Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin

Andrew Phillip*, Jennifer Chan, Shelton Peiris

School of Mathematics and Statistics, The University of Sydney, Sydney, Australia

ARTICLE INFO

Article history:
Received 11 January 2018
Revised 22 October 2018
Accepted 22 October 2018
Available online xxx

JEL classification:
C5
C22
G1

Keywords:
Gegenbauer long memory
Stochastic volatility
Leverage
Heavy tails
Cryptocurrency
Bitcoin

ABSTRACT

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 is presented. This is inspired by the deep rooted characteristics found in cryptocurrencies. Until recently their econometric properties have not been thoroughly investigated. Thus, 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 and 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 forecasts focused specifically on Bitcoin (while other forecasts are available as supplementary material). 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. Proposed models are compared with many popular models including those commonly used in industry. The models are applied in a Value-at-Risk (VaR) context and several measures are used to assess model performance.

1. Introduction

Academic interests in anonymous communications research date back to the early 1980s (Chaum, 1981),[1] 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 underlying 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 goals is to avoid central control so each person owns their own copy of the ledger. To date, this has manifested itself into a growing cryptocurrency

* Corresponding author.
E-mail addresses: andrew.phillip@sydney.edu.au (A. Phillip), jennifer.chan@sydney.edu.au (J. Chan), shelton.periris@sydney.edu.au (S. Peiris).

[1] Additional empirical out of sample results on ETH, XRP, NEM and DASH are available as supplementary material.

https://doi.org/10.1016/j.ecosta.2018.10.003
2452-3062/

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.

Cryptocurrencies have gained large media attention as of late and deserve some specialized studies. 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.

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. Details are provided on a large investable sample of cryptocurrencies, and some important stylized facts are conditionally measured. Due to these unique stylized facts, the proposed model contributes to the literature by measuring these effects.

To date, there are more than 2000 investable cryptocurrencies. Unlike their fiat counterparts, different exchange rates 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 are discussed later in Section 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 kth-lag autocovariance. This class of time series generalizes the usual Box-Jenkins ARIMA model by modeling long term correlation structures as suggested by Mandelbrot and Van Ness (1968). An appealing generalization of traditional long memory models are generalized 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 et al. (1989). The novelty in such polynomials lies in its orthogonality and recursion properties. A comprehensive review can be found in Dissanayake et al. (2018).

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 effect 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). 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 and hence leading to higher expected future volatility (Engle and Ng, 1993).

The most heated debate about cryptocurrencies is 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 non-Gaussian heavy-tailed distributions (Chib et al., 2002; Berg et al., 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 modeling 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 generalized autoregressive conditional heteroscedastic (GARCH) model of Bollerslev (1986) because it adequately displays the main stylized-facts found in the daily returns of financial returns (Carnero et al., 2004). See Engle (1995) and Shephard (2005) for a detailed account and comparison of the two approaches. Here, the proposed Gegenbauer long memory stochastic volatility model with leverage and bivariate Student's t model is a generalization of 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 et al. (2014).

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) who used a generalized method of moments to price currency options, Harvey et al. (1994) who applied a quasi maximum likelihood approach in a multivariate context, and also Ait-Sahalia and Kimmel (2007) who performed Monte Carlo simulations to estimate short-dated at-the-money options. Over the last twenty years, the Bayesian approach has become popular due to cheap computing power with many efficient samplers, for example, the Metropolis algorithm (Jacquier et al., 1994), importance sampling (Shephard and Pitt, 1997) and algorithm using normal mixtures (Kim et al., 1998). Recently Chan and Grant (2016) 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

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

upon in the last ten years due to the Global Financial Crisis. A VaR forecasting exercise is also conducted here to compare the relative performance of the proposed model to alternative popular VaR models including the RiskMetrics approach. Cryptocurrency risk modeling in general is an important aspect not only for modeling cryptocurrencies, but also for practical reasons due to financial institutions increased risk exposure to these new financial assets (Hotz-Behofsits et al., 2018; Catania et al., 2018; Hencic and Gouriéroux, 2015).

An efficient Bayesian estimation procedure that models Gegenbauer long memory, stochastic volatility, leverage and a bivariate Student's t-distribution is proposed here to model cryptocurrencies. A practical example involving a large investable sample of cryptocurrencies is also discussed. 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.

The remainder of this article is organized as follows: the model is introduced in Section 2, and its estimation procedure is detailed in Section 3. The results are discussed in Sections 4 and 5, and concluded with Section 6.

2. The model

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

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

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

$$\begin{pmatrix} \varepsilon_t^* \\ \eta_{t+1}^* \end{pmatrix} \sim t_\nu \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} e^{h_t} & \sigma\rho e^{h_t/2} \\ \sigma\rho e^{h_t/2} & \sigma^2 \end{pmatrix} \right), \tag{3}$$

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

The process y_t is assumed to be 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 et al. (2016) for details. The class of time series generated by (1)–(3) is similar to the GARFIMA(p, q)-SV time series process of Phillip et al. (2018) but includes additional important features.

A leverage effect is further assumed between the errors of the observation Eq. (1) and the latent Eq. (2) such that $\mathbb{E}[\varepsilon_t^* \eta_{t+1}^*] = \rho$. Further, they are assumed to follow a bivariate t-distribution in Eq. (3). Clearly, h_t is the log-volatility, which evolves according to the state Eq. (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. It is assumed $|\beta| < 1$ so h_t is stationary.

For simplicity, the generalized fractional stochastic volatility noise process is discussed when $\phi(B) = \psi(B) = 1$ such that $(1 - 2uB + B^2)^d y_t = \varepsilon_t^*$. This proposed (0,0) order model is called the GARFIMA-SV leverage heavy common tail (GMA-SV-LVG-HC) model. Under the assumption that y_t is causal, the following MA(∞) representation

$$y_t = (1 - 2uB + B^2)^{-d} \varepsilon_t^* = \sum_{j=0}^{\infty} \lambda_j \varepsilon_{t-j}^*, \tag{4}$$

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

$$\lambda_j = 2u\left(\frac{d-1}{j} + 1\right)\lambda_{j-1} - \left(\frac{2(d-1)}{j} + 1\right)\lambda_{j-2}, \quad j \geq 2. \tag{5}$$

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 (4) arises from truncating at lag J so that

$$y_t = (1 - 2uB + B^2)^{-d} \varepsilon_t^* \approx \sum_{j=0}^{J_t^*} \lambda_j \varepsilon_{t-j}^*, \tag{6}$$

where $J_t^* = \min(t, J)$. For further discussion on the choice of J, see Phillip et al. (2018). The power spectrum of (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 $\omega = \cos^{-1}(u)$ is the Gegenbauer frequency.

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

3. Bayesian inference

In order to estimate models (1)–(3), the fractional noise case is considered in order not to detract from the main features of the model, noting that alternative mean structures such as ARMA(p,q) can easily be implemented. Two important modifications in order to operationalize the model are presented, and they are order invariant.

Firstly, the Student-t distribution presented as Scale Mixtures of Normals (SMN) is introduced (Andrews and Mallows, 1974). 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 positive real support $\Re^+$, then the pdf of $\boldsymbol{X}$ is said to have 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

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

or hierarchically

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

A SMN is used to redefine the bivariate t-error distribution, so that (3) can be rewritten as:

$$\begin{pmatrix} \varepsilon_t^* \\ \eta_{t+1}^* \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \xi_{t+1}\begin{pmatrix} e^{h_t} & \sigma\rho e^{h_t/2} \\ \sigma\rho e^{h_t/2} & \sigma^2 \end{pmatrix}\right), \tag{7}$$

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

Secondly, the uncorrelated marginal distributions are derived. 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 (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 Eq. (6) can be expressed as

$$y_t|h_{t+1}, h_t = \sum_{j=0}^{J_t^*}\lambda_j e^{h_{t-j}/2}\frac{\rho}{\sigma}(h_{t+1-j} - \alpha - \beta(h_{t-j} - \alpha)) + \sum_{j=0}^{J_t^*}\lambda_j e^{h_{t-j}/2}\xi_{t+1-j}^{\frac{1}{2}}\sqrt{1-\rho^2}\varepsilon_{t-j}, \tag{8}$$

$$h_{t+1} = \alpha + \beta(h_t - \alpha) + \xi_{t+1}^{\frac{1}{2}}\sigma\eta_{t+1}, \tag{9}$$

where ε_t and η_{t+1} are the uncorrelated $N(0, 1)$ 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 $\mathbf{V} = \mathrm{diag}(\boldsymbol{W} \circ \boldsymbol{W} \circ \boldsymbol{\xi})$, $\boldsymbol{\xi} = (\xi_2, \ldots, \xi_{T+1})$ 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

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

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

$$\log f(\boldsymbol{Y}|\boldsymbol{h}, \boldsymbol{G}_J) = -\frac{T}{2}\log(2\pi(1-\rho^2)) - \frac{1}{2}\sum_{t=1}^{T}(h_t + \log \xi_{t+1}) - \frac{1}{2}(\boldsymbol{Y}-\boldsymbol{\mu})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}) \tag{10}$$

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 B.

4. Simulation studies

A comprehensive simulation study is now outlined in order to assess the performance of the 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 $\boldsymbol{\theta}$.

4.1. Parameter estimation

Data is generated according to (1)–(3) and the parameters are estimated subsequently. A value of $J = 1000$ is used in order to simulate a time series as close as possible to an infinite moving average representation. The parameter values $\nu = [5, \dots, 15, 20]$ and $\rho = [-0.6, -0.3]$ are also used. 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

$$(1 - 1.6B + B^2)^{0.4} y_t = \varepsilon_t^*, \tag{11}$$

$$h_{t+1} = 0.97 h_t + \eta_{t+1}^*, \tag{12}$$

$$\begin{pmatrix} \varepsilon_t^* \\ \eta_{t+1}^* \end{pmatrix} \sim t_\nu \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} e^{h_t} & 0.025\rho e^{h_t/2} \\ 0.025\rho e^{h_t/2} & 0.025^2 \end{pmatrix} \right). \tag{13}$$

The process is simulated to be $T = 1500$ to ensure the Gegenbauer parameters are estimated accurately (see Phillip et al., 2018 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 it is assumed $T = 1500$. The initial starting values are chosen arbitrarily and are purposely chosen to be far away from the true values. Several fixed starting values are also tested to ensure that different MCMC chains converge within similar value ranges.

The hyperparameters are set as follows:

$$u \sim N(0, 0.1)[-1, 1], \; d \sim N(0.125, 0.05)[0, 0.25]\mathbb{1}_{ud} + N(0.25, 0.05)[0, 0.5](1 - \mathbb{1}_{ud}), \; \nu \sim U[3, 23]$$

$$\rho \sim N(-0.1, 0.05), \; \alpha \sim N(0, 0.05), \; \beta \sim N(0.99, 0.2).$$

The process in Eqs. (11)–(13) is simulated $\Omega = 1,000$ times, and the parameters are estimated each time using the model. The estimated mean of each parameter, the root mean squared error (RMSE) and the mean of the standard errors in parentheses are reported in Tables 1 and 2. The number of iterates used is $M = 10,000$ after a burn-in period of 10,000. The burn-in period is purposely chosen to be of half the total number of iterations under the advice of Gelman et al. (2013).

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

Table 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	5.203	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)						

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

Table 2
Simulation study results when true value of $\rho = -0.6$.

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)						

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

Table 3
Table comparing the parameter estimates of the modeling procedures of Wang et al. (2011) and Choy and Chan (2000) to the 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).

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

As shown below in Tables 1 and 2, the estimates of $\hat{\nu}$ are more accurate with lower RMSE when the true value of ν is low. Conversely, 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 h). 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.

4.2. Methodology comparison

In Table 3, the procedure is compared with the methodologies of Wang et al. (2011) and Choy and Chan (2000). The authors propose model (1–3) without long memory effects, which is of course a special case of the proposed model when $d = 0$. The authors in Wang et al. (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, which is also adopted in this paper. Their priors and parameters are replicated as much as possible so that all three methodologies are directly comparable. First, Wang et al. (2011) assigned a vague prior to α, so the Normal prior used is $N(0, 10)$. The prior of $\beta^* = \frac{\beta+1}{2}$ is assumed to be Beta(20, 1.5) so the prior used is $N(0.95, 0.05)$ 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 the proposed methodology. A uniform prior $U(-1, 1)$ is assigned to ρ, so the truncated Normal prior of ρ is set 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, which are also adopted. The authors use a burn-in period of 50,000 loops and follow up with an additional 250,000 MCMC loops. A much smaller number of 10,000 loops, follow up with an additional 10,000 loops is used and it is found that all parameters have converged. The same true parameter values: $\alpha = -10$, $\beta = 0.8$, $\sigma^2 = 0.2$, $\rho = 0.8$ and $\nu = 5$ are used in both papers and the results are reported below in Table 3.

In essence, it is clear from Table 3 that the proposed method dominates for all parameters. In all cases, the standard deviations of parameter estimates for the proposed model, 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 the proposed model. The estimation of the shape parameter ν is noticeably different and the

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

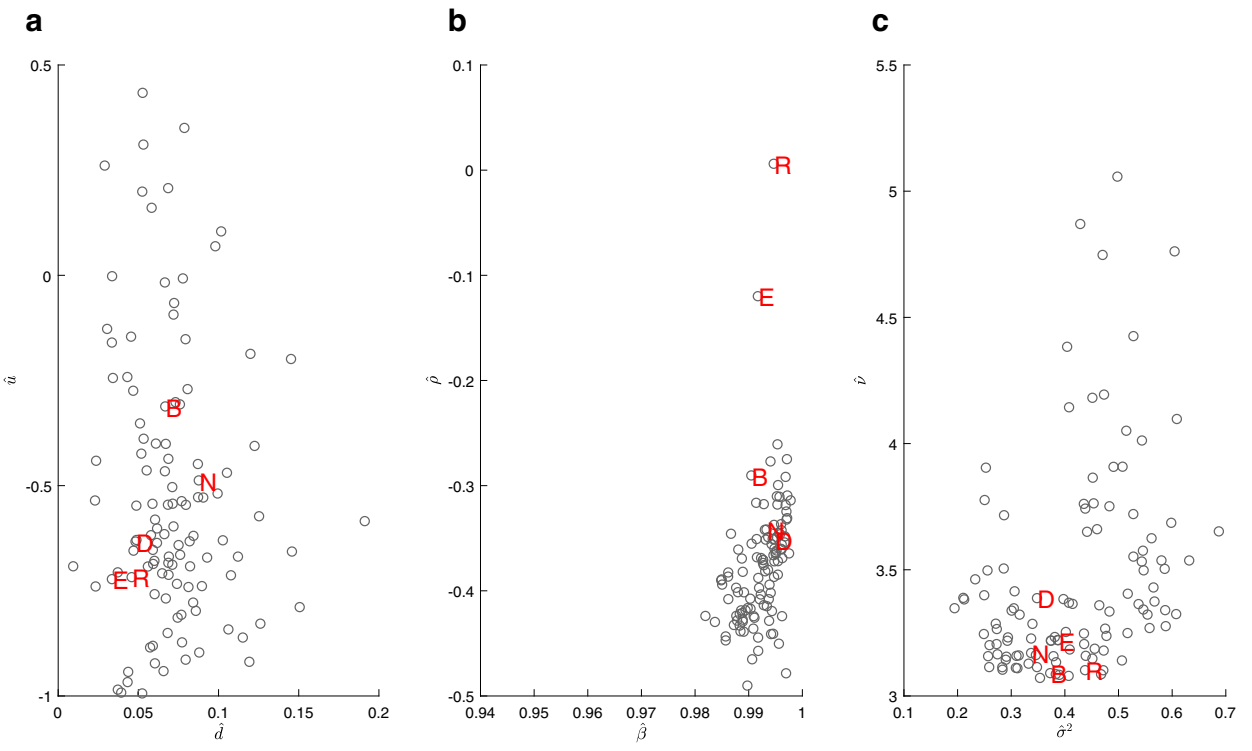

Fig. 1. Scatter plots of parameter estimates: (a) $[\hat{u}, \hat{d}]$; (b) $[\hat{\rho}, \hat{\beta}]$; (c) $[\hat{\nu}, \hat{\sigma}^2]$, of 114 different cryptocurrency data sets under the GLM-SV-LVG-HC model. B: Bitcoin, E: Ethereum, R: Ripple, N: NEM, D: Dash.

proposed model estimates are closer to the true values. Most notably, the authors were unable to correctly estimate ρ, but the proposed model is able to estimate ρ to be very close to the true parameter value.

5. Empirical data analysis

In this section a set of empirical data is considered to illustrate the proposed models. A general broad scope on an investable basket of 114 cryptocurrencies is presented by discussing model parameter estimates. A focus is then provided on five of the most popular cryptocurrencies, namely Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), NEM, and Dash, with model comparison, and followed 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 2796 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 2796 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, 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.

5.1. In-sample fitting

5.1.1. All 114 Cryptocurrencies

The proposed model in Eqs. (1)–(3) are first fitted 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 Fig. 1.

It is clear from Fig. 1(a) that cryptocurrencies in general show signs of generalized 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.

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

Table 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. Note: they all end on the 30th of April, 2018 but the start date varies, as per the number of observations listed.

	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 (< .0001)	255.8224 (< .0001)	3426.8910 (< .0001)
Ethereum	2	67.0341	995	0.0043	0.0715	−0.2145	7.3622	−0.3959	0.3293	281.7172 (< .0001)	141.7939 (< .0001)	796.5499 (< .0001)
Ripple	3	26.0533	1489	0.0000	0.0781	−1.3992	31.8162	−0.8844	0.6393	559.8652 (< .0001)	103.5672 (< .0001)	52003.6272 (< .0001)
NEM	12	3.0601	1117	0.0026	0.0877	0.4502	9.3332	−0.4656	0.6443	291.9821 (< .0001)	136.7968 (< .0001)	1904.4886 (< .0001)
Dash	14	2.6920	1483	0.0018	0.0727	0.0821	10.6803	−0.4881	0.5232	851.7943 (< .0001)	591.9198 (< .0001)	3646.5302 (< .0001)

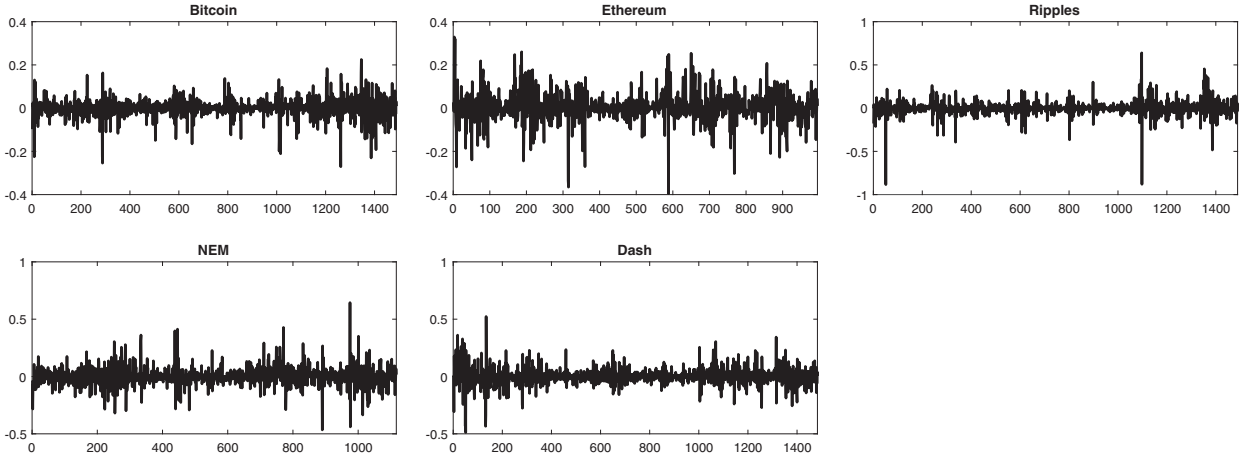

Fig. 2. Plots of daily returns for the top five cryptocurrencies by market capitalization.

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 the 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 et al., 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 Fig. 1(c). Interestingly, two main regimes are clear which depict the variability characteristics of cryptocurrencies. It is found 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.

5.1.2. Five of the most popular cryptocurrencies

In this subsection, the focus is on five of the most popular cryptocurrencies as per Table 4, which are plotted in Fig. 2.

The descriptive statistics of each time-series transformed by taking the daily price percentage change are shown in Table 4. Similar to their fiat counterparts, 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.

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 h to be sent from one user to another. Two particular cryptocurrencies which attempt to overcome this issue are ETH and Dash.

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

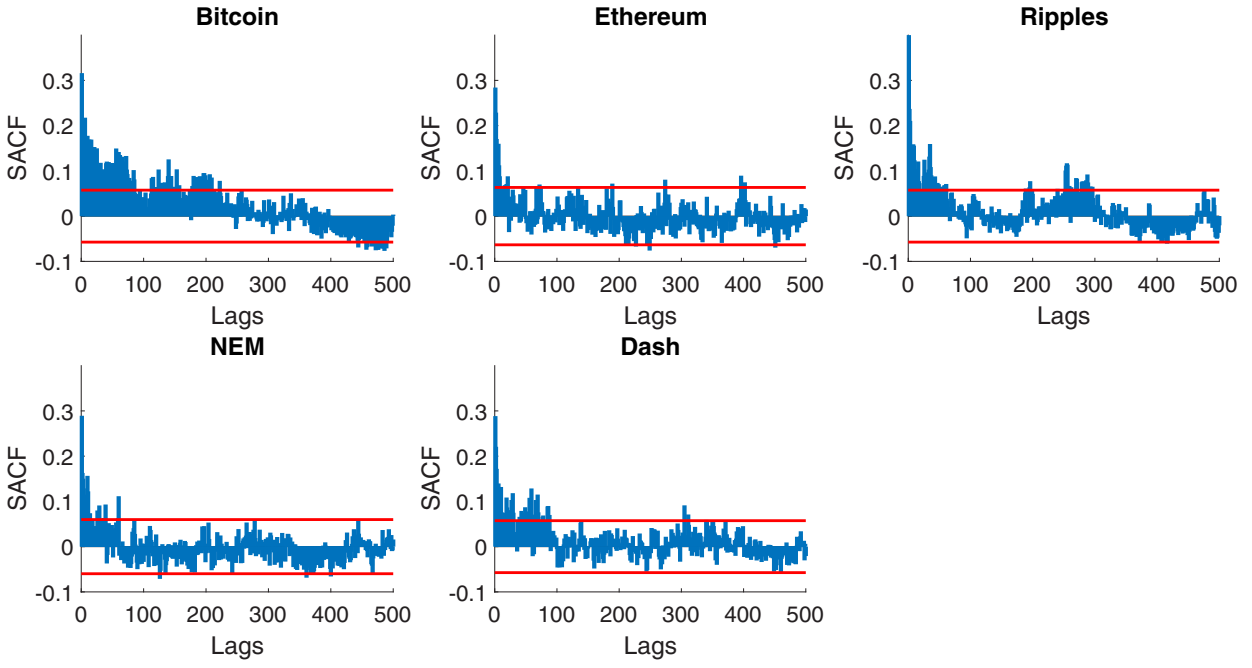

Fig. 3. ACFs of absolute returns of the top five cryptocurrencies by market capitalization as of the 31st of July, 2017.

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.

Fig. 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, Bitcoin displays the most recurrent behavior as evidenced by its strong cyclicality.

The autocorrelations of the squared returns as depicted in Fig. 4 also displays interesting behaviors of cryptocurrencies. It is shown that BTC, ETH and NEM display properties of dependent errors as typically seen in the literature. Notably, 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 Fig. 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, it has higher dependence amongst its squared returns.

In order to further inspect the properties of the cryptocurrencies listed in Table 4, each of the five most popular cryptocurrency data sets are fitted according to model (1)–(3) called the GMA-SV-LVG-HC, and five further nested sub-models including the Stochastic Volatility (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). It is particularly noted the sub-models GMA-SV-LVG and GMA-SV-HC models are nested variations of the proposed 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: $\text{DIC} = -2\log p(y|\hat{\theta}) + 2p_E$, 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.

For each model, the DIC is calculated and reported relative to the full GMA-SV-LVG-HC model in Table 5. Since all of the DICs are negative, the model with larger ratio is better. This table can be interpreted in conjunction with Fig. 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.

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

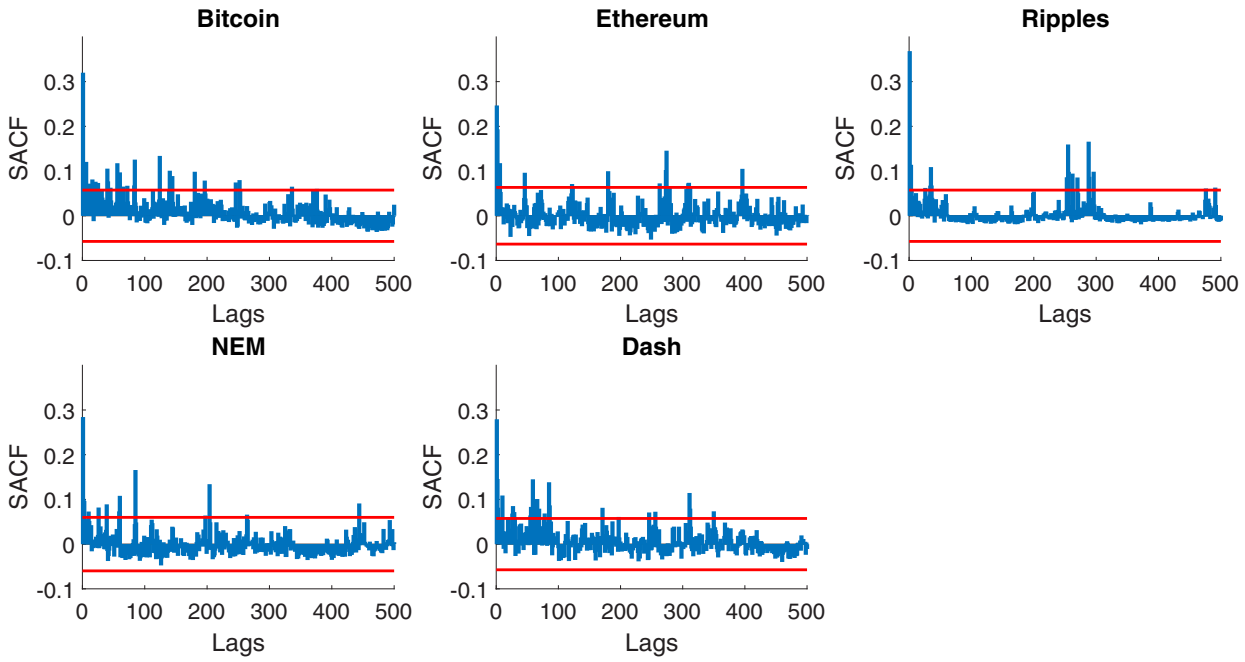

Fig. 4. ACFs of squared returns of the top five cryptocurrencies by market capitalization as of the 31st of July, 2017.

Table 5
Ratio of DICs. Since they are all negative values, larger is better.

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**

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. Fig. 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 Fig. 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.

These assertions are confirm to be consistent with many subtle facts of cryptocurrencies demonstrated by the proposed models. 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.

5.1.3. Bitcoin

This section is concluded by reporting in details parameter estimates of the models for Bitcoin in Table 6 and providing an in-depth discussion of the results.

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

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

Table 6
Analysis of BTC data.

Model	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	

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.

5.2. Out-of-sample forecast with Bitcoin

In this section, a forecasting exercise using Bitcoin is conducted for demonstrative purposes (forecasts for other four cryptocurrencies are available from supplementary material). Bitcoin is only reported 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. First the predictive density is defined to be the forecasted density at some future time point t denoted by $p(y_t|\mathcal{F}_{t-1}, \boldsymbol{\theta})$. Thus, for each MCMC iterate of $\boldsymbol{\theta}^i$, it is straightforward to 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

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

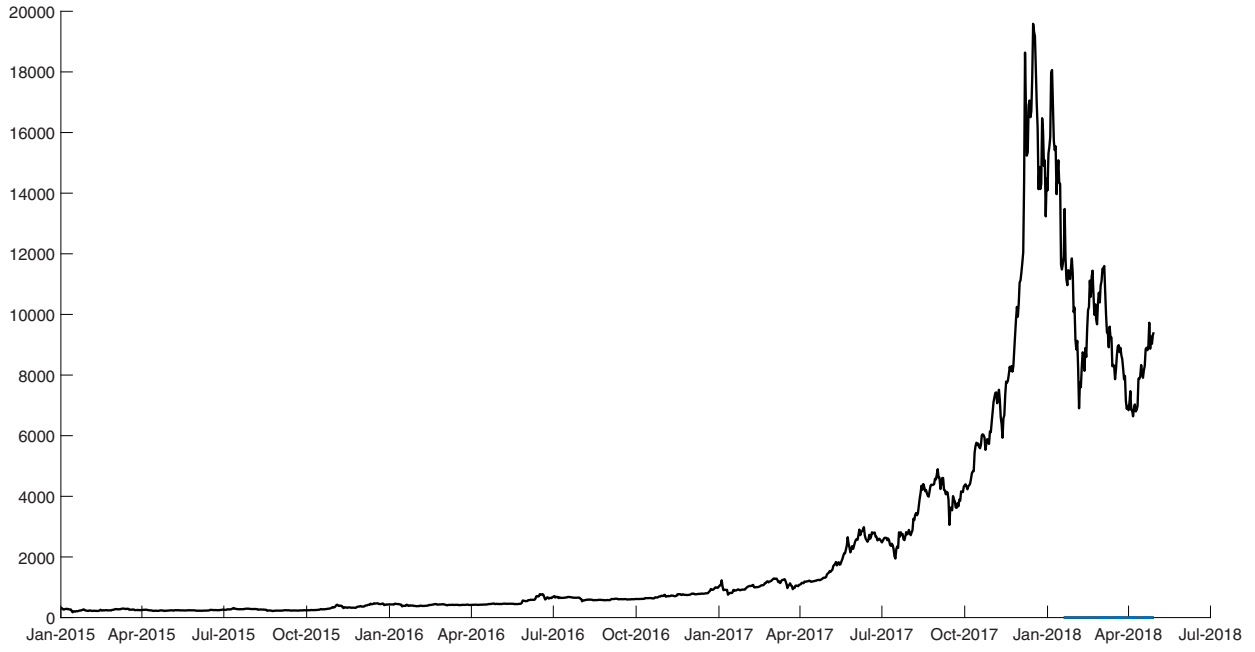

Fig. 5. Plot of Bitcoin price data from the 1st of January, 2015 to the 30th of April, 2018.

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 $> \alpha$). The most popular regulator's loss functions which are surveyed in Abad et al. (2015) are 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 1489 data points and the training dataset is 95% of the available data which contains 1415 data points and these are used to slide one day ahead to forecast VaR for the remaining 74 days (Fig. 5).

The first two rows of Table 7 report the average DIC and LL across the forecast horizon relative to those of SV-GMA-LVG-HC model. It is clear from these two measures, the inclusion of Gegenbauer long memory, stochastic volatility and heavy common tail are favored. All violation rates and all deviations are the smallest in general for the two smallest percentiles under the SV-GMA-LVG-HC specification.

Fig. 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.

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

Table 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.

		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**

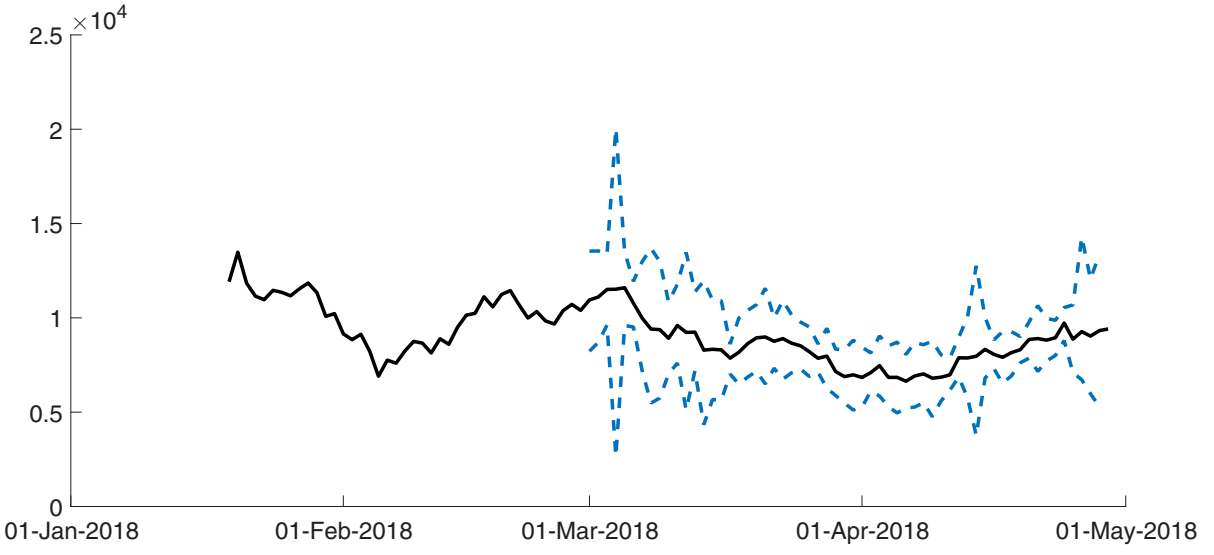

Fig. 6. Black line: Bitcoin price. Blue dotted lines: 1% and 99% one-step ahead VaR forecasts using the GMA-SV-LVG-HC model.

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 the aim of this work is 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 behaviors.

In the previously discussed work, a realistic investable basket of cryptocurrencies is explained and detailed. 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 these results, and are now able to provide a statistical handle on some of these matters. The most cited and known cryptocurrency, Bitcoin, is a prime

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

example of this. It is shown that its behaviors 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. It is proven, for the first time, the differences in risks amongst the most popular and debated cryptocurrencies. One important example that is measured, and scientifically proved, 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.

Appendix A. Model definitions

Model definitions. Let $y_t, t = 1, 2, \ldots, T$ be a stochastic process satisfying some model equations.

1. Stochastic volatility (SV):

$$\begin{aligned} y_t &= \varepsilon_t, \\ h_t &= \alpha + \beta(h_{t-1} - \alpha) + \eta_t, \end{aligned}$$

where $\varepsilon_t \sim \mathrm{N}(0, e^{h_t})$, $\eta_t \sim \mathrm{N}(0, \sigma^2)$ and $\mathbb{E}[\varepsilon_t \eta_t] = 0$.

2. Stochastic volatility model with leverage (SV-LVG):

$$\begin{aligned} y_t &= \varepsilon_t, \\ h_{t+1} &= \alpha + \beta(h_t - \alpha) + \eta_{t+1}, \\ \begin{pmatrix} \varepsilon_t \\ \eta_{t+1} \end{pmatrix} &\sim \boldsymbol{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} e^{h_t} & \sigma\rho e^{h_t/2} \\ \sigma\rho e^{h_t/2} & \sigma^2 \end{pmatrix} \right). \end{aligned}$$

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 $\varepsilon_t \sim \mathrm{N}(0, e^{h_t})$, $\eta_t \sim \mathrm{N}(0, \sigma^2)$ and $\mathbb{E}[\varepsilon_t \eta_t] = 0$.

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 $\mathbb{E}[\varepsilon_t \eta_t] = 0$.

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 B. Bayesian analysis

Sampling for u and d:

In order to estimate u and d, two independent truncated normal priors are considered with support in the region where generalized long-memory holds such that $p(u) \sim \mathrm{N}(\mu_u, \sigma_u^2)\mathbb{1}_{ud}$ and $p(d) \sim \mathrm{N}(\mu_d, \sigma_d^2)\mathbb{1}_{ud}$ 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 Gegenbauer long-memory stationarity is imposed through the prior distributions of u and d.

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

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, an approximation based on posterior modes from Gelman et al. (2013) is used, coupled with a proposal distribution precision tuning algorithm which is conducted only within the burn-in period. 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. Maximize the log posterior of u and d to find the modes $\tilde{u}$ and $\tilde{d}$ respectively. The log posterior modes are found by maximizing

$$\log p_u(u|d, \boldsymbol{h}) = \log f(\boldsymbol{Y}|d, \boldsymbol{h}) + \log \mathrm{N}(\mu_u, \sigma_u^2)\mathbb{1}_{ud},$$
$$\log p_d(d|u, \boldsymbol{h}) = \log f(\boldsymbol{Y}|u, \boldsymbol{h}) + \log \mathrm{N}(\mu_{d_1}, \sigma_d^2)\mathbb{1}_{ud} + \log \mathrm{N}(\mu_{d_2}, \sigma_d^2)(1 - \mathbb{1}_{ud}).$$

   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 $(\{-1 < u^* < 1, 0 < d < 0.5\} \cup \{|u^*| = 1, 0 < d < 0.25\})$. Practically, when $u^* \geq 0.99$, then $u^* = 1$ is set in order to give the event $|u| = 1$ non-zero probabilities. 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 and 3 by replacing d with d^*, u^* with u and c_u with c_d.

If u^* and d^* are accepted, then $u^{(m+1)} = u^*$ and $d^{(m+1)} = d^*$ are updated respectively. Next $\boldsymbol{G}_J$ is updated using the new Gegenbauer parameters.

Sampling for $\boldsymbol{h}$:

The estimation procedure of the latent variable $\boldsymbol{h} = [h_1, \ldots, h_T]$ which has its roots in Chan and Grant (2016) and Chan and Strachan (2012) is now discussed. Let $\boldsymbol{Y}^* = \boldsymbol{G}_J^{-1}\boldsymbol{Y}$ where $\boldsymbol{Y}^* = [y_1^*, \ldots, y_T^*]$. First, the density of $h_{t+1}|h_t$

$$h_{t+1}|h_t \sim \begin{cases} \mathrm{N}\left(\alpha, \frac{\sigma^2\xi_1}{1-\beta^2}\right), & t = 0 \\ \mathrm{N}(\alpha + \beta(h_t - \alpha), \sigma^2\xi_{t+1}), & t \neq 0 \end{cases}$$

is considered and it can be written in matrix notation as $\boldsymbol{H}_\phi \boldsymbol{h} = \tilde{\boldsymbol{\alpha}} + \boldsymbol{\omega}$ where

$$\boldsymbol{\omega} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_h), \quad \boldsymbol{\Sigma}_h = \mathrm{diag}\left(\frac{\sigma^2\xi_1}{1-\beta^2}, \sigma^2\xi_2, \ldots, \sigma^2\xi_{T+1}\right), \quad \tilde{\boldsymbol{\alpha}} = (\alpha, \alpha(1-\beta), \ldots, \alpha(1-\beta))',$$

$$\boldsymbol{H}_\phi = \begin{pmatrix} 1 & 0 & \ldots & \ldots & \ldots \\ -\beta & 1 & \ldots & \ldots & \ldots \\ 0 & -\beta & 1 & \ldots & \ldots \\ \vdots & 0 & -\beta & \ldots & \ldots \\ 0 & \vdots & \vdots & \ldots & \ldots \end{pmatrix}.$$

Therefore, the density of $\boldsymbol{h}$ is $\boldsymbol{h}|\xi_1, \boldsymbol{\xi}, \alpha, \beta, \sigma^2 \sim \mathbf{N}(\boldsymbol{H_\phi}^{-1}\tilde{\boldsymbol{\alpha}}, (\boldsymbol{H}_\phi'\boldsymbol{\Sigma}_{\boldsymbol{h}}^{-1}\boldsymbol{H}_\phi)^{-1})$. However, it is clear 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 $\log f(\boldsymbol{Y^*}|\boldsymbol{h}, \boldsymbol{\xi}, \alpha, \beta, \sigma^2, \rho)$, Taylor series expansion is applied around the neighborhood $\tilde{\boldsymbol{h}}$ of $\boldsymbol{h} \in \Re^T$, such that

$$\begin{aligned} \log f(\boldsymbol{Y^*}|\boldsymbol{h}, \boldsymbol{\xi}, \alpha, \beta, \sigma^2, \rho) &\approx \log f(\boldsymbol{Y^*}|\tilde{\boldsymbol{h}}, \boldsymbol{\xi}, \alpha, \beta, \sigma^2, \rho) + (\boldsymbol{h} - \tilde{\boldsymbol{h}})'\boldsymbol{f} - \frac{1}{2}(\boldsymbol{h} - \tilde{\boldsymbol{h}})'\boldsymbol{G}(\boldsymbol{h} - \tilde{\boldsymbol{h}}) \\ &\approx \boldsymbol{h}'\boldsymbol{f} - \frac{1}{2}\left[\boldsymbol{h}'\boldsymbol{G}\boldsymbol{h} - \boldsymbol{h}'\boldsymbol{G}\tilde{\boldsymbol{h}} - \tilde{\boldsymbol{h}}\boldsymbol{G}\boldsymbol{h}\right] \\ &\approx -\frac{1}{2}\left[\boldsymbol{h}'\boldsymbol{G}\boldsymbol{h} - 2\boldsymbol{h}'(\boldsymbol{f} + \boldsymbol{G}\tilde{\boldsymbol{h}})\right] \end{aligned}$$

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

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 $t = 2, \dots, T+1$,

$$f(y_t^*|h_{t+1}, h_t, \xi_{t+1}, \alpha, \beta, \sigma^2, \rho) = f(y_t^*) = \frac{1}{\sqrt{2\pi\xi_{t+1}(1-\rho^2)e^{h_t/2}}} \exp\left\{ -\frac{\left\{y_t^* - \frac{\rho}{\sigma}e^{h_t/2}[h_{t+1} - \alpha - \beta(h_t - \alpha)]\right\}^2}{2(1-\rho^2)\xi_{t+1}e^{h_t}} \right\},$$

$$f_1 = \frac{\partial \log f(y_1^*)}{\partial h_1}, \quad f_t = \frac{\partial}{\partial h_t}(\log f(y_t^*) + \log f(y_{t-1}^*)),$$

$$G_{11} = -\frac{\partial^2 \log f(y_1^*)}{\partial h_1^2}, \quad G_{tt} = -\frac{\partial^2}{\partial h_t^2}(\log f(y_t^*) + \log f(y_{t-1}^*)), \quad G_{t-1,t} = -\frac{\partial^2 \log f(y_t^*)}{\partial h_t \partial h_{t+1}}$$

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 $\tilde{\boldsymbol{h}}$:
   The mode $\tilde{\boldsymbol{h}}$ is first found with a Newton–Raphson scheme. Noting that in higher dimensions, the Newton–Raphson method can be generalized 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

   $$\begin{aligned}
   -\frac{1}{2}\frac{d}{d\boldsymbol{h}}(\boldsymbol{h}'\boldsymbol{K}_h\boldsymbol{h} - 2\boldsymbol{h}'\boldsymbol{k}_h) &= -\boldsymbol{h}'\boldsymbol{K}_h + \boldsymbol{k}_h \\
   -\frac{1}{2}\frac{d^2}{d\boldsymbol{h}^2}(\boldsymbol{h}'\boldsymbol{K}_h\boldsymbol{h} - 2\boldsymbol{h}'\boldsymbol{k}_h) &= -\boldsymbol{K}_h
   \end{aligned}$$

   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. A good trade-off between computational time and accuracy is found with $\max\{|\tilde{\boldsymbol{h}}^{n+1} - \tilde{\boldsymbol{h}}^n|\} < 10^{-4}$.

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

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 $p(\boldsymbol{h}|\boldsymbol{Y}^*, \boldsymbol{\theta}) \propto f(\boldsymbol{Y}^*|\boldsymbol{h}, \boldsymbol{\theta})f(\boldsymbol{h}|\boldsymbol{\theta})$. A draw $\boldsymbol{h}^*$ is first made and accepted with probability α_{AR} as

$$\alpha_{AR} = \min\left[\frac{p(\boldsymbol{h}^*|\boldsymbol{Y}^*, \boldsymbol{\theta})}{p_N(h^*|\tilde{\boldsymbol{h}}, \boldsymbol{K}_h^{-1})}, 1\right]$$

where $p_N(\cdot|\boldsymbol{m}, \boldsymbol{C})$ is a Gaussian proposal with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{C}$. Step 2 is repeated 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

$$\alpha_{MH} = \frac{p_N(\boldsymbol{h}|\tilde{\boldsymbol{h}}, \boldsymbol{K}_h^{-1})}{p(\boldsymbol{h}|\boldsymbol{Y}^*, \boldsymbol{\theta})}$$

(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:

The prior $\alpha \sim N(\mu_\alpha, \sigma_\alpha^2)$ is used. Note that a vague prior is most commonly used in the literature. It is assumed that the time series are percentage log returns so that $\mu_\alpha = 0$, and use $\sigma_\alpha^2 = \sqrt{10}$ similar to Kim et al. (1998). The posterior distribution of α is easily derived as

$$p(\alpha|h_1, \boldsymbol{h}, \xi_1, \boldsymbol{\xi}, \beta, \sigma^2, \rho, \boldsymbol{Y}^*) \propto f(h_1|\xi_1, \alpha, \beta, \sigma^2)\prod_{t=1}^{T} f(h_{t+1}|h_t, \xi_{t+1}, \alpha, \beta, \sigma^2)\prod_{t=1}^{T} f(y_t^*|h_{t+1}, h_t, \xi_{t+1}, \alpha, \beta, \sigma^2, \rho)\, p(\alpha).$$

Thus $p(\alpha|h_1, \boldsymbol{h}, \xi_1, \boldsymbol{\xi}, \beta, \sigma^2, \rho, \boldsymbol{Y}^*) \sim \mathrm{N}(V_\alpha M_\alpha, V_\alpha)$ where

$$M_\alpha = \frac{(1-\beta^2)h_1}{\sigma^2\xi_1} + \frac{1-\beta}{\sigma^2}\sum_{t=1}^{T}\frac{h_{t+1}-\beta h_t}{\xi_{t+1}} + \frac{(1-\beta)\rho}{(1-\rho^2)\sigma}\sum_{t=1}^{T}\frac{1}{\xi_{t+1}}\left[\frac{\rho}{\sigma}(h_{t+1}-\beta h_t) - \frac{y_t^*}{e^{h_t/2}}\right] + \frac{\mu_\alpha}{\sigma_\alpha^2}$$

$$V_\alpha = \left(\frac{1-\beta^2}{\sigma^2\xi_1} + \frac{(1-\beta)^2}{\sigma^2}\sum_{t=1}^{T}\frac{1}{\xi_{t+1}} + \frac{(1-\beta)^2\rho^2}{(1-\rho^2)\sigma^2}\sum_{t=1}^{T}\frac{1}{\xi_{t+1}} + \frac{1}{\sigma_\alpha^2}\right)^{-1}.$$

It is found 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. An adaptive Random-Walk Metropolis algorithm is instead favored, which has efficient results.

Sampling of β:

The unconditional likelihood of β is intractable due to the inclusion of the marginal likelihood of $f(h_1|\xi_1, \alpha, \beta, \sigma^2)$. 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

$$p(\beta|h_1, \boldsymbol{h}, \xi_1, \boldsymbol{\xi}, \alpha, \sigma^2, \rho, \boldsymbol{Y}^*) \propto \prod_{t=1}^{T} f(h_{t+1}|h_t, \xi_{t+1}, \alpha, \beta, \sigma^2)\prod_{t=1}^{T} f(y_t^*|h_{t+1}, h_t, \xi_{t+1}, \alpha, \beta, \sigma^2, \rho)\, p(\beta).$$

Hence, $p(\beta|h_1, \boldsymbol{h}, \xi_1, \boldsymbol{\xi}, \alpha, \sigma^2, \rho, \boldsymbol{Y}^*) \sim \mathrm{N}(V_\beta M_\beta, V_\beta)$ where:

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

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

The proposal β' is accepted with probability $\min\{\frac{q(\beta')}{q(\beta^{(i-1)})}, 1\}$, where $q(\cdot)$ is

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

$$q(x) = p(h_1|\xi_1, \alpha, x, \sigma^2) = \frac{(1-x^2)^{\frac{1}{2}}}{\sqrt{2\pi}\sqrt{\xi_1}\sigma} \exp\left\{-\frac{(h_1-\alpha)(1-x^2)}{2\xi_1\sigma^2}\right\} \propto (1-x^2)^{\frac{1}{2}} \exp\left\{-\frac{(h_1-\alpha)(1-x^2)}{2\xi_1\sigma^2}\right\}$$

The most typical scenario found in financial time series is β to be close to 1, and as such the prior is set to $\beta \sim N(0.99, 0.2)$.

Sampling of σ^2:

To sample σ^2, a modified version of the maximization at posterior (MAP) sampler is used, which is outlined below:

1. Derive the posterior log-density of σ^2 as

$$\begin{aligned}
&\log p(\sigma^2|h_1, \boldsymbol{h}, \xi_1, \boldsymbol{\xi}, \alpha, \beta, \rho, \boldsymbol{Y}^*) \\
&\approx \log f(h_1|\xi_1, \alpha, \beta, \sigma^2) + \sum_{t=1}^{T} \log f(h_{t+1}|h_t, \xi_{t+1}, \alpha, \beta, \sigma^2) + \sum_{t=1}^{T} \log f(y_t^*|h_{t+1}, h_t, \xi_{t+1}, \alpha, \beta, \sigma, \rho) + c_2 \\
&\approx -\frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2\xi_1} - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\frac{[h_{t+1}-\alpha-\beta(h_t-\alpha)]^2}{\xi_{t+1}} - (T+1)\log\sigma \\
&\quad -\sum_{t=1}^{T}\frac{\{\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}[h_{t+1}-\alpha-\beta(h_t-\alpha)]\}^2}{2(1-\rho^2)\xi_{t+1}}.
\end{aligned}$$

   Then, perform non-linear least squares optimization 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, a Gaussian prior is used such that $\rho \sim \mathrm{N}(\mu_\rho, \sigma_\rho^2)$. It is simple to see that

$$\begin{aligned}
&p(\rho|h_1, \boldsymbol{h}, \xi_1, \boldsymbol{\xi}, \alpha, \beta, \sigma^2, \boldsymbol{Y}^*) \\
&\propto f(h_1|\xi_1, \alpha, \beta, \sigma^2)\prod_{t=1}^{T} f(h_{t+1}|h_t, \xi_{t+1}, \alpha, \beta, \sigma^2)\prod_{t=1}^{T} f(y_t^*|h_{t+1}, h_t, \xi_{t+1}, \alpha, \beta, \sigma, \rho)\, f(\rho) \\
&= \frac{(1-\beta^2)^{-\frac{1}{2}}\xi_1^{-\frac{1}{2}}}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(h_1-\alpha)^2(1-\beta^2)}{2\sigma^2\xi_1}\right\} \times \frac{1}{(\sqrt{2\pi}\sigma)^T}\exp\left\{-\sum_{t=1}^{T}\frac{[h_{t+1}-\alpha-\beta(h_t-\alpha)]^2}{2\sigma^2}\right\} \\
&\quad \times \frac{1}{[2\pi(1-\rho^2)]^{T/2}\prod_{t=1}^{T}\xi_{t+1}}\exp\left\{-\sum_{t=1}^{T}\frac{\{\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}[h_{t+1}-\alpha-\beta(h_t-\alpha)]\}^2}{2(1-\rho^2)\xi_{t+1}}\right\} \times \frac{1}{\sqrt{2\pi}\sigma_\rho}\exp\left\{-\frac{(\rho-\mu_\rho)^2}{2\sigma_\rho^2}\right\} \\
&\propto \frac{1}{(1-\rho^2)^{T/2}}\exp\left\{-\sum_{t=1}^{T}\frac{\{\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}[h_{t+1}-\alpha-\beta(h_t-\alpha)]\}^2}{2(1-\rho^2)\xi_{t+1}} - \frac{\rho^2}{2\sigma_\rho^2} + \frac{2\mu_\rho\rho}{2\sigma_\rho^2}\right\}.
\end{aligned}$$

Then ρ is estimated using the MAP sampler, and accepted/rejected using the Metropolis Hastings sampling scheme. It is assumed that the Gaussian prior is $\rho \sim \mathrm{N}(-0.1, 0.05)$.

Sampling of $\boldsymbol{\xi}$:

To sample $(\xi_1, \boldsymbol{\xi})$ first note that each element is independent and $\xi_t \sim$ 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:

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

Similarly, when $t > 0$, then

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

$$\xi_{t+1}|h_{t+1}, h_t, \alpha, \beta, \sigma^2, \rho, \nu, y_t^* \sim \mathrm{IG}\left(\frac{\nu+1}{2}, \frac{[h_{t+1}-\alpha-\beta(h_t-\alpha)]^2}{2\sigma^2} + \frac{\{\frac{y_t^*}{e^{h_t/2}} - \frac{\rho}{\sigma}[h_{t+1}-\alpha-\beta(h_t-\alpha)]\}^2}{2(1-\rho^2)} + \frac{\nu}{2}\right)$$

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

Sampling of ν:

In order to sample ν, an adaptive independence-chain Metropolis–Hastings algorithm is implemented. Assuming the prior $\nu \sim \mathrm{U}[\nu^-, \nu^+]$, the density of the posterior distribution of ν is

$$p(\nu|\boldsymbol{\xi}) \propto f(\boldsymbol{\xi}|\nu)\, f(\nu)$$
$$\propto \prod_{t=1}^{T}\left[\frac{(\frac{\nu}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}\xi_t^{-\frac{\nu}{2}-1}\exp\left(-\frac{\nu}{2\xi_t}\right)\right] = \frac{(\frac{\nu}{2})^{\frac{T\nu}{2}}}{[\Gamma(\frac{\nu}{2})]^T}\left(\prod_{t=1}^{T}\xi_t^{-\frac{\nu}{2}-1}\right)\exp\left(-\sum_{t=1}^{T}\frac{\nu}{2\xi_t}\right) I(0<\nu<b_\nu)$$

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. The density is maximized in order to find the mode $\tilde{\nu}$ using routine optimization methods. Although an Inverse Gamma proposal is typically used, it is found this leads to boundary trap issues. Superior results are instead found 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 hyperparameters are $\nu^- = 3$ to avoid infinite 'explosive' variance issues, and $\nu^+ = 23$.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecosta.2018.10.003.

References

Abad, P., Muela, S.B., Martín, C.L., 2015. The role of the loss function in value-at-risk comparisons. J. Risk Model Valid. 9 (1), 1.
Ait-Sahalia, Y., Kimmel, R., 2007. Maximum likelihood estimation of stochastic volatility models. J. Financ. Econ. 83 (2), 413–452.
Andrews, D.F., Mallows, C.L., 1974. Scale mixtures of normal distributions. J. R. Stat. Soc. Series B Stat. Methodol. 36 (1), 99–102.
Berg, A., Meyer, R., Yu, J., 2004. Deviance information criterion for comparing stochastic volatility models. J. Bus. Econ. Stat. 22 (1), 107–120.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econom. 31 (3), 307–327.
Bos, C.S., Koopman, S.J., Ooms, M., 2014. Long memory with stochastic variance model: a recursive analysis for US inflation. Comput. Stat. Data Anal. 76 (0), 144–157.
Carnero, M.A., Peña, D., Ruiz, E., 2004. Persistence and kurtosis in GARCH and stochastic volatility models. J. Financ. Econ. 2 (2), 319–342.
Catania, L., Grassi, S., 2017. Modelling Crypto-Currencies Financial Time-Series. SSRN Working Paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3028486.
Catania, L., Grassi, S., Ravazzolo, F., 2018. Forecasting cryptocurrencies financial time series, https://brage.bibsys.no/xmlui/bitstream/handle/11250/2489408/WP_CAMP_5_2018.pdf.
Chan, J.C.C., Grant, A.L., 2016. Modeling energy price dynamics: GARCH versus stochastic volatility. Energy Econ. 54, 182–189.
Chan, J. C. C., Strachan, R. W., 2012. Estimation in Non-Linear Non-Gaussian State space Models with Precision-Based Methods, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3028486.
Chaum, D.L., 1981. Untraceable electronic mail, return addresses, and digital pseudonyms. Commun. ACM 24 (2), 84–90.
Cheah, E.-T., Fry, J., 2015. Speculative bubbles in bitcoin markets? an empirical investigation into the fundamental value of Bitcoin. Econ. Lett. 130 (Supplement C), 32–36.
Chib, S., Greenberg, E., 1994. Bayes inference in regression models with ARMA (p, q) errors. J. Econom. 64 (1-2), 183–206.
Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. Am. Stat. 49 (4), 327–335.
Chib, S., Nardari, F., Shephard, N., 2002. Markov chain monte carlo methods for stochastic volatility models. J. Econom. 108 (2), 281–316.
Choy, S.T.B., Chan, C.M., 2000. Bayesian estimation of stochastic volatility model via scale mixture distributions. In: Statistics and Finance. Imperial College Press and World Scientific Publishing Co., pp. 185–204.
Christie, A.A., 1982. The stochastic behavior of common stock variances: value, leverage and interest rate effects. J. Financ. Econ. 10 (4), 407–432.
Dissanayake, G., Peiris, M., Proietti, T., 2016. State space modeling of Gegenbauer processes with long memory. Computat. Stat. Data Anal. 100, 115–130.
Dissanayake, G., Peiris, M., Proietti, T., et al., 2018. Fractionally differenced gegenbauer processes with long memory: a review. Stat. Sci. 33 (3), 413–426.
Engle, R., 1995. ARCH: Selected Readings. Oxford University Press.
Engle, R.F., Ng, V.K., 1993. Measuring and testing the impact of news on volatility. J. Financ. 48 (5), 1749–1778.
Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis, 3rd CRC Press, Hoboken.
Granger, C., Joyeux, R., 1980. An introduction to long-memory time series models and fractional differencing. J. Time Ser. Anal. 1 (1), 15–29.
Gray, H.L., Zhang, N.-F., Woodward, W.A., 1989. On generalized fractional processes. J. Time Ser. Anal. 10 (3), 233–257.
Harvey, A., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Rev. Econ. Stud. 61 (2), 247–264.
Hencic, A., Gouriéroux, C., 2015. Noncausal autoregressive model in application to Bitcoin/USD exchange rates. In: Huynh, V.-N., Kreinovich, V., Sriboonchitta, S., Suriya, K. (Eds.), Econometrics of Risk. Springer International Publishing, Cham, pp. 17–40.
Hosking, J.R.M., 1981. Fractional differencing. Biometrika 68 (1), 165–176.
Hotz-Behofsits, C., Huber, F., Zörner, T. O., 2018. Predicting crypto-currencies using sparse non-gaussian state space models, https://onlinelibrary.wiley.com/doi/full/10.1002/for.2524.
Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. J. Bus. Econ. Stat. 12 (4), 371–389.

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Rev. Econ. Stud. 65 (3), 361–393.
Mandelbrot, B., Van Ness, J., 1968. Fractional Brownian motions, fractional noises and applications. SIAM Rev. 10 (4), 422–437.
Melino, A., Turnbull, S.M., 1990. Pricing foreign currency options with stochastic volatility. J. Econom. 45 (1), 239–265.
Meyer, R., Yu, J., 2000. BUGS for a Bayesian analysis of stochastic volatility models. Econom. J. 3 (2), 198–215.
Nakamoto, S., 2008. Bitcoin: a peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf.
Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59 (2), 347–370.
Omori, Y., Chib, S., Shephard, N., Nakajima, J., 2007. Stochastic volatility with leverage: fast and efficient likelihood inference. J. Econom. 140 (2), 425–449.
Omori, Y., Watanabe, T., 2008. Block sampler and posterior mode estimation for asymmetric stochastic volatility models. Comput. Stat. Data Anal. 52 (6), 2892–2910.
Phillip, A., Chan, J., Peiris, S., 2018. Bayesian estimation of Gegenbauer long memory processes with stochastic volatility: methods and applications. Studies in Nonlinear Dynamics and Econometrics, De Gruyter 22 (3), 1–29.
Rainville, E.D., 1960. The Macmillan Co. New York 1960.
Shephard, N., 2005. Stochastic Volatility: Selected Readings. Oxford University Press.
Shephard, N., Pitt, M.K., 1997. Likelihood analysis of non-gaussian measurement time series. Biometrika 84 (3), 653–667.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Series B Stat. Methodol. 64 (4), 583–639.
Taylor, S., 1986. Modelling Financial Time Series. Wiley, Chichester [West Sussex];New York;.
Urquhart, A., 2016. The inefficiency of Bitcoin. Econ. Lett. 148 (Supplement C), 80–82.
Urquhart, A., 2017. Price clustering in Bitcoin. Econ. Lett. 159 (Supplement C), 145–148.
Wang, J.J.J., Chan, J.S.K., Choy, S.T.B., 2011. Stochastic volatility models with leverage and heavy-tailed distributions: a Bayesian approach using scale mixtures. Comput. Stat. Data Anal. 55 (1), 852–862.
Yu, J., 2005. On leverage in a stochastic volatility model. J. Econom. 127 (2), 165–178.

Please cite this article as: A. Phillip, J. Chan and S. Peiris, On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2018.10.003