p(\xi_{t+1}|h_{t+1}, h_t, \alpha, \beta, \sigma^2, \rho, \nu, y_t^*) \propto f(h_{t+1}|h_t, \xi_{t+1}, \alpha, \beta, \sigma^2) f(y_t^*|h_{t+1}, h_t, \xi_{t+1}, \alpha, \beta, \sigma, \rho)\, f(\xi_{t+1})$$

$$\xi_{t+1}|h_{t+1}, h_t, \alpha, \beta, \sigma^2, \rho, \nu, y_t^* \ \sim \text{IG}\left(\frac{\nu}{2}+1, S_\xi\right)$$

where $S_\xi = \frac{[h_{t+1}-\alpha-\beta(h_t-\alpha)]^2}{2\sigma^2} + \frac{\{\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}[h_{t+1}-\alpha-\beta(h_t-\alpha)]\}^2}{2(1-\rho^2)} + \frac{\nu}{2}$.

Sampling of ν

In order to sample ν, we implement an adaptive independence-chain Metropolis-Hastings algorithm. Assuming the prior $\nu \ \sim \text{U}[\nu^-, \nu^+]$, the density of the posterior distribution of ν is

$$\begin{aligned} p(\nu|\boldsymbol{\xi}) &\propto f(\boldsymbol{\xi}|\nu)\, f(\nu) \\ &\propto \prod_{t=1}^{T}\left[\frac{(\frac{\nu}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}\xi_t^{-\frac{\nu}{2}-1}\exp\left(-\frac{\nu}{2\xi_t}\right)\right] = \frac{(\frac{\nu}{2})^{\frac{T\nu}{2}}}{[\Gamma(\frac{\nu}{2})]^T}\left(\prod_{t=1}^{T}\xi_t^{-\frac{\nu}{2}-1}\right)\exp\left(-\sum_{t=1}^{T}\frac{\nu}{2\xi_t}\right)\ I(0<\nu<b_\nu) \end{aligned}$$

so that the log posterior density of ν is

$$\log p(\nu|\boldsymbol{\xi}) \approx \frac{T\nu}{2}\log\left(\frac{\nu}{2}\right) - T\log\Gamma\left(\frac{\nu}{2}\right) - \left(\frac{\nu}{2}+1\right)\sum_{t=1}^{T}\log\xi_t - \frac{\nu}{2}\sum_{t=1}^{T}\xi_t^{-1}$$

where

$$\frac{d \log p(\nu|\boldsymbol{\xi})}{d\nu} = \frac{T}{2}\log\left(\frac{\nu}{2}\right) + \frac{T}{2} - \frac{T}{2}\psi\left(\frac{\nu}{2}\right) - \frac{1}{2}\sum_{t=1}^{T}\log\xi_t - \frac{1}{2}\sum_{t=1}^{T}\xi_t^{-1}$$
$$\frac{d^2 \log p(\nu|\boldsymbol{\xi})}{d\nu^2} = \frac{T}{2\nu} - \frac{T}{4}\psi_1\left(\frac{\nu}{2}\right)$$

where ψ and ψ_1 are the digamma and trigamma functions respectively. We maximise the density in order to find the mode $\tilde{\nu}$ using routine optimisation methods. Although an Inverse Gamma proposal is typically used, we find this leads to boundary trap issues. We instead find superior results using a Gaussian proposal $\nu \sim \mathrm{N}(\tilde{\nu}, V_\nu c_\nu)$, where $V_\nu = 1$ and c_ν is a tuning parameter initialized as 1. The choice of hyperparmeters are $\nu^- = 3$ to avoid infinite 'explosive' variance issues, and $\nu^+ = 23$.

Bibliography

Abad, Pilar, Sonia Muela, and Carmen Martin (2015). "The role of the loss function in value-at-risk comparisons". In: *The Journal of Risk Model Validation* 9.1, p. 1.

Ait-Sahalia, Yacine and Robert Kimmel (2007). "Maximum likelihood estimation of stochastic volatility models". In: *Financial Econometrics* 83.2, pp. 413–452.

Alizadeh, Sassan, Michael W Brandt, and Francis X Diebold (2002). "Range-Based Estimation of Stochastic Volatility Models". In: *The Journal of Finance* 57.3, pp. 1047–1091.

Andersen, Torben G and Tim Bollerslev (1998). "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts". In: *International Economic Reviews.* 39.4, pp. 885–905.

Andersen, Torben G, Hyung-Jin Chung, and Bent E Sørensen (1999). "Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study". In: *The Journal of Econometrics* 91.1, pp. 61–87.

Andersen, Torben G et al. (2001). "The Distribution of Realized Exchange Rate Volatility". In: *The Journal of the American Statistical Association.* 96.453, pp. 42–55.

– (2003). "Modeling and Forecasting Realized Volatility". In: *Econometrica* 71.2, pp. 579–625.

Andrews, D F and C L Mallows (1974). "Scale Mixtures of Normal Distributions". In: *The Journal of the Royal Statistics Society Series B Statistical Methodology* 36.1, pp. 99–102.

Artiach, Miguel and Josu Arteche (2012). "Doubly fractional models for dynamic heteroscedastic cycles". In: *Computational Statistics & Data Analysis* 56.6, pp. 2139–2158.

Asai, Manabu (2008). "Autoregressive stochastic volatility models with heavy-tailed distributions: A comparison with multifactor volatility models". In: *The Journal of Empirical Finance* 15.2, pp. 332–341.

Asai, Manabu and Michael McAleer (2005). "Dynamic Asymmetric Leverage in Stochastic Volatility Models". In: *Econometric Reviews* 24.3, pp. 317–332.

Atkeson, Andrew and Lee E Ohanian (2001). "Are Phillips curves useful for forecasting inflation?" In: *Federal Reserve Bank of Minneapolis. Quarterly Review-Federal Reserve Bank of Minneapolis* 25.1, p. 2.

Aye, Goodness C. et al. (2014). "Predicting BRICS stock returns using ARFIMA models". In: *Applied Financial Economics.* 24.17, pp. 1159–1166.

Baillie, Richard T. (1996). "Analysing inflation by the fractionally integrated ARFIMA-GARCH model". In: *The Journal of Applied Econometrics* 11.1, pp. 23–40.

Baillie, Richard T., Tim Bollerslev, and Hans O. Mikkelsen (1996). "Fractionally integrated generalized autoregressive conditional heteroskedasticity". In: *The Journal of Econometrics* 74.1, pp. 3–30.

Ball, Clifford A and Walter N Torous (1985). "On Jumps in Common Stock Prices and Their Impact on Call Option Pricing". In: *The Journal of Finance* 40.1, pp. 155–173.

Bardet, Jean-March et al. (2003). "Theory and Applications of Long-Range Dependance". In: pp. 579–623.

Bariviera, Aurelio F (2017). "The inefficiency of Bitcoin revisited: A dynamic approach". In: *Economic Letters* 161, pp. 1–4.

Barndorff-Nielsen, Ole E (1997). "Normal inverse Gaussian distributions and stochastic volatility modelling". In: *Scandinavian Journal of statistics* 24.1, pp. 1–13.

Barndorff-Nielsen, O E and others (2001). "Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics". In: *The Journal of the Royal Statistical Society*.

Bates, David S (1996). "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options". In: *Review of Financial Studies* 9.1, pp. 69–107.

Bensoussan, Alain, Michel Crouhy, and Dan Galai (1994). "Stochastic equity volatility related to the leverage effect". In: *The Journal of Applied Mathematical Finance* 1.1, pp. 63–85.

Berg, Andreas, Renate Meyer, and Jun Yu (2004). "Deviance Information Criterion for Comparing Stochastic Volatility Models". In: *The Journal of Business Economics and Statistics* 22.1, pp. 107–120.

Besag, J E and J C York (1989). "Bayesian restoration of images". In: *Analysis of Statistical Information*, pp. 491–507.

Bhardwaj, Geetesh and Norman R. Swanson (2006). "An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series". In: *The Journal of Econometrics* 131.1–2, pp. 539–578.

Bianchi, Daniele (June 2018). "Cryptocurrencies As an Asset Class? An Empirical Assessment". In:

Black, F (1976). "Studies in stock price volatility changes". In: *Proceedings of the 1976 Business meeting of the business and economic statistics association*, pp. 177–181.

Black, Fischer and Myron Scholes (1973). "The Pricing of Options and Corporate Liabilities". In: *The Journal of Political Economics.* 81.3, pp. 637–654.

Bollerslev, T and J M Wooldridge (1992). "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances". In: *Econometric Review*.

Bollerslev, Tim (1986). "Generalized autoregressive conditional heteroskedasticity". In: *The Journal of Econometrics* 31.3, pp. 307–327.

Bordignon, Silvano, Massimiliano Caporin, and Francesco Lisi (2007). "Generalised long-memory GARCH models for intra-daily volatility". In: *Computational Statistics & Data Analysis* 51.12, pp. 5900–5912.

Bos, Charles S., Siem Jan Koopman, and Marius Ooms (2014a). "Long memory with stochastic variance model: A recursive analysis for US inflation". In: *Computational Statistics & Data Analysis* 76, pp. 144–157.

Bos, Charles S, Siem Jan Koopman, and Marius Ooms (2014b). "Long memory with stochastic variance model: A recursive analysis for US inflation". In: *Computational Statistics and Data Analysis* 76.0, pp. 144–157.

Bouchaud, J P, A Matacz, and M Potters (2001). "Leverage effect in financial markets: the retarded volatility model". In: *Physics Review Letters* 87.22, p. 228701.

Box, George E P and George C Tiao (2011). *Bayesian Inference in Statistical Analysis*. John Wiley & Sons.

Capobianco, E (1996). "State-space stochastic volatility models: A review of estimation algorithms". In: *Applied Stochastic Models and Data Analysis*.

Carlos, Juan C. and Luis A. Gil-Alana (2016). "Testing for long memory in the presence of non-linear deterministic trends with Chebyshev polynomials." In: *Studies in Nonlinear Dynamics & Econometrics* 20.1, pp. 57–74. ISSN: 10811826.

Carnero, M A, D Pena, and E Ruiz (2003). "Why is GARCH more persistent and conditionally leptokurtic than Stochastic Volatility". In: *Unpublished, Universidad Carlos III de Madrid*.

Carnero, M Angeles, Daniel Peña, and Esther Ruiz (2004). "Persistence and Kurtosis in GARCH and Stochastic Volatility Models". In: *The Journal of Financial Econometrics* 2.2, pp. 319–342.

Catania, Leopoldo and Stefano Grassi (Aug. 2017). "Modelling Crypto-Currencies Financial Time-Series". In: *SSRN working paper*.

Catania, Leopoldo, Stefano Grassi, and Francesco Ravazzolo (2018). "Forecasting Cryptocurrencies Financial Time Series". In: *Unpublished*.

Chan, Jennifer SK, ST Boris Choy, and Connie PY Lam (2014). "Modeling electricity price using a threshold conditional autoregressive geometric process jump model". In: *Communications in Statistics-Theory and Methods* 43.10-12, pp. 2505–2515.

Chan, Joshua C C and Angelia L Grant (2016a). "Modeling energy price dynamics: GARCH versus stochastic volatility". In: *Energy Economics* 54, pp. 182–189.

– (2016b). "Modeling energy price dynamics: GARCH versus stochastic volatility". In: *Energy Economics* 54, pp. 182–189.

Chan, Joshua C C and Cody Y L Hsiao (2014). "Estimation of Stochastic Volatility Models with Heavy Tails and Serial Dependence". In: *Bayesian Inference in the Social Sciences*. John Wiley & Sons, Inc., pp. 155–176.

Chan, Joshua C C and Rodney W Strachan (2012). "Estimation in Non-Linear Non-Gaussian State Space Models with Precision-Based Methods". In: *Unpublished, in the SSRN Electronic Journal*.

Chan, Joshua C.C. (2013). "Moving average stochastic volatility models with application to inflation forecast". In: *The Journal of Econometrics* 176.2, pp. 162–172.

Chan, Ngai H. and Wilfredo Palma (1998). "State Space Modeling of Long-Memory Processes". In: *The Annals of Statistics* 26.2, pp. 719–740.

Chaum, David L (1981). "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms". In: *Communications in ACM* 24.2, pp. 84–90.

Cheah, Eng-Tuck and John Fry (May 2015). "Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin". In: *Economic Letters* 130, pp. 32–36.

Chen, Cathy W S, Feng-Chi Liu, and Mike K P So (2008a). "Heavy-tailed distributed threshold stochastic volatility models in financial time series". In: *The Australian and New Zealand Journal of Statistics* 50.1, pp. 29–51.

Chen, Cathy W S and Buu-Chau Truong (2016). "On double hysteretic heteroskedastic model". In: *The Journal of Statistical Computation and Simulation*. 86.13, pp. 2684–2705.

Chen, Cathy WS, Feng-Chi Liu, and Mike KP So (2008b). "Heavy-tailed-distributed threshold stochastic volatility models in financial time series". In: *Australian & New Zealand Journal of Statistics* 50.1, pp. 29–51.

Chernov, Mikhail et al. (Sept. 2003). "Alternative models for stock price dynamics". In: *The Journal of Econometrics* 116.1, pp. 225–257.

Cheung, Yin-Wong (1993). "Long Memory in Foreign-Exchange Rates". In: *The Journal of Business & Economic Statistics* 11.1, pp. 93–101.

Chib, Siddhartha and Edward Greenberg (1994). "Bayes inference in regression models with ARMA(p, q) errors". In: *The Journal of Econometrics* 64.1–2, pp. 183–206.

– (1995). "Understanding the Metropolis-Hastings Algorithm". In: *The Journal of the American Statistical Association* 49.4, pp. 327–335.

Chib, Siddhartha, Federico Nardari, and Neil Shephard (2002). "Markov chain Monte Carlo methods for stochastic volatility models". In: *The Journal of Econometrics* 108.2, pp. 281–316.

Choy, Boris and Jennifer S K Chan (2008). "Scale mixtures distributions in statistical modelling". In: *The Australian and New Zealand Journal of Statistics* 50.2, pp. 135–146.

Choy, S T B and C M Chan (2000). "Bayesian estimation of stochastic volatility model via scale mixture distributions". In: *Statistics and Finance*. Published by Imperial College press and distributed by world scientific publishing co., pp. 185–204.

Choy, S T Boris, Cathy W S Chen, and Edward M H Lin (July 2014). "Bivariate asymmetric GARCH models with heavy tails and dynamic conditional correlations". In: *Quant. Finance* 14.7, pp. 1297–1313.

Christie, Andrew A (1982). "The stochastic behavior of common stock variances: Value, leverage and interest rate effects". In: *The Journal of Financial Economics* 10.4, pp. 407–432.

Clark, Peter K (1973). "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices". In: *Econometrica* 41.1, pp. 135–155.

Conrad, Christian and Menelaos Karanasos (2005). "Dual Long Memory in Inflation Dynamics across Countries of the Euro Area and the Link between Inflation Uncertainty and Macroeconomic Performance." In: *Studies in Nonlinear Dynamics & Econometrics* 9.4, pp. 1–36.

Corsi, Fulvio (2004). "A Simple Long Memory Model of Realized Volatility". In:

– (2009). "A Simple Approximate Long-Memory Model of Realized Volatility". In: *The Journal of Financial Econometrics* 7.2, pp. 174–196.

Crato, Nuno and Philip Rothman (1994). "Fractional integration analysis of long-run behavior for US macroeconomic time series". In: *Economics Letters* 45.3, pp. 287–291.

Dissanayake, G.S., M.S. Peiris, and T. Proietti (2016). "State space modeling of Gegenbauer processes with long memory". In: *Computational Statistics & Data Analysis* 100, pp. 115–130.

Doornik, Jurgen A. and Henrik Hansen (2008). "An Omnibus Test for Univariate and Multivariate Normality*". In: *Oxford Bulletin of Economics and Statistics* 70, pp. 927–939. ISSN: 1468-0084.

Engle, Robert (1995). *ARCH: Selected Readings*. Oxford University Press.

Engle, Robert F (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* 50.4, pp. 987–1007.

Engle, Robert F and Victor K Ng (1993). "Measuring and Testing the Impact of News on Volatility". In: *The Journal of Finance* 48.5, pp. 1749–1778.

Fei-xue, Huang, Zhao Yan, and Hou Tie-shan (2009). "Long memory and leverage effect of Euro exchange rate based on ARFIMA-FIEGARCH". In: *Management Science and Engineering*. Pp. 1416–1421.

Gallant, A Ronald, David A Hsieh, and George E Tauchen (1991). "On fitting a recalcitrant series: the pound/dollar exchange rate, 1974-1983". In: *Nonparametric and semiparametric methods in econometrics and statistics*, pp. 199–240.

Garman, Mark B and Michael J Klass (1980). "On the estimation of security price volatilities from historical data". In: *Journal of business*, pp. 67–78.

Gelfand, Alan E and Adrian F M Smith (1990). "Sampling-Based Approaches to Calculating Marginal Densities". In: *The Journal of the American Statistical Association*. 85.410, pp. 398–409.

Gelman, Andrew (Sept. 2006). "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)". In: *Bayesian Analysis* 1.3, pp. 515–534.

Gelman, Andrew et al. (2013). *Bayesian Data Analysis*. 3rd. Hoboken: CRC Press.

Geman, S and D Geman (1984). "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images". In: *IEEE Transnational Pattern Analysis.* 6.6, pp. 721–741.

Geweke, John and Gianni Amisano (2011). "Hierarchical Markov normal mixture models with applications to financial asset returns". In: *The Journal of Applied Econometrics* 26.1, pp. 1–29.

Geweke, John and Susan Porter-Hudak (1983). "The estimation and application of long-memory time series models". In: *The Journal of Time Series Analysis* 4.4, pp. 221–238.

Geyer, Charles J (1991). "Markov chain Monte Carlo maximum likelihood". In:

Ghysels, E and J Jasiak (1994). *Stochastic Volatility and time Deformation: an Application of trading Volume and Leverage Effects*. papyrus.bib.umontreal.ca.

Ghysels, Eric and Pierre Perron (1996). "The effect of linear filters on dynamic time series with structural change". In: *The Journal of Econometrics* 70.1, pp. 69–97.

Giacomini, Raffaella and Ivana Komunjer (Oct. 2005). "Evaluation and Combination of Conditional Quantile Forecasts". In: *J. Bus. Econ. Stat.* 23.4, pp. 416–431.

Gil-Alana, Luis A. (2002). "Modelling the Persistence of Unemployment in Canada". In: *The International Review of Applied Economics.* 16.4, pp. 465–477.

Gil-Alana, L. A. and J. Toro (2002). "Estimation and testing of ARFIMA models in the real exchange rate". In: *The International Journal of Finance & Economics.* 7.4, pp. 279–292.

Goldman, Elena et al. (2013). "Regimes and long memory in realized volatility." In: *Studies in Nonlinear Dynamics & Econometrics* 17.5, pp. 521–549.

Granger, Clive W.J. (1980). "Long memory relationships and the aggregation of dynamic models". In: *The Journal of Econometrics* 14.2, pp. 227–238.

Granger, Clive W.J. and Roselyne Joyeux (1980). "An introduction to long-memory time series and fractional differencing". In: *The Journal of Time Series Analysis* 1.1, pp. 15–29.

Gray, Henry L., Nien-Fan Zhang, and Wayne A. Woodward (1989a). "On generalized fractional processes". In: *The Journal of Time Series Analysis* 10.3, pp. 233–257.

Gray, Henry L, Nien-Fan Zhang, and Wayne A Woodward (May 1989b). "On generalized fractional processes". In: *The Journal of Time Series Analysis* 10.3, pp. 233–257.

Guégan, Dominique (2005). "How can we Define the Concept of Long Memory? An Econometric Survey". In: *Econometric Reviews* 24.2, pp. 113–149.

Haario, Heikki, Eero Saksman, and Johanna Tamminen (2001). "An adaptive Metropolis algorithm". In: *Bernoulli* 7.2, pp. 223–242.

Harvey, Andrew, Esther Ruiz, and Neil Shephard (1994). "Multivariate Stochastic Variance Models". In: *Review of Economic Studies* 61.2, pp. 247–264.

Hastings, W K (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1, p. 97.

Hauser, Michael and Robert Kunst (1998). "Fractionally Integrated Models With ARCH Errors: With an Application to the Swiss 1-Month Euromarket Interest Rate". In: *Review of Quantitative Finance and Accounting* 10.1, pp. 95–113.

Hencic, Andrew and Christian Gouriéroux (2015). "Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rates". In: *Econometrics of Risk*. Ed. by Van-Nam Huynh et al. Cham: Springer International Publishing, pp. 17–40.

Hosking, J.R.M (1981). "Fractional differencing". In: *Biometrika* 68.1, pp. 165–176.

Hotz-Behofsits, Christian, Florian Huber, and Thomas O Zörner (Jan. 2018). "Predicting crypto-currencies using sparse non-Gaussian state space models". In: arXiv: `1801.06373 [econ.EM]`.

Hwang, Eunju and Dong W. Shin (2014). "Infinite-order, long-memory heterogeneous autoregressive models". In: *Computational Statistics & Data Analysis* 76, pp. 339–358.

Iglesias, Pilar, Hector Jorquera, and Wilfredo Palma (2006). "Data analysis using regression models with missing observations and long-memory: an application study". In: *Computational Statistics & Data Analysis* 50.8, pp. 2028–2043.

Jacquier, Eric, Nicholas G Polson, and Peter E Rossi (1994). "Bayesian Analysis of Stochastic Volatility Models". In: *The Journal of Business Economics and Statistics* 12.4, pp. 371–389.

Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi (2004). "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors". In: *The Journal of Econometrics* 122.1, pp. 185–212.

Jiang, Yonghong, He Nie, and Weihua Ruan (2017). "Time-varying long-term memory in Bitcoin market". In: *Finance Research Letters*.

Kass, Robert E. and Adrian E. Raftery (1995). "Bayes Factors". In: *The Journal of the American Statistical Association* 90.430, pp. 773–795.

Katsiampa, Paraskevi (2017). "Volatility estimation for Bitcoin: A comparison of GARCH models". In: *Economic Letters* 158, pp. 3–6.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998). "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models". In: *Review of Economic Studies* 65.3, pp. 361–93.

Koopman, Siem Jan, Borus Jungbacker, and Eugenie Hol (June 2005). "Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements". In: *The Journal of Empirical Finance* 12.3, pp. 445–475.

Lahiani, A. and O. Scaillet (2009). "Testing for threshold effect in ARFIMA models: Application to US unemployment rate data". In: *The International Journal of Forecasting* 25.2, pp. 418–428.

Lahmiri, Salim, Stelios Bekiros, and Antonio Salvi (2018). "Long-range memory, distributional variation and randomness of bitcoin volatility". In: *Chaos Solitons Fractals* 107, pp. 43–48.

Li, W K and K Lam (1995). "Modelling Asymmetry in Stock Returns by a Threshold Autoregressive Conditional Heteroscedastic Model". In: *The Journal of the Royal Statistical Society. Series D (The Statistician)* 44.3, pp. 333–341.

Liesenfeld, Roman and Robert C Jung (2000). "Stochastic Volatility Models: Conditional Normality versus Heavy-Tailed Distributions". In: *The Journal of Applied Econometrics* 15.2, pp. 137–160.

Lillo, Fabrizio and J. Doyne Farmer (2004). "The Long Memory of the Efficient Market." In: *Studies in Nonlinear Dynamics & Econometrics* 8.3, pp. 1–33.

Lopes, Sílvia R.C. and Taiane S. Prass (2013). "Seasonal FIEGARCH processes". In: *Computational Statistics & Data Analysis* 68, pp. 262–295.

Mahieu, Ronald and Peter Schotman (1994). *Stochastic volatility and the distribution of exchange rate news*. Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis.

Mandelbrot, Benoit (1969). "Long-Run Linearity, Locally Gaussian Process, H-Spectra and Infinite Variances". In: *International Economic Review* 10.1, pp. 82–111.

Mandelbrot, Benoit B. and John W.V. Ness (1968). “Fractional Brownian Motions, Fractional Noises and Applications”. In: *SIAM Review* 10.4, pp. 422–437.

Markowitz, H (1952). “Portfolio selection”. In: *The Journal of Finance*.

Mascagni, Michael and Ashok Srinivasan (2004). “Parameterizing parallel multiplicative lagged-Fibonacci generators”. In: *Parallel Computing* 30.7, pp. 899–916.

McAleer, Michael and Marcelo C Medeiros (Nov. 2008). “A multiple regime smooth transition Heterogeneous Autoregressive model for long memory and asymmetries”. In: *The Journal of Econometrics* 147.1, pp. 104–119.

Melino, Angelo and Stuart M Turnbull (1990). “Pricing foreign currency options with stochastic volatility”. In: *The Journal of Econometrics* 45.1, pp. 239–265.

Merton, Robert C (1976). “Option pricing when underlying stock returns are discontinuous”. In: *The Journal of Financial Economics* 3.1, pp. 125–144.

Metropolis, Nicholas et al. (1953). “Equation of state calculations by fast computing machines”. In: *The Journal of Chemical Physics* 21.6, pp. 1087–1092.

Meyer, Renate and Jun Yu (2000). “BUGS for a Bayesian analysis of stochastic volatility models”. In: *The Journal of Econometrics.* 3.2, pp. 198–215.

Mikhail, O., C. J. Eberwein, and J. Handa (2006). “Estimating persistence in Canadian unemployment: evidence from a Bayesian ARFIMA”. In: *Applied Economics* 38.15, pp. 1809–1819.

Nakajima, Jouchi and Yasuhiro Omori (2012). “Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student’s t-distribution”. In: *Computational Statistics & Data Analysis* 56.11, pp. 3690–3704.

Nakamoto, Satoshi (2008). *Bitcoin: A peer-to-peer electronic cash system*. www.bitcoin.org.

Nelson, Daniel B (1991). “Conditional Heteroskedasticity in Asset Returns: A New Approach”. In: *Econometrica* 59.2, pp. 347–370.

Omori, Yasuhiro and Toshiaki Watanabe (2008). “Block sampler and posterior mode estimation for asymmetric stochastic volatility models”. In: *Computational Statistics and Data Analysis* 52.6, pp. 2892–2910.

Omori, Yasuhiro et al. (2007). “Stochastic volatility with leverage: Fast and efficient likelihood inference”. In: *The Journal of Econometrics* 140.2, pp. 425–449.

Parkinson, Michael (1980). “The extreme value method for estimating the variance of the rate of return”. In: *Journal of business*, pp. 61–65.

Phillip, Andrew, Jennifer Chan, and Shelton Peiris (2017). ". (in press). Bayesian estimation of Gegenbauer long memory processes with stochastic volatility: methods and applications". In: *Studies in Nonlinear Dynamics and Econometrics*.

Phillip, Andrew, Jennifer S K Chan, and Shelton Peiris (Feb. 2018). "A new look at Cryptocurrencies". In: *Economic Letters* 163, pp. 6–9.

Raggi, Davide and Silvano Bordignon (2012). "Long memory and nonlinearities in realized volatility: A Markov switching approach". In: *Computational Statistics & Data Analysis* 56.11, pp. 3730–3742.

Rainville, Earl D. (1960). *Special functions*. New York: Macmillan.

Reisen, Valderio A., Alexandre L. Rodrigues, and Wilfredo Palma (2006). "Estimation of seasonal fractionally integrated processes". In: *Computational Statistics & Data Analysis* 50.2, pp. 568–582.

Ritter, Christian and Martin A Tanner (1992). "Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler". In: *The Journal of the American Statistical Association* 87.419, pp. 861–868.

Robert, ChristianP. (1995). "Simulation of truncated normal variables". In: *Statistics and Computing* 5.2, pp. 121–125.

Roberts, Gareth O., Andrew Gelman, and Walter R. Gilks (1997). "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms". In: *The Annals of Applied Probability* 7.1, pp. 110–120.

Roberts, Gareth O. and Jeffrey S. Rosenthal (2001). "Optimal Scaling for Various Metropolis-Hastings Algorithms". In: *Statistical Science* 16.4, pp. 351–367.

Rogers, L Christopher G and Stephen E Satchell (1991). "Estimating variance from high, low and closing prices". In: *The Annals of Applied Probability*, pp. 504–512.

Rue, Håvard, Ingelin Steinsland, and Sveinung Erland (2004). "Approximating hidden Gaussian Markov random fields". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66.4, pp. 877–892.

Ruiz, Esther (1994). "Quasi-maximum likelihood estimation of stochastic volatility models". In: *The Journal of Econometrics* 63.1, pp. 289–306.

Shephard, N (1996). "Statistical aspects of ARCH and stochastic volatility". In: *Monographs on Statistics and Applied Probability*.

Shephard, Neil (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press.

Shephard, Neil and Michael K Pitt (1997). "Likelihood analysis of non-Gaussian measurement time series". In: *Biometrika* 84.3, pp. 653–667.

Shirota, Shinichiro, Takayuki Hizu, and Yasuhiro Omori (2014). "Realized stochastic volatility with leverage and long memory". In: *Computational Statistics & Data Analysis* 76, pp. 618–641.

Smith, A F M and G O Roberts (1993). "Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods". In: *The Journal of the Royal Statistical Society, Series B Statistical Methodology.* 55.1, pp. 3–23.

So, Mike K P and C Y Choi (2009). "A threshold factor multivariate stochastic volatility model". In: *The Journal of Forecasting* 28.8, pp. 712–735.

So, Mike K.P., K. Lam, and W. K. Li (1998). "A Stochastic Volatility Model With Markov Switching". In: *The Journal of Business & Economic Statistics* 16.2, pp. 244–253.

– (2002). "A threshold stochastic volatility model". In: *The Journal of Forecasting* 21.7, pp. 473–507.

Sowell, Fallaw (1992). "Modeling long-run behavior with the fractional ARIMA model". In: *The Journal of Monetary Economics.* 29.2, pp. 277–302.

Spiegelhalter, David J et al. (2002). "Bayesian measures of model complexity and fit". In: *The Journal of the Royal Statistics Society Series B Statistical Methodology* 64.4, pp. 583–639.

Stock, James H and Mark W Watson (2007). "Why Has U.S. Inflation Become Harder to Forecast?" In: *The Journal of Money Credit Bank* 39, pp. 3–33.

Strasser, Helmut (1975). "The asymptotic equivalence of Bayes and maximum likelihood estimation". In: *The Journal of Multivariate Analysis.* 5.2, pp. 206–226.

Takahashi, Makoto, Yasuhiro Omori, and Toshiaki Watanabe (2009). "Estimating stochastic volatility models using daily returns and realized volatility simultaneously". In: *Computational Statistics and Data Analysis* 53.6, pp. 2404–2426.

Tanner, Martin A and Wing Hung Wong (1987). "The Calculation of Posterior Distributions by Data Augmentation". In: *The Journal of the American Statistical Association.* 82.398, pp. 528–540.

Tauchen, George E and Mark Pitts (1983). "The Price Variability-Volume Relationship on Speculative Markets". In: *Econometrica* 51.2, pp. 485–505.

Taylor, Stephen (1986). *Modelling financial time series*. Wiley.

Taylor, Stephen J (1994). "Modeling stochastic volatility: a review and comparative study". In: *Mathematical Finance* 4.2, pp. 183–204.

Terasvirta, T and H M Anderson (1992). "Characterizing nonlinearities in business cycles using smooth transition autoregressive models". In: *Journal of Applied Econometrics.* 7.

Tierney, Luke (1994). "Markov chains for exploring posterior distributions". In: *the Annals of Statistics*, pp. 1701–1728.

Tong, Howell (1990). *Non-linear time series: a dynamical system approach*. Oxford, UK: Oxford University Press, p. 564.

Tropp, Joel A and Stephen J Wright (2010). "Computational methods for sparse solution of linear inverse problems". In: *Proceedings of the IEEE* 98.6, pp. 948–958.

Turkyilmaz, Serpil and Mesut Balibey (2014). "Long Memory Behavior in the Returns of Pakistan Stock Market: ARFIMA-FIGARCH Models". In: *The International Journal of Economics and Financial Issues* 4.2, pp. 400–410.

Urquhart, Andrew (2016). "The inefficiency of Bitcoin". In: *Economic Letters* 148, pp. 80–82.

– (2017). "Price clustering in Bitcoin". In: *Economic Letters* 159, pp. 145–148.

Walker, Stephen G and Eduardo Gutiérrez-Pena (1999). "Robustifying Bayesian procedures". In: *Bayesian statistics* 6.6, pp. 685–710.

Wang, Joanna J J, Jennifer S K Chan, and S T Boris Choy (2011). "Stochastic volatility models with leverage and heavy-tailed distributions: A Bayesian approach using scale mixtures". In: *Computational Statistics and Data Analysis* 55.1, pp. 852–862.

Wang, Joanna J.J., S. T. Boris Choy, and Jennifer S.K. Chan (2013a). "Modelling stochastic volatility using generalized t-distribution". In: *The Journal of Statistical Computation and Simulation* 83.2, pp. 340–354.

Wang, Joanna JJ, ST Boris Choy, and Jennifer SK Chan (2013b). "Modelling stochastic volatility using generalized t distribution". In: *The Journal of Statistical Computation and Simulation* 83.2, pp. 340–354.

Wirjanto, Tony S, Adam W Kolkiewicz, and Zhongxian Men (2016). "Bayesian Analysis of a Threshold Stochastic Volatility Model". In: *The Journal of Forecasting* 35.5, pp. 462–476.

Woodward, Wayne A, Q C Cheng, and H L Gray (1998). "A k-Factor GARMA Long-memory Model". In: *The Journal of Time Series Analysis* 19.4, pp. 485–504.

Yamauchi, Yuta, Yasuhiro Omori, and Others (2016). *Multivariate Stochastic Volatility Model with Realized Volatilities and Pairwise Realized Correlations*. Tech. rep. CIRJE, Faculty of Economics, University of Tokyo.

Yu, Jun (2005). “On leverage in a stochastic volatility model”. In: *The Journal of Econometrics* 127.2, pp. 165–178.

Zhu, Ke, Philip L H Yu, and Wai Keung Li (2014). “Testing for the buffered autoregressive processes”. In: *Statistica Sinica*.