Cross, J. L., Hou, C., & Trinh, K. (2021). Returns, volatility and the cryptocurrency bubble of 2017–18. *Economic Modelling*, *104*, 105643. https://doi.org/10.1016/j.econmod.2021.105643

Returns, Volatility and the Cryptocurrency Bubble of 2017-18*

Jamie L. Cross† Chenghan Hou‡ Kelly Trinh§

September 8, 2021

Abstract:

Research on cryptocurrencies has focused on price and volatility formation in isolation, however knowledge about their interdependence is important for risk management and asset allocation. We investigate the existence and nature of such a relationship in four commonly traded cryptocurrencies: Bitcoin, Ethereum, Litecoin and Ripple, during the cryptocurrency bubble of 2017-18. Using a generalized asset pricing model, we find evidence of a risk premium effect in Litecoin and Ripple during the boom of 2017, and that adverse news effects were an important driver of the cryptocurrency crash of 2018 in all four cryptocurrencies. In an out-of-sample forecasting exercise, we find that allowing for stochastic volatility and a heavy tailed distribution provides more accurate return and volatility forecasts compared to a random walk benchmark. This suggests that cryptocurrency markets were not weak-form efficient during this period.

Keywords: Cryptocurrencies, Returns and volatility, Stochastic volatility, Time-varying parameter model, Forecasting

JEL-Classification: C52, C53, C58

*We thank the editor Professor Sushanta Mallick, as well as the associate editor and anonymous reviewers for all of their constructive feedback throughout the review process of the journal. We also thank Professor Francesco Ravazzolo for his very helpful comments in the development of this paper. This work is part of the research activities at the Centre for Applied Macroeconomics and Commodity Prices (CAMP) at the BI Norwegian Business School. Chenghan Hou would like to acknowledge financial support by the National Natural Science Foundation of China (72003064).

†BI Norwegian Business School.

‡Hunan University. Corresponding author: chenghan.hou@hotmail.com.

§James Cook University.

1 Introduction

Over the past decade, cryptocurrencies have emerged as a new asset class that has received much attention from investors, regulators, and academics. A large academic research has focused on price formation (Cheah and Fry, 2015; Urquhart, 2016; Nadarajah and Chu, 2017; Balcilar et al., 2017; Stavroyiannis and Babalos, 2019), and the drivers of volatility (Balcilar et al., 2017; Catania et al., 2018), however, little is known about the relationship between expected returns and volatility in cryptocurrencies. This is surprising because the link between expected returns and volatility in other asset classes, such as equities, is an important factor in both risk management and asset allocation: see, e.g. French et al. (1987) and Schwert (1989). On the one hand, increased price volatility may incur a risk premium, leading to a positive correlation between volatility and returns. On the other hand, adverse news effects that result in decreased returns may induce more uncertainty and higher volatility. Identifying such relationships remains an important step in correctly understanding returns and volatility dynamics in cryptocurrency markets.

In this paper, we contribute to this rapidly expanding literature by investigating whether the expected returns and volatility in four commonly traded cryptocurrencies: Bitcoin, Ethereum, Litecoin and Ripple, were time-varying and interdependent during the 2017-18 cryptocurrency bubble.[1] The major hurdle in conducting our analysis is that no suitable methodological framework to jointly model the relationship between possibly time-varying expected returns and volatility of assets currently exists. This is because standard asset pricing models that capture the link between expected returns and volatility, such as the commonly used stochastic volatility with leverage model (Omori et al., 2007), do not allow for the possibility of time-varying coefficients in the returns equation, which have been shown to be a key modelling feature when analyzing investor behaviour in cryptocurrency markets (Stavroyiannis and Babalos, 2019).[2] Moreover, testing the null hypothesis that the coefficients are constant within a time-varying parameter model results in a non-regular testing problem, for which conventional hypothesis testing procedures, e.g. Zhang et al. (2018), can not be applied.

A methodological contribution of this paper is that we show how to overcome both of these difficulties. In the first instance, we extend the state-of-the-art stochastic volatility with leverage and Student's-t distributed errors model developed in Jacquier et al. (2004), to a time-varying parameter framework, and provide an efficient Markov chain Monte Carlo

[1] From January 2017 to January 2019, the combined total market capitalization of these four currencies was between 70 and 90 percent making them representative of the entire market.

[2] A relationship between asset returns and volatility, is widely known as the *leverage effect* in finance (Black, 1976; Christie, 1982). For this reason, models that allow for correlation between returns and volatility are commonly referred to as leverage models. Since we are modeling cryptocurrencies, our intent is not to examine the presence of a leverage effect, however we are interested in generalizing this class of models to test for correlations between time varying returns and volatility within cryptocurrencies.

(MCMC) sampler for estimating this new model. Second, we build upon recent developments of hypothesis testing procedures within state-space models (Frühwirth-Schnatter and Wagner, 2010; Kastner and Frühwirth-Schnatter, 2014; Chan, 2018), to show how a non-centered parameterization of our model facilitates the computation of a Savage-Dickey density (SD) ratio—a common method for model comparison in Bayesian statistics. The SD ratio can then be used to test whether the expected returns and volatility of the cryptocurrencies in question are time-varying or constant, and for a correlation between the returns and volatility, within a unified framework.

Using the cryptocurrency bubble of 2017-18 as a testing ground for our analysis, the main results from our analysis can be summarized as follows. First, when looking at the entire period, we find that time-varying volatility is an important modeling feature of expected returns, however there is little evidence to support the existence of time-varying returns or a link between returns and volatility. Second, when splitting the period into two subperiods—i.e. the cryptocurrency boom in 2017 and the bust in 2018—we find a strong positive relationship between returns and volatility for Litecoin and Ripple during the boom period of 2017, however no such evidence is found in either Bitcoin or Ethereum during this period. This suggests that Litecoin and Ripple incurred a risk premium by investors during the boom of 2017. A possible explanation as to why Bitcoin and Ethereum did not incur such a premium is that they were the two largest cryptocurrencies in terms of market capitalization thereby making them more trustworthy in the eyes of investors. In contrast, the returns and volatility of all four cryptocurrencies exhibited a negative correlation during the crash of 2018, suggesting that adverse news effects were an important driver of cryptocurrency returns during this period. Moreover, this transition from a positive to negative relationship between returns and volatility caused these effects to either partially or fully cancel each other out when combining the two periods. Third, we find no evidence of herding behavior of investors during this period, thereby corroborating the recent results in Stavroyiannis and Babalos (2019).

Finally, in an out-of-sample forecast exercise, we find that our proposed model specification can outperform the random walk in both returns and volatility forecasts (both in terms of mean squared forecast errors and predictive likelihoods). Thus suggests that the cryptocurrencies under investigation were not weak-form efficient during this period. We also find that time-varying parameters and correlation between returns and volatility further improve upon the forecasts of volatility, however, the extent of these improvements is quite heterogeneous across the different cryptocurrencies.

The paper is structured as follows. We discuss related literature in Section 2, present the methodology in Section 3, results in Section 4 and conclude in Section 5. To enhance the readability of the paper, we defer all technical details to Appendix.

2 Literature review

A rapidly expanding literature on cryptocurrencies has emerged over the past decade. Early papers in this literature established that cryptocurrencies can not only be viewed as virtual currencies, but also as financial assets.[3] For instance, Woo et al. (2013) argue that the original cryptocurrency, Bitcoin, has money-like properties and consequently has a market value. This sentiment is corroborated by Glaser et al. (2014) who also find that most of the interest in Bitcoin is due to its potential returns as opposed to its use as a currency. More recently, Bouri et al. (2017a,b) have highlighted various characteristics of Bitcoin that suggest it should be viewed as an asset. In line with this evidence, we view cryptocurrencies as an asset throughout this paper, and propose a hypothesis testing framework to test for time-varying returns and volatility in an asset pricing model.

Having established that cryptocurrencies can be viewed as financial assets, scholars have began to investigate the price formation in cryptocurrency markets. An early study by Kristoufek (2013) finds evidence that the price of a Bitcoin is closely related to the number of search queries for Bitcoin on Google Trends and Wikipedia. The prevalence of such speculative actions is further supported by Cheah and Fry (2015) and Baur et al. (2018). A possible explanation for this behavior offered by Baur and Dimpfl (2018), is that uninformed investors suffer from a "fear of missing out" (FOMO) and the existence of pump and dump schemes. A related explanation put forth by Bouri et al. (2019), is the existence of herding behaviour, however Stavroyiannis and Babalos (2019) find that herding behaviour is no longer present when the more robust time-varying model is employed. This suggests that there is a need for the development of hypothesis testing procedure to determine whether or not cryptocurrency returns are in fact time-varying. A key contribution of our paper is the development of such a framework.

The first paper directly relating to cryptocurrency returns and volatility is Balcilar et al. (2017). They employ a non-parametric causality-in-quantiles test to analyse the causal relation between trading volume and Bitcoin returns and volatility. Their main conclusion is that volume cannot help predict the volatility of Bitcoin returns at any point of the conditional distribution. This does not imply that returns and volatility are unrelated. For instance, Baur and Dimpfl (2018) use the asymmetric threshold GARCH (TGARCH) model to examine the volatility effects and report that the positive shocks increase the volatility by more negative shocks. The usefulness of a GARCH methodology is questioned by Catania and Grassi (2017), however, who provide evidence that the more flexible model stochastic volatility framework is generally preferred to the GARCH

[3]This perspective is not without its critics. For instance Hanley (2013) and Yermack (2015) independently argue that Bitcoin has no intrinsic value. That being said, Yermack (2015) also concedes that the extreme volatility of Bitcoin price implies that it behaves more like a speculative investment than a currency.

model when modeling cryptocurrency returns. We build on Catania and Grassi (2017) by providing a framework that shows how to facilitate hypothesis testing of time-varying parameters within a newly developed asset pricing model with stochastic volatility.

Another strand of related literature has investigated the hedge, safe haven and diversification potential of cryptocurrencies against gold and crude oil (Charfeddine et al., 2020; Urom et al., 2020), stock markets (Jiang et al., 2021), and global economic policy uncertainty (Qin et al., 2021). Their empirical results suggest that cryptocurrencies should generally not be used as a hedge or safe haven due to their dynamic relationship with other factors. For example, Jiang et al. (2021) find that the dependence between cryptocurrencies and stock markets are significantly positive for some periods, and rarely negative. This leads them to conclude that cryptocurrencies do not provide a strong hedge or safe haven for stock markets.

Finally, two recent studies have aimed at forecasting cryptocurrency returns (Catania et al., 2019) and volatility (Catania et al., 2018). The present paper builds on both studies by providing a unified framework to forecast cryptocurrency returns and volatility.

3 Methodology

In this section, we present a novel methodology for modeling the relationship between returns and volatility within cryptocurrencies. We begin with the generalization of a popular asset pricing model with stochastic volatility to allow for time-varying parameters. We then show how adopting a non-centered parameterization of the model can overcome the difficulties of testing for time-variation in the returns and volatility equations with conventional hypothesis testing tools. This results in a novel framework to test for a connection between cryptocurrency returns and volatility.[4]

3.1 A time-varying parameter asset pricing model

A popular way to model the correlation between asset returns and volatility is the autoregressive stochastic volatility with leverage (AR-SVL) model, first proposed in Harvey and Shephard (1996).[5] The model has since been extended by Jacquier et al. (2004) to have Student's-t distributed errors in the returns equation (AR-SVL-t). In this section we build on this framework by proposing a time-varying parameter AR-SVL-t (TVP-AR-SVL-t) model.

[4]Note that while the methods developed in this paper were created to model the relationship between returns and volatility of cryptocurrencies, they can also be applied to asset prices more generally.

[5]An alternative approach to modeling volatility are the autoregressive conditional heteroscedasticity (ARCH) models (Engle, 1982)—or their generalized (GARCH) counterparts (Bollerslev, 1986). SV models have been shown to provide increased flexibility and superior model fit compared to GARCH models when modeling cryptocurrencies (Catania and Grassi, 2017). For this reason we adopt the SV approach to volatility modeling in this paper.

Let y_t denote the (excess) return of a cryptocurrency. The TVP-AR-SVL-t model is defined by

$$y_t = \mu_t + \alpha_t y_{t-1} + \varepsilon_t^y, \tag{3.1}$$

$$h_{t+1} = \mu_h + \phi_h(h_t - \mu_h) + \varepsilon_{t+1}^h, \tag{3.2}$$

$$\begin{pmatrix} \varepsilon_t^y \\ \varepsilon_{t+1}^h \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \eta_t e^{h_t} & \rho\eta_t^{1/2} e^{h_t/2}\omega_h \\ \rho\eta_t^{1/2} e^{h_t/2}\omega_h & \omega_h^2 \end{pmatrix}\right). \tag{3.3}$$

Equation (3.1) is the *returns equation*, in which μ_t captures the mean returns of y_t and α_t reflects the relationship between returns today y_t and yesterday y_{t-1}. The conditional expectation of this equation therefore captures expected returns. Equation (3.2) is the *volatility equation*, in which the log-volatility h_t is assumed to follow a stationary AR(1) process (3.2) with $|\phi_h| < 1$, and the initial condition is drawn from its stationary distribution $h_1 \sim \mathcal{N}(\mu_h, \omega_h^2/(1-\phi_h^2))$. Finally, equation (3.3) shows the joint distribution of the error terms. Following Jacquier et al. (2004), we utilize the scale mixture representation of the Student's-t distribution. Specifically, the scale-mixing variables $\eta_1, \ldots, \eta_T$ are assumed to be independent and follow the inverse-Gamma distribution $\mathcal{IG}(\nu/2, \nu/2)$ in which ν is the degree of freedom parameter (Geweke, 1993). To enhance the papers readability, we defer details of the prior distributions and our efficient MCMC sampler to Appendix A.

The correlation between the returns and volatility is modeled by the parameter ρ, which relates the disturbances from the return equation ε_t^y to those in the volatility equation ε_{t+1}^h. There are two opposing relations between volatility and returns. On the one hand, asset volatility may incur a risk premium, leading to a positive correlation between volatility and returns. On the other hand, the news effect (or leverage effect in the case of equities), acts in the opposite direction, leading to a negative returns increase volatility. This parameter is therefore central in addressing our research question.

There are two differences between the model specification in (3.1)-(3.3) and the conventional SVL-t model considered in Jacquier et al. (2004). First, Jacquier et al. (2004) define the relationship between returns and volatility to be a contemporaneous correlation between the return and volatility errors at date t, i.e. $\text{corr}(\varepsilon_t^y, \varepsilon_t^h) = \rho$. However here we adopt the more conventional definition in Harvey and Shephard (1996) of intertemporal dependence between the return and volatility errors, i.e. $\text{corr}(\varepsilon_t^y, \varepsilon_{t+1}^h) = \rho$.[6] This is motivated by Yu (2005) who shows that this specification is superior from both theoretical and empirical perspectives. From a theoretical perspective, the Harvey and

[6]It's worth noting that Harvey and Shephard (1996) use the notation ε_t^h for the error term in the volatility equation. Here we follow Yu (2005) and instead use ε_{t+1}^h to make clear that the correlation between returns and volatility relates to the intertemporal errors as opposed to the contemporaneous errors.

Shephard (1996) specification is appealing because it is the Euler approximation to the continuous time asymmetric SV model which is widely used in finance. As discussed in Yu (2005), one implication of this difference is that the Harvey and Shephard (1996) specification is a martingale difference sequence, implying that the model is a priori consistent with the efficient market hypothesis, whereas the Jacquier et al. (2004) specification is not. From an empirical perspective, Yu (2005) finds in a study of daily returns of the S&P500, that if the correlation between returns and volatility were estimated from the Jacquier et al. (2004) model, then it would be underestimated in magnitude by about 20%.

The second difference in model specification is our adoption of time-varying coefficients in the returns equation, i.e. (3.1). In particular, we assume that μ_t and α_t follow independent random walks of the form

$$\mu_t = \mu_{t-1} + \varepsilon_t^{\mu}, \qquad \varepsilon_t^{\mu} \sim \mathcal{N}(0, \omega_{\mu}^2), \tag{3.4}$$

$$\alpha_t = \alpha_{t-1} + \varepsilon_t^{\alpha}, \qquad \varepsilon_t^{\alpha} \sim \mathcal{N}(0, \omega_{\alpha}^2), \tag{3.5}$$

where the initial conditions are drawn from the distributions $\mu_1 \sim \mathcal{N}(\mu_0, V_\mu)$ and $\alpha_1 \sim \mathcal{N}(\alpha_0, V_\alpha)$, in which the hyperparameters are assumed to be known. This extension is motivated by the possibility that expected returns may have changed over the sample period. If they did, then failing to account for such time-variation would result in unreliable inference. Of course, if they did not, then specifying a time-varying parameter model will also result in unreliable inference. Before addressing our first research question, it is therefore essential that we start by determining whether the conditional mean of (3.1) is constant or time-varying. Such a test is discussed in the next section.

3.2 Hypothesis Testing: Bayes Factor and Savage-Dickey Density Ratio

In this section, we show how to formally test whether time-variation in the returns or volatility equations of the TVP-AR-SVL-t model exist in the data. For concreteness, suppose we wish to test if there is time-variation in the returns. This amounts to comparing the model in (3.1)-(3.5) to a restricted version in which the conditional mean of (3.1) is constant, i.e. (1) $\mu_t = \mu_0$ and (2) $\alpha_t = \alpha_0$ for $t = 1, \ldots, T$. For expository purposes, denote the unrestricted model as M_u and the restricted model as M_r. Then, given the data, one way to formally test the hypothesis that the unrestricted model is more favorable than the restricted one is with the Bayes factor (BF). Under the assumption of equal prior model probabilities, the BF of M_u in favor of M_r is defined as

$$\text{BF}_{ur} = \frac{p(y|M_u)}{p(y|M_r)}, \tag{3.6}$$

where $p(y|M_i), i \in \{u, r\}$, is the marginal (data) likelihood given the model M_i evaluated at the observed data $y = (y_1, y_2, \ldots, y_T)$. The Bayes factor quantifies the relative likelihood of the observed data under each model. For example, if $\text{BF}_{ur} = 2$, then the observed data are twice as likely to have been generated by the unrestricted model.

A desirable property of using the BF for hypothesis testing is that it has a probabilistic interpretation given equal prior model probabilities i.e. $p(M_u) = p(M_r)$.[7] Thus, if $\text{BF}_{ur} = 2$, then M_u is 2 times more likely than model M_r given the data. Despite this nice interpretation, however, the BF is not problem-free. First, it is well known that the BF is sensitive to the choice of prior distribution, and is not well-defined when an improper prior distribution is used.[8] This problem arises because the marginal likelihood is computed by integrating the likelihood function over the prior distribution. Thus, when the prior is not a proper probability density function, this integral is not well-defined. Second, the BF is subject to the well-known *Jeffrey-Lindley's paradox*. This paradox states that when a proper prior with large spread is used, the BF tends to favor the restricted model even when the sample size is large. In a recent attempt to overcome this problem, Li et al. (2014, 2015) have proposed a new Bayesian hypothesis testing framework in which the test statistics are well-defined under improper priors and are thereby immune to the Jeffrey-Lindley's paradox. However, these hypothesis testings are not without their own shortcomings. For instance, they rely on the asymptotic distributions of the test statistics and their finite sample properties are remain unknown. Furthermore, unlike the BF, these new proposed test statistics do not have a natural probabilistic interpretation. In this paper, we therefore conduct hypothesis testing based on the BF with proper priors used for all model parameters that are comparable to those used in previous studies such as Yu (2005).

While the Bayes factor has a natural interpretation, the major difficulty in obtaining the marginal likelihood of the TVP-AR-SVL-t model, is that it requires us to integrate out each of the remaining latent states: μ_t, α_t and h_t for all dates $t = 1, \ldots, T$, and this integral does not have an analytical solution. Fortunately, when models are nested, such as in our case, the Bayes factor can be computed using the Savage-Dickey density ratio (Verdinelli and Wasserman, 1995). Specifically, the Bayes factor of the unrestricted model in favor of the restricted one, can be obtained using the ratio (ignoring two complexities which are discussed below):

$$\text{BF}_{ur} = \frac{p(\Phi = 0)}{p(\Phi = 0|y)}, \tag{3.7}$$

where the numerator denotes the (joint) prior density of the set of parameters that we

[7] This is because the BF is equal to posterior odd ratio $\frac{p(M_u|y)}{p(M_r|y)}$ when $p(M_u) = p(M_r)$.

[8] An improper prior distribution is one that is not a proper probability distribution, i.e., in the case of continuous random variables, the prior density does not integrate to one.

wish to set to zero, and the denominator is the joint posterior density evaluated at zero. Intuitively, if $\text{BF}_{ur} < 1$ then the restricted model is more likely under the posterior density relative to the prior density, which is viewed as evidence in favor of the time-invariant model, and vice versa. Thus, the Savage-Dickey density ratio reduces the computational complexity of the Bayes factor from an analytically intractable integral, to evaluating two densities at a point. For example, to test whether time-varying coefficients are zero, we would set $\Phi = (\omega_\mu^2, \omega_\alpha^2)$. Alternatively, we could test whether there was a significant relationship between returns and volatility by simply setting $\Phi = \rho$.

Despite its simple theoretical foundations, there are two important practical complexities that arise in computing the Savage-Dickey density ratio when testing for time-varying returns or volatility. First, the null hypothesis of a zero variance value lies on the boundary of the parameter space resulting in an infeasible evaluation of the posterior distribution at that point. Second, the commonly employed inverse-gamma distribution for the prior of the variance terms has zero density at zero. Fortunately, we can overcome both of these difficulties by adopting a non-centered specification, which we detail in the next section.

3.3 Non-centered parameterization of the extended SVL model

In this section we show how the non-centered parameterization for state-space models proposed in Frühwirth-Schnatter and Wagner (2010) can applied to our TVP-AR-SVL-t model, thus allowing for evaluation of the Savage-Dickey density ratio.

Let $\widetilde{\mu}_t = \frac{(\mu_t - \mu_0)}{\omega_\mu}$, $\widetilde{\alpha}_t = \frac{(\alpha_t - \alpha_0)}{\omega_\alpha}$, $\widetilde{h}_t = \frac{(h_t - \mu_h)}{\omega_h}$, and, for notational convenience, we define $\theta_0 = (\mu_0, \alpha_0)'$, $\widetilde{\theta}_t = (\widetilde{\mu}_t, \widetilde{\alpha}_t)'$, $\omega_\theta = (\omega_\mu, \omega_\alpha)'$ and $x_t = (1, y_{t-1})$. Then, the non-centered parameterization of the TVP-AR-SVL-t model in (3.1)-(3.5) is given by

$$y_t = x_t\theta_0 + x_t \text{diag}(\widetilde{\theta}_t)\omega_\theta + \sqrt{\eta_t}\exp\left(\frac{1}{2}(\mu_h + \omega_h\tilde{h}_t)\right)\widetilde{\epsilon}_t^y, \tag{3.8}$$

$$\tilde{h}_{t+1} = \phi_h\tilde{h}_t + \widetilde{\epsilon}_t^h, \tag{3.9}$$

$$\widetilde{\theta}_t = \widetilde{\theta}_{t-1} + \widetilde{\epsilon}_t^\theta, \tag{3.10}$$

$$\widetilde{\epsilon}_t^\theta \sim \mathcal{N}(0, I_2), \quad \eta_t \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2}), \tag{3.11}$$

$$\begin{pmatrix} \widetilde{\epsilon}_t^y \\ \widetilde{\epsilon}_{t+1}^h \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \tag{3.12}$$

where $\text{diag}(\widetilde{\theta}_t)$ is a diagonal matrix with $\widetilde{\theta}_t$ on its diagonal, $\widetilde{\theta}_0 \sim \mathcal{N}(0, V_{\theta_0})$ and $\widetilde{h}_1 \sim \mathcal{N}(0, 1/(1-\phi_h^2))$.

The model specification is completed by specifying prior distributions for the parameters,

for which complete details are provided in Appendix.[9] Here we comment on the use of zero-mean, independent Gaussian priors, for the standard deviations of the latent states. This has two main advantages. First, it is conjugate and therefore facilitates fast and simple computation. Specifically, since the conditional distribution of the states are Gaussian, the posterior density can be directly estimated via Monte Carlo methods. Second, having both Gaussian priors and posteriors allows for straightforward computation of the Bayes factor using the Savage-Dickey density ratio discussed in the previous section.

4 Data and empirical findings

4.1 Data

We consider the four major cryptocurrencies: Bitcoin, Ethereum, Ripple and Litecoin from the 21st January 2017 - 20th January 2019. The rapid growth in total market capitalization of these cryptocurrencies is shown in Figure 4.1. In 2017, the market boomed, with total market capitalization increasing from \$16 billion to a peak of \$535 billion, an annual growth rate of just over 3200%. In January 2018, however, the market crashed, losing over \$400 billion of market capitalization in just one year. This cryptocurrency crash of 2018 has resulted in cryptocurrencies becoming notorious for their returns and volatility.

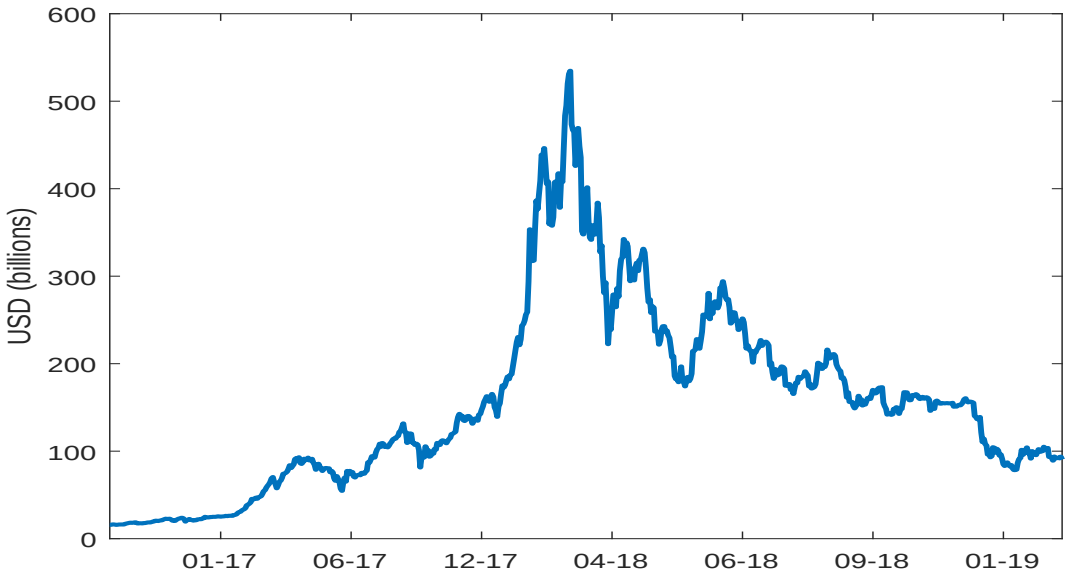

Figure 4.1: Total market capitalization (US Dollars) of four major cryptocurrencies: Bitcoin, Ethereum, Litecoin and Ripple, from January 2017 to January 2019. Source: coinmarketcap.com

Since the cryptocurrency market is open 24 hours a day, seven days a week, we convert

[9] Appendix A provides details for the proposed algorithm to estimate the model parameters, while Appendix C presents a Monte Carlo simulation to demonstrate the performance of our proposed algorithm.

the prices into daily returns using the closing price at midnight (UTC). This gives us a total of 730 observations for each series. All data was collected from CoinMarketCap.[10]

Figure 4.2 summarizes the returns series of the cryptocurrencies. We highlight three key points. First, the returns for each cryptocurrency do not seem to fluctuate around a constant mean, but instead revert to an evolving mean which seem to exhibit a sinosudinal pattern around zero. Second, each of the returns series exhibits large spikes, especially in the 2017 calendar year. Finally, the returns all appear to be heteroscedastic. These three observations suggest that our TVP-AR-SVL-t model may provide superior model fit to a conventional SVL model with Gaussian errors and time-invariant coefficients in the returns equation.

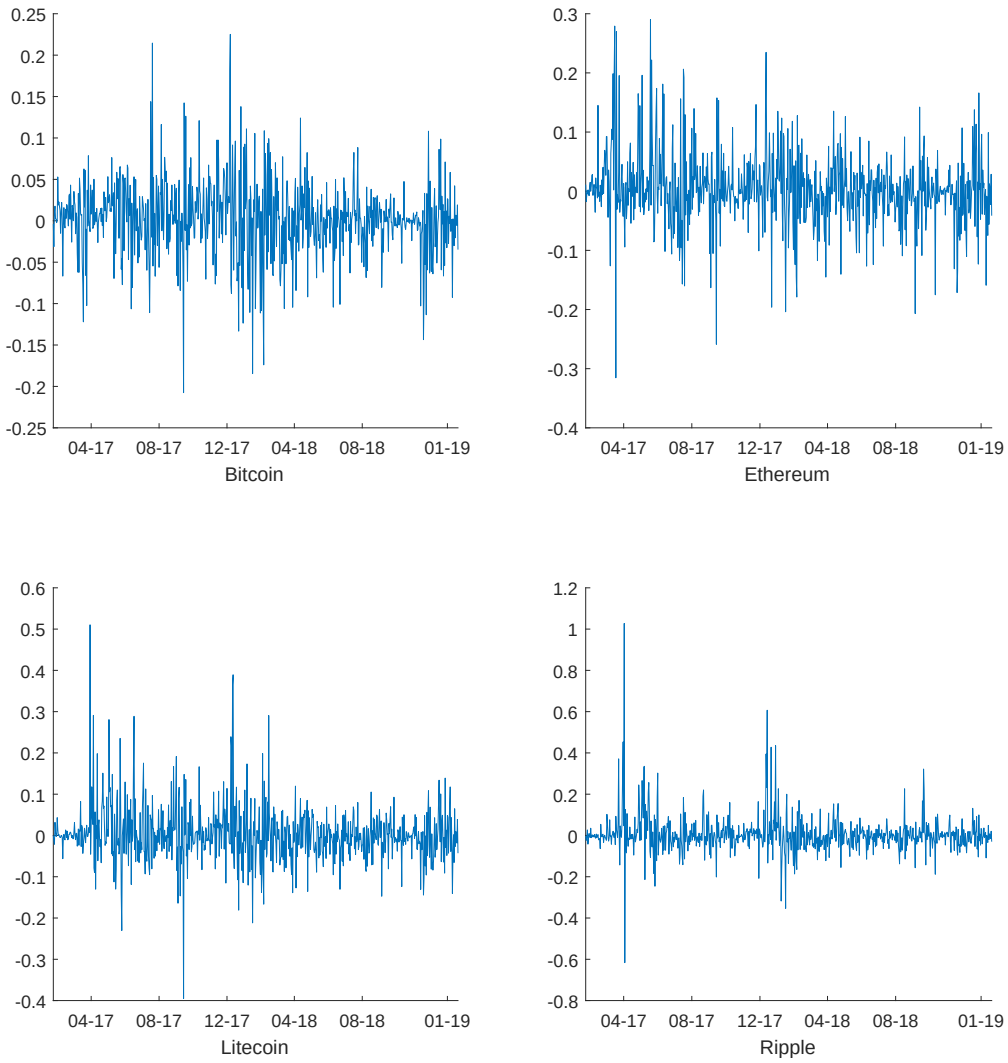

Figure 4.2: Daily returns of cryptocurrencies from 21st January 2017 - 20th January 2019.

[10]The data that support the findings of this study are openly available in https://coinmarketcap.com/.

4.2 Hypothesis Testing Results

Were expected (excess) returns and volatility in cryptocurrencies time-varying and interdependent during the cryptocurrency bubble of 2017-18? Was there evidence of news or risk premium effects or herding behaviour during this period? In this section, we address these questions using the hypothesis testing framework detailed in Section 3.2. The excess returns are here defined as the log-return of a cryptocurrency minus the risk-free return $r_{f,t}$, i.e. $y_t = log(P_t/P_{t-1}) - r_{f,t}$ where P_t is the closing price of the cryptocurrency at period t and $r_{f,t}$ is proxied by 1-month treasury constant maturity rate (Series ID: DGS1MO), provided by the Federal Reserve Bank of St. Louis Economic Database (FRED).[11]

As discussed in Section 3.2, examining time variation in the parameters μ_t, α_t and h_t is equivalent to testing whether or not the standard deviations associated with these time-varying states are different from zero. In other words, we wish to test the hypothesis

$$\begin{aligned} &\mathrm{H}_0 : \omega_r = 0, \quad \text{against}, \\ &\mathrm{H}_1 : \omega_r \neq 0, \end{aligned}$$

where ω_r, $r \in \{\mu, \alpha, h\}$, is the standard deviation of the error term in a given state equation. To test this hypothesis, we use the following variant of the Bayes factor in (3.7)

$$\mathrm{BF}_{ur} = \frac{p(\omega_r = 0)}{p(\omega_r = 0|y)},$$

where the numerator is the marginal prior density of ω_r evaluated at 0, which often can be evaluated analytically. The denominator is the marginal posterior of ω_r evaluated at 0, which in general has no closed form solution. However, given the posterior draws of the model parameters, the denominator can be easily computed using Monte Carlo methods. We defer details of this computation to Appendix B.

The estimated log BF for each restriction are in Table 4.1. For interpretation purposes, a value greater than zero implies the unrestricted model is preferred to the restricted model, and vice versa. Following Kass and Raftery (1995) values between zero and one are *not worth more than a bare mention*, one and three provides *positive* evidence in favor of the unrestricted model, three to five provides *strong* evidence and anything greater than five is *very strong*. We conduct the hypothesis-tests on data over the full period 2017-18 as well as data on 2017 (referred to as the "boom" period), 2018 (referred to as the "bust" period) for the robustness analysis.

[11] The use of excess returns is necessary for the interpretation of a risk premium effect in the returns-volatility relationship.

In the case of time-varying parameters, the results in the first two rows of each panel suggest positive to very strong evidence for the restricted model. One exception is the intercept for the full period, which is found to be time-varying. Moreover, when testing for time-varying volatility, the results in row three of each panel provide very strong evidence for the unrestricted volatility model in both sub-samples and across the entire sample. In other words, the results suggest that while stochastic volatility is an important modeling feature of cryptocurrency returns, the evidence for time-varying coefficients is less compelling.

Table 4.1: Estimated log Bayes factors and associated numerical standard errors.

Restriction	**Bitcoin**	**Ethereum**	**Litecoin**	**Ripple**
(A)	**Boom period: Jan/2017-Jan/2018**			
$\omega_\mu = 0$	-5.123	-4.174	-5.105	-4.831
	(0.003)	(0.009)	(0.012)	(0.009)
$\omega_\alpha = 0$	-2.1423	-2.139	-1.470	-2.267
	(0.002)	(0.004)	(0.012)	(0.004)
$\omega_h = 0$	13.40	9.126	6.268	17.037
	(1.817)	(0.776)	(0.149)	(2.220)
$\rho = 0$	-1.328	-1.161	2.191	1.133
	(0.006)	(0.011)	(0.060)	(0.041)
(B)	**Bust period: Jan/2018-Jan/2019**			
$\omega_\mu = 0$	-4.842	-4.422	-5.168	-4.923
	(0.007)	(0.016)	(0.006)	(0.012)
$\omega_\alpha = 0$	-2.256	-2.345	-2.487	-2.101
	(0.003)	(0.007)	(0.003)	(0.008)
$\omega_h = 0$	5.96	1.738	2.417	0.492
	(0.174)	(0.183)	(0.045)	(0.093)
$\rho = 0$	-0.236	-0.077	0.581	-0.070
	(0.017)	(0.031)	(0.042)	(0.030)
(C)	**Full period: Jan/2017-Jan/2019**			
$\omega_\mu = 0$	39.127	17.175	19.795	7.49
	(0.837)	(0.165)	(0.243)	(0.252)
$\omega_\alpha = 0$	-2.793	-3.049	-2.779	-2.775
	(0.002)	(0.0028)	(0.003)	(0.005)
$\omega_h = 0$	96.50	56.767	79.73	91.061
	(3.921)	(1.852)	(4.791)	(6.829)
$\rho = 0$	-1.398	-1.588	-1.073	-0.511
	(0.013)	(0.007)	(0.016)	(0.024)

Notes: The rows in each panel contain Bayes factors associated with the ratio of prior and posterior densities of the unrestricted model evaluated at the point given by the first column. The numbers in parenthesis are numerical standard errors that are obtained from estimating the Bayes factors on 10 parallel chains.

The next question is whether a relationship between returns and volatility of the cryptocurrencies exists. This relationship is captured by ρ. The results in row four of each panel in Table 4.1 are mixed. While the full period results provide weak or positive evidence of no relationship between returns and volatility across all cryptocurrencies, the sub-period results suggest a different story. Specifically, the Bayes Factors for each cryptocurrency during the boom of 2017 provide evidence of a news effect in Litecoin and Ripple but equally strong evidence is provided against such a risk premium effect in both Bitcoin and Ethereum. The strength of this evidence then dissipates during the crash of 2018. To further investigate these mixed results we plot the posterior distributions of ρ in Figure 4.3.

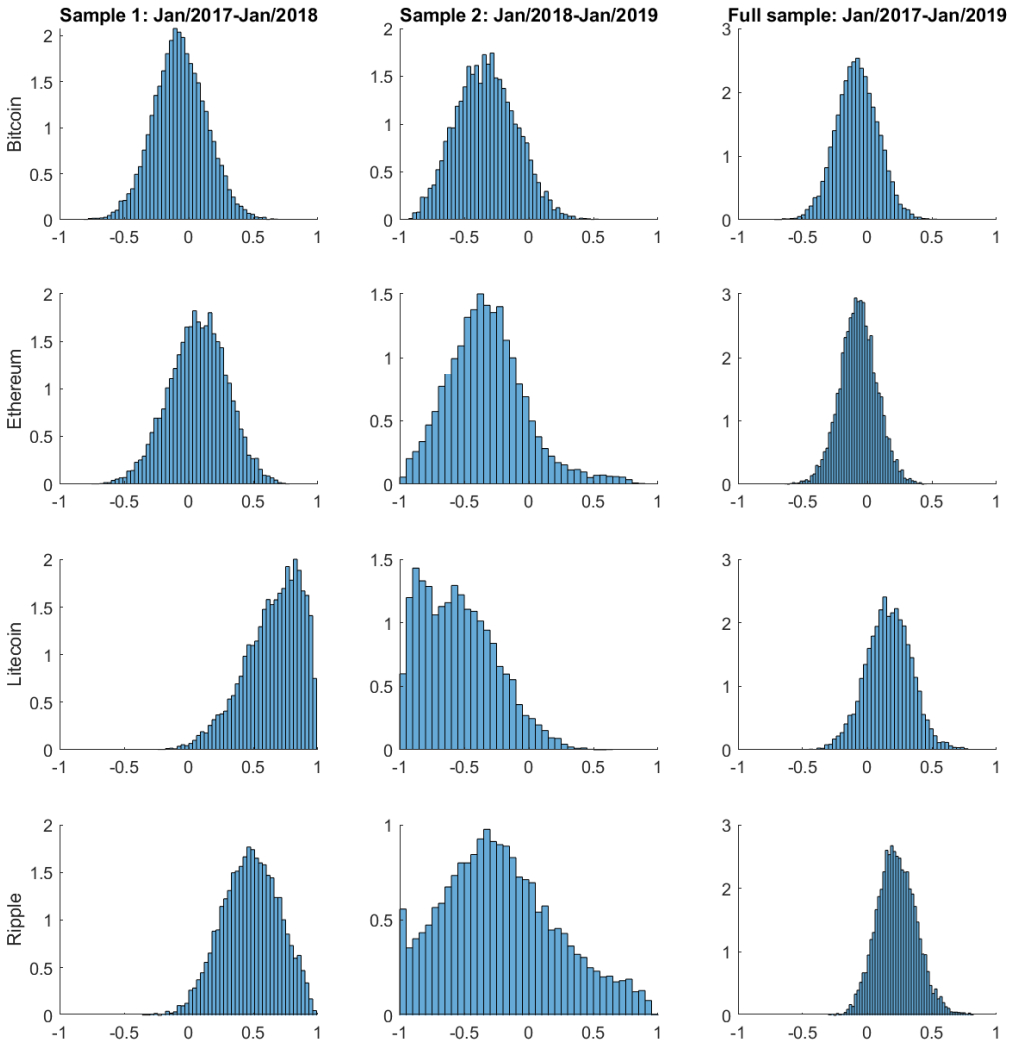

Figure 4.3: Posterior distribution for ρ.

During the boom period (column one), the posterior distributions of ρ for Litecoin and

Ripple suggest a strong positive correlation between returns and volatility, however no such evidence is found in either Bitcoin or Ethereum. In contrast, the returns and volatility of all four cryptocurrencies exhibited a strong negative correlation during the bust (column two). Interestingly, the full-period (column three) distribution for Bitcoin and Ethereum are fairly centered at zero, but skewed to the right for Litecoin and Ripple. These results suggests that the volatility of Litecoin and Ripple incurred a risk premium by investors during the boom of 2017, however no such phenomenon existed for Bitcoin or Ethereum. This could be due to the fact that Bitcoin and Ethereum are the two largest cryptocurrencies in terms of market capitalization thereby making them more trustworthy in the eyes of investors. In contrast, adverse news effects were an important driver of the cryptocurrency crash of 2018. Moreover, this transition from a positive to negative relationship between returns and volatility caused these effects to either partially or fully cancel each other out over the full period.

To test for herding behaviour, we consider the cross-sectional absolute deviation (CSAD) proposed by Chang et al. (2000) which is based on capital asset pricing model (CAPM) to study the returns' dispersion during the periods of extreme market movements

$$CSAD_{\mathfrak{i},t} = \gamma_0 + \gamma_1 |R_{m,t}| + \gamma_2 R_{m,t}^2 + \varepsilon_t \tag{4.1}$$

where $CSAD_{\mathfrak{i},t} = \frac{1}{N}\sum_{\mathfrak{i}=1}^{N} |R_{\mathfrak{i},t} - R_{m,t}|$ with $R_{\mathfrak{i},t}$ is the return of asset $\mathfrak{i}$ in period t and $R_{m,t}$ is the cross-sectional average return of the market portfolio in time t and N is the number of cryptocurrencies.

If investors tend to suppress their own analysis and follow a herding behaviour during the extreme periods of market movements, then the estimated γ_2 is expected to be negative and statistically significant. In this analysis, we extend the Chang et al. (2000) by allowing time-varying, volatility, and interdependencies between the returns and volatility in CSAD to examine the herding behaviour. The estimated γ_2 is presented in Figure 4.4, which indicates that no herding behaviour exists in the market, thereby corroborating the evidence presented in Stavroyiannis and Babalos (2019), who use a similar time-varying parameter model that does not account for correlation between the returns and volatility equations.

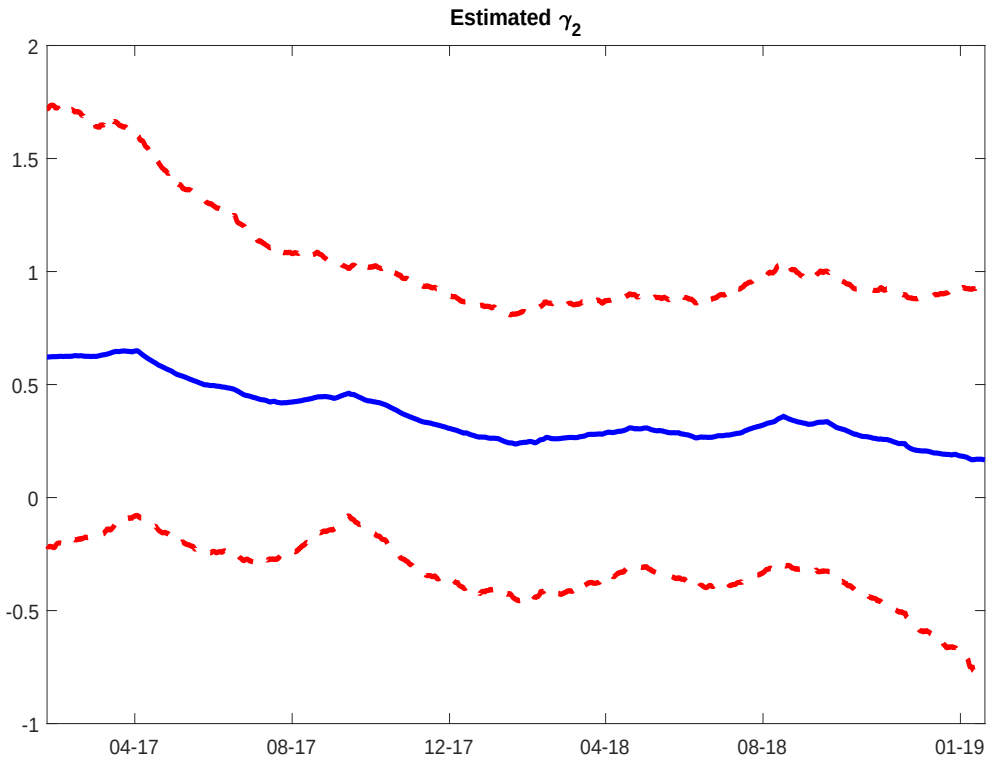

Figure 4.4: Time evolution of the estimated γ_2 of the CSAD model and its 95% posterior intervals.

4.3 Forecast Results

Is accounting for the relationship between returns and volatility documented in Section 4.2 useful for forecasting purposes? We investigate this question by comparing the out-of-sample forecast performance of the generalized TVP-AR-SVL-t model, to various nested versions, specifically: a random walk autoregressive model (RW), an autoregressive model with stochastic volatility (AR-SV), and an AR-SV with correlation between returns and volatility (AR-SVL), each with and without time-varying parameters in the returns equation and with either Gaussian of Student's-t distributed errors. The complete list of models is summarized in Table 4.2.

Table 4.2: A list of competing models.

Model	Description
RW	Random walk with homoscedastic Gaussian errors
AR	Constant parameter AR(1) model with homoscedastic Gaussian errors
AR-SV	AR with Gaussian SV errors
AR-SVL	AR with homoscedastic Gaussian errors and correlation between returns and volatility
AR-t	AR with Student's-t errors
AR-SV-t	AR with Student's-t SV errors
AR-SVL-t	AR with Student's-t SV errors and correlation between returns and volatility
TVP-AR	AR with time-varying parameters
TVP-AR-SV	AR-SV with time-varying parameters
TVP-AR-SVL	AR-SVL with time-varying parameters
TVP-AR-t	AR-t with time-varying parameters
TVP-AR-SV-t	AR-SV-t with time-varying parameters
TVP-AR-SVL-t	AR-SVL-t with time-varying parameters

At the outset, we note that our goal is not to propose the best possible forecasting model. Instead, the nested models in Table 4.2 are used to explore whether the inclusion of time-varying parameters, heavy-tailed errors, or the correlation between returns and volatility can help to improve forecast accuracy. Such knowledge can be used to determine whether these cryptocurrencies were weak-form efficient throughout the cryptocurrency crash of 2018, and may provide some insights to investors who are generally interested in forecasting both returns of portfolios, and volatility as a measurement of risk.[12]

The details of the forecasting study are summarized as follows. First, we divide the data into a training period that contains 50% of the data, i.e. 21st January 2017 - 20th January 2018, and the evaluation period that contains the remaining 50% of the data. Next, the training period is used to obtain initial parameter estimates for each model. These parameters are then used to predict returns and volatility of cryptocurrencies using forecast horizons from $s = 1$ to $s = 8$ days. This procedure is then repeated by recursively adding additional observations over the evaluation period using an expanding window approach.

A well known difficulty in evaluating the accuracy of volatility forecasts is that the true volatility process is not observed. To overcome this problem, we use the volatility proxy of Parkinson (1980), which is defined as

$$h_t^0 = \frac{(\log(H_t) - \log(L_t))^2}{4\log(2)}, \tag{4.2}$$

[12]It is also worth noting that our forecast exercise can also be used to address the issue of model selection. This is because the one-step-ahead predictive likelihood is the multiplicative updating factor applied to the marginal likelihood as new observations become available. As a result, the model with the best one-step-ahead predictive likelihood also provides the best in-sample fit. This relationship dates back to Geisel (1973) who used it to compare macroeconomic models. An excellent textbook discussion on the topic of model comparison is provided by Geweke (2005).

where H_t, L_t are the daily high and low prices respectively. Several studies have found that this estimator performs well with real data (see e.g. Alizadeh et al. (2002) and references therein).[13]

To assess the accuracy of the point and density forecasts, we use the mean squared forecast error (MSFE) and the average log-predictive-likelihood (ALPL). To this end, let z^o_{t+s} denote Parkinson's volatility proxy or the observed prices of cryptocurrencies, and $\hat{z}_{t+s}$ be the estimated posterior means of volatility or the predicted returns obtained from one of the above models. The MSFE is defined as

$$\text{MSFE} = \frac{1}{T-s-t_0+1}\sum_{t=t_0}^{T-s}(z^o_{t+s} - \hat{z}_{t+s})^2,$$

while the ALPL is defined as

$$\text{ALPL} = \frac{1}{T-s-t_0+1}\sum_{t=t_0}^{T-s}\log p_{t+s}(z_{t+s} = z^0_{t+s}|z_{1:t}).$$

The forecast results for the returns and volatility are respectively provided in Table 4.3, Table 4.4 and Table 4.5 for the forecast horizons $s = 1, 2, 4$ and 8 days. To facilitate our comparison of the results, in each table we first present the MSFE and ALPL of the benchmark model (AR-SV model for volatility, and AR model for returns), and then present relative MSFE and relative ALPL values for the remaining models to the benchmark. Since we place the benchmark in the denominator, a relative MSFE value smaller 1 indicates that the given model is more accurate than the benchmark model. The same conclusion holds for density forecasts when the relative ALPL is larger than 0. In order to statistically compare the differences in the point and density forecasts (i.e. MSFE and ALPL), we conduct pairwise tests of equal predictive accuracy by Diebold and Mariano (1995) relative to the random walk benchmark.

While no model systematically dominates across both returns and volatility forecasts, a few general observations can be made. When forecasting returns (see Tables 4.3-4.4), even a simple Gaussian AR model can outperform a random walk benchmark. Thus, in contrast to Nadarajah and Chu (2017), we find that the cryptocurrency market was not weak-form efficient during this period. Moreover, we find that allowing for time-varying coefficients, correlation between returns and volatility, or a heavy tailed error distribution generates further improvements upon the simple Gaussian AR model. More specifically, the TVP-AR-SV-t model tends to provide the most accurate predictions for Bitcoin and Litecoin, while the TVP-AR-SVL-t model does a good job when predicting Ethereum.

[13] A popular alternative to the Parkinson estimator is the index of squared returns. It is well known however, that while this measure results in an unbiased index, it is extremely noisy (Andersen and Bollerslev, 1998). We therefore use the Parkinson estimator.

A stronger result holds for volatility forecasts (see Table 4.5). In that case, specifying a heavy tailed error distribution systematically improves upon the forecast performance of a Gaussian error distribution. In contrast, solely allowing for correlation between returns and volatility or time-varying parameters to the benchmark model—i.e. the AR-SVL or TVP-AR-SV specifications—will generally worsen the forecast performance. From a practical perspective, the TVP-AR-SV-t model tends to be most accurate for Bitcoin and Ripple, while the AR-SVL-t model does a good job for Ethereum and Litecoin.

In summary, the general takeaway is that the cryptocurrency market was not weak-form efficient during this period. From a modeling perspective, allowing for stochastic volatility and a heavy tailed distribution provides more accurate return and volatility forecasts compared to a simple AR benchmark. Specifying time-varying parameters or correlation between returns and volatility may further improve upon the returns forecasts, however, the extent of these improvements is quite heterogeneous across the different series.

Table 4.3: Relative MSFE (panel a), ALPL (panel b) for forecasting the returns of Bitcoin, and Ethereum

	a) Relative MSFE				b) Relative ALPL			
Bitcoin								
Models	s=1	s=2	s=4	s=8	s=1	s=2	s=4	s=8
RW	0.003	0.003	0.003	0.003	0.970	0.672	0.342	0.004
AR	0.485	0.546	0.498	0.493	0.097	0.390	0.719	1.055
AR-SV	0.490	0.554	0.499	0.498	0.918	1.219	1.564	1.853
AR-SVL	0.490	0.553	0.499	0.497	0.920	1.221	1.562	1.854
AR-t	0.486	0.547	0.493	0.492	0.090	0.382	0.711	1.048
AR-SV-t	0.489	0.555	0.500	0.498	0.942	1.240	1.566	1.871
AR-SVL-t	0.490	0.556	0.500	0.499	0.939	1.237	1.563	1.864
TVP-AR	0.482	**0.541**	0.492	0.492	0.093	0.372	0.696	1.030
TVP-AR-SV	0.484	0.544	**0.491**	0.492	0.931	1.231	1.579	1.869
TVP-AR-SVL	0.483	0.542	0.491	**0.491**	0.932	1.234	1.579	1.869
TVP-AR-t	**0.479**	**0.541**	0.493	**0.491**	0.086	0.364	0.688	1.022
TVP-AR-SV-t	0.483	0.544	**0.491**	**0.491**	**0.957**	**1.255**	**1.581**	**1.885**
TVP-AR-SVL-t	0.483	0.543	**0.491**	0.648	0.942	1.237	1.561	1.863
Ethereum								
RW	0.006	0.006	0.006	0.006	0.820	0.552	0.229	-0.102
AR	0.495	0.568	0.492	0.492	0.151	0.417	0.738	1.068
AR-SV	0.481	0.554	0.477	0.477	0.694	0.968	1.288	1.604
AR-SVL	0.481	0.554	0.477	0.478	0.699	0.972	1.292	1.609
AR-t	0.495	0.568	0.489	0.491	0.145	0.411	0.732	1.061
AR-SV-t	0.475	0.550	0.474	0.475	**0.724**	0.992	1.303	1.616
AR-SVL-t	0.477	0.550	**0.473**	0.475	**0.724**	**0.994**	1.306	1.621
TVP-AR	0.484	0.551	0.480	0.481	0.149	0.407	0.725	1.053
TVP-AR-SV	0.479	0.546	0.475	0.478	0.695	0.972	1.293	1.604
TVP-AR-SVL	0.479	0.547	0.474	0.476	0.704	0.977	1.297	1.609
TVP-AR-t	0.483	0.548	0.479	0.480	0.142	0.400	0.718	1.046
TVP-AR-SV-t	**0.473**	0.547	0.474	**0.473**	0.722	0.991	1.303	1.612
TVP-AR-SVL-t	0.476	**0.545**	0.474	0.475	0.721	**0.994**	**1.307**	**1.622**

Note: Results for RW benchmark are actual values, and the results for the remaining models are reported relative to the benchmark. Values in bold indicates the best relative MSFE and ALPL. Gray cells indicate the significance difference of the predictive accuracy between an alternative model and the benchmark RW, at 10% level of significance using the test in Diebold and Mariano (1995).

Table 4.4: Relative MSFE (panel a), ALPL (panel b) for forecasting the returns of Litecoin, and Ripple

	a) Relative MSFE				b) Relative ALPL			
Litecoin								
Models	s=1	s=2	s=4	s=8	s=1	s=2	s=4	s=8
RW	0.006	0.006	0.006	0.006	0.794	0.513	0.190	-0.143
AR	0.483	0.532	0.493	0.501	0.160	0.439	0.761	1.093
AR-SV	0.465	0.515	0.475	0.485	0.755	1.033	1.362	1.686
AR-SVL	0.465	0.515	0.475	0.485	0.758	1.035	1.364	1.687
AR-t	0.482	0.531	0.489	0.500	0.158	0.437	0.759	1.091
AR-SV-t	**0.462**	**0.512**	**0.474**	0.483	**0.767**	**1.037**	1.357	1.678
AR-SVL-t	0.464	0.515	**0.474**	0.486	0.758	1.027	1.348	1.672
TVP-AR	0.476	0.523	0.483	0.493	0.157	0.427	0.746	1.076
TVP-AR-SV	0.467	0.513	0.475	0.485	0.754	1.035	**1.365**	**1.688**
TVP-AR-SVL	0.467	0.514	**0.474**	0.485	0.756	1.036	**1.365**	**1.688**
TVP-AR-t	0.475	0.520	0.481	0.490	0.155	0.424	0.743	1.073
TVP-AR-SV-t	0.467	**0.512**	**0.474**	**0.482**	0.764	**1.037**	1.358	1.678
TVP-AR-SVL-t	1.236	1.357	1.247	1.271	0.386	0.643	0.950	1.261
Ripple								
RW	0.007	0.006	0.007	0.007	0.642	0.348	0.020	-0.316
AR	0.530	0.577	0.518	0.516	0.203	0.496	0.822	1.157
AR-SV	0.507	0.552	**0.493**	0.493	0.848	1.117	1.435	1.774
AR-SVL	**0.506**	0.551	0.495	**0.492**	0.853	1.124	1.440	1.779
AR-t	0.522	0.571	0.510	0.507	0.232	0.525	0.851	1.186
AR-SV-t	0.514	**0.547**	0.494	0.496	**0.861**	1.134	1.452	1.781
AR-SVL-t	0.581	0.611	0.552	0.607	0.846	1.105	1.438	1.745
TVP-AR	0.525	0.563	0.511	0.511	0.201	0.479	0.799	1.131
TVP-AR-SV	0.516	0.553	0.498	0.497	0.841	1.114	1.434	1.771
TVP-AR-SVL	0.517	0.554	0.497	0.497	0.846	1.124	1.439	1.777
TVP-AR-t	0.520	0.556	0.504	0.500	0.227	0.508	0.830	1.163
TVP-AR-SV-t	0.530	**0.547**	0.504	0.500	0.853	1.129	1.446	1.772
TVP-AR-SVL-t	0.522	0.554	0.499	0.498	0.860	**1.136**	**1.453**	**1.787**

Note: Results for RW benchmark are actual values, and the results for the remaining models are reported relative to the benchmark. Values in bold indicates the best relative MSFE and ALPL. Gray cells indicate the significance difference of the predictive accuracy between an alternative model and the benchmark RW, at 10% level of significance using the test in Diebold and Mariano (1995).

Table 4.5: Relative MSFE (panel a), ALPL (panel b) for forecasting the volatilities of Bitcoin, Ethereum, Litecoin, and Ripple

	a) Relative MSFE				b) Relative ALPL			
Bitcoin								
Models	s=1	s=2	s=4	s=8	s=1	s=2	s=4	s=8
AR-SV	1.945	2.065	2.152	2.456	-4.146	-4.104	-3.610	-3.390
AR-SVL	1.047	1.044	1.046	1.048	0.227	0.291	0.087	0.029
AR-SV-t	0.925	0.923	0.945	0.957	1.985	1.896	1.353	1.067
AR-SVL-t	0.927	0.921	0.947	0.961	2.056	1.963	1.415	1.111
TVP-AR-SV	0.978	0.978	0.976	0.978	0.208	0.217	0.162	0.230
TVP-AR-SVL	1.008	1.007	1.009	1.010	0.441	0.428	0.272	0.179
TVP-AR-SV-t	**0.901**	**0.900**	**0.924**	**0.936**	**2.092**	**1.967**	**1.452**	**1.175**
TVP-AR-SVL-t	0.954	0.958	0.931	0.981	1.791	1.420	1.346	0.678
Ethereum								
AR-SV	1.832	1.971	2.112	2.367	-3.004	-2.989	-2.789	-2.577
AR-SVL	1.038	1.032	1.029	1.020	0.173	0.145	0.042	-0.091
AR-SV-t	**0.869**	0.869	0.881	0.882	**1.308**	**1.227**	**0.956**	0.672
AR-SVL-t	0.873	**0.865**	**0.866**	**0.852**	1.292	1.227	0.965	**0.681**
TVP-AR-SV	1.002	1.001	1.000	0.998	0.021	-0.005	0.001	-0.066
TVP-AR-SVL	1.037	1.027	1.025	1.014	0.191	0.122	-0.003	-0.160
TVP-AR-SV-t	0.874	0.870	0.880	0.880	1.295	1.220	0.955	0.668
TVP-AR-SVL-t	0.900	0.893	0.896	0.886	1.108	1.041	0.778	0.495
Litecoin								
AR-SV	1.409	1.536	1.610	1.709	-2.399	-2.440	-2.204	-1.993
AR-SVL	1.014	1.010	1.008	1.009	0.214	0.200	0.087	0.017
AR-SV-t	0.890	0.879	0.889	0.895	0.871	0.860	0.583	0.347
AR-SVL-t	0.895	**0.875**	**0.867**	**0.859**	0.864	0.864	**0.605**	**0.369**
TVP-AR-SV	0.998	0.996	0.993	0.993	0.021	0.046	-0.001	0.024
TVP-AR-SVL	1.017	1.012	1.010	1.012	0.196	0.194	0.095	0.016
TVP-AR-SV-t	**0.888**	0.876	0.883	0.891	**0.874**	**0.866**	0.598	0.356
TVP-AR-SVL-t	1.462	1.376	1.342	1.343	-1.80	-2.05	-1.73	-2.28
Ripple								
AR-SV	1.730	1.986	2.200	2.387	-2.307	-2.403	-2.279	-2.120
AR-SVL	1.013	1.004	1.005	0.996	0.136	0.113	0.053	-0.007
AR-SV-t	0.964	0.917	0.892	0.887	0.627	0.665	0.495	0.301
AR-SVL-t	0.971	0.914	0.872	**0.845**	0.616	0.664	0.504	**0.319**
TVP-AR-SV	1.000	1.002	1.003	1.000	0.013	0.022	-0.013	-0.002
TVP-AR-SVL	1.025	1.013	1.009	0.996	0.134	0.123	0.041	0.000
TVP-AR-SV-t	**0.949**	**0.911**	**0.881**	**0.881**	**0.640**	**0.676**	**0.513**	0.308
TVP-AR-SVL-t	0.956	0.928	0.907	0.891	0.509	0.511	0.346	0.192

Note: Results for AR-SV benchmark are actual values, and the results for the remaining models are reported relative to the benchmark. Values in bold indicates the best relative MSFE and ALPL. Gray cells indicate the significance difference of the predictive accuracy between an alternative model and the benchmark AR-SV, at 10% level of significance using the test in Diebold and Mariano (1995).

5 Conclusion

We have questioned whether time-variation and inter-dependencies between expected returns and volatility in four commonly traded cryptocurrencies: Bitcoin, Ethereum, Litecoin and Ripple, can explain their dynamics during the cryptocurrency bubble of 2017-18. To this end, we generalized a popular asset pricing model with correlation between the returns and volatility equations, to allow for time-varying parameters, and showed how to formally test for the significance of time-varying returns and volatility within this new model. An efficient algorithm for estimating this new model was also provided.

This suggests that the volatility of Litecoin and Ripple incurred a risk premium by investors during the boom of 2017, however no such relationship was found in either Bitcoin or Ethereum during this period. A possible explanation for this is that Bitcoin and Ethereum are the two largest cryptocurrencies in terms of market capitalization thereby making them more trustworthy in the eyes of investors. Second, the returns and volatility of all four cryptocurrencies exhibited a negative relationship during the bust of 2018-2019. This suggests that adverse news effects were an important driver of cryptocurrency returns during the cryptocurrency crash of 2018. Moreover, this transition from a positive to negative relationship between returns and volatility caused these effects to either partially or fully cancel each other out over the full period. Third, we found no evidence of herding behaviour of investors during this period. Fourth, in a recursive out-of-sample forecast exercise, we found that allowing for stochastic volatility and a heavy tailed distribution will generally provide more accurate return and volatility forecasts compared to a random walk benchmark. This suggests that cryptocurrency markets have not been weak-form efficient during this period. Finally, allowing for time-varying parameters and correlation between volatility and returns may further improve the accuracy of volatility forecasts, however, the extent of these improvements differs across the set of cryptocurrencies. For instance, time-varying parameters generally improve upon the volatility forecasts for Bitcoin and Ripple, while accounting for the link between volatility and returns is important for Ethereum and Litecoin.

The present study was aimed at examining the relation between volatility and returns within single cryptocurrency series. For future research it would be useful to consider a multivariate approach that can be used for the entire cryptocurrency market, and a panel approach to test for links between volatility and returns in cryptocurrency markets and other financial markets, e.g. equity markets. It would also be interesting to investigate the importance of other predictors on the relationship between returns and volatilities, such as the volume of daily purchases, searches on the Internet, and mining costs of the various coins.

References

Alizadeh, S., Brandt, M. W., and Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, 57(3):1047–1091.

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905.

Balcilar, M., Bouri, E., Gupta, R., and Roubaud, D. (2017). Can volume predict bitcoin returns and volatility? A quantiles-based approach. *Economic Modelling*, 64:74 – 81.

Baur, D. G. and Dimpfl, T. (2018). Asymmetric volatility in cryptocurrencies. *Economics Letters*, 173:148–151.

Baur, D. G., Hong, K., and Lee, A. D. (2018). Bitcoin: Medium of exchange or speculative assets? *Journal of International Financial Markets, Institutions and Money*, 54:177–189.

Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section*, pages 177–181.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Bouri, E., Azzi, G., and Dyhrberg, A. H. (2017a). On the return-volatility relationship in the bitcoin market around the price crash of 2013. *Economics-The Open-Access, Open-Assessment E-Journal*, 11:1–16.

Bouri, E., Gupta, R., and Roubaud, D. (2019). Herding behaviour in cryptocurrencies. *Finance Research Letters*, 29:216–221.

Bouri, E., Molnár, P., Azzi, G., Roubaud, D., and Hagfors, L. I. (2017b). On the hedge and safe haven properties of bitcoin: Is it really more than a diversifier? *Finance Research Letters*, 20:192–198.

Catania, L. and Grassi, S. (2017). Modelling crypto-currencies financial time-series.

Catania, L., Grassi, S., and Ravazzolo, F. (2018). Predicting the volatility of cryptocurrency time-series. In *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pages 203–207. Springer.

Catania, L., Grassi, S., and Ravazzolo, F. (2019). Forecasting cryptocurrencies under model and parameter instability. *International Journal of Forecasting*, 35(2):485 – 501.

Chan, J. C. C. (2018). Specification tests for time-varying parameter models with stochastic volatility. *Econometric Reviews*, 0(0):1–17.

Chan, J. C. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120.

Chang, E. C., Cheng, J. W., and Khorana, A. (2000). An examination of herd behavior in equity markets: An international perspective. *Journal of Banking & Finance*, 24(10):1651–1679.

Charfeddine, L., Benlagha, N., and Maouchi, Y. (2020). Investigating the dynamic relationship between cryptocurrencies and conventional assets: Implications for financial investors. *Economic Modelling*, 85:198–217.

Cheah, E.-T. and Fry, J. (2015). Speculative bubbles in bitcoin markets? an empirical investigation into the fundamental value of bitcoin. *Economics letters*, 130:32–36.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial Economics*, 10(4):407–432.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of financial Economics*, 19(1):3–29.

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for gaussian and partial non-gaussian state space models. *Journal of Econometrics*, 154(1):85 – 100.

Geisel, M. S. (1973). Bayesian comparisons of simple macroeconomic models. *Journal of Money, Credit and Banking*, 5(3):751–772.

Geweke, J. (1993). Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics*, 8(S1):S19–S40.

Geweke, J. (2005). *Contemporary Bayesian econometrics and statistics*. John Wiley & Sons.

Glaser, F., Zimmermann, K., Haferkorn, M., and Weber, M. C. (2014). Bitcoin - asset or currency? revealing users' hidden intentions. *International Finance eJournal*.

Hanley, B. P. (2013). The false premises and promises of bitcoin. *arXiv preprint arXiv:1312.2048*.

Harvey, A. C. and Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics*, 14(4):429–434.

Hosszejni, D. and Kastner, G. (2019). Approaches toward the bayesian estimation of the stochastic volatility model with leverage. *Working paper*.

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

Jiang, Y., Lie, J., Wang, J., and Mu, J. (2021). Revisiting the roles of cryptocurrencies in stock markets: A quantile coherency perspective. *Economic Modelling*, 95:21–34.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kastner, G. and Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting mcmc estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76:408 – 423. CFEnetwork: The Annals of Computational and Financial Econometrics.

Kristoufek, L. (2013). Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific reports*, 3(1):1–7.

Li, Y., Liu, X.-B., and Yu, J. (2015). A bayesian chi-squared test for hypothesis testing. *Journal of Econometrics*, 189(1):54–69.

Li, Y., Zeng, T., and Yu, J. (2014). A new approach to bayesian hypothesis testing. *Journal of Econometrics*, 178:602–612.

McCausland, W. J., Miller, S., and Pelletier, D. (2011). Simulation smoothing for state-space models: A computational efficiency analysis. *Computational Statistics and Data Analysis*, 55(1):199–212.

Nadarajah, S. and Chu, J. (2017). On the inefficiency of bitcoin. *Economics Letters*, 150:6–9. cited By 55.

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

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *The Journal of Business*, 53(1):61–65.

Qin, M., Su, C.-W., and Tao, R. (2021). Bitcoin: A new basket for eggs? *Economic Modelling*, 94:896–907.

Schwert, G. W. (1989). Why does stock market volatility change over time? *The journal of finance*, 44(5):1115–1153.

Stavroyiannis, S. and Babalos, V. (2019). Herding behavior in cryptocurrencies revisited: Novel evidence from a tvp model. *Journal of Behavioral and Experimental Finance*, 22:57–63.

Urom, C., Abid, I., Guesmi, K., and Chevallier, J. (2020). Quantile spillovers and dependence between bitcoin, equities and strategic commodities. *Economic Modelling*, 93:230–258.

Urquhart, A. (2016). The inefficiency of bitcoin. *Economics Letters*, 148:80–82. cited By 86.

Verdinelli, I. and Wasserman, L. (1995). Computing bayes factors using a generalization of the savage-dickey density ratio. *Journal of the American Statistical Association*, 90(430):614–618.

Woo, D., Gordon, I., and Laralov, V. (2013). Bitcoin: A first assessment fx and rates. *Global Bank of America, Merrill Lynch*.

Yermack, D. (2015). Is bitcoin a real currency? an economic appraisal. In *Handbook of digital currency*, pages 31–43. Elsevier.

Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics*, 127(2):165–178.

Zhang, J.-Y., Chen, Z.-T., and Li, Y. (2018). Bayesian testing for leverage effect in stochastic volatility models. *Computational Economics*.

A Priors and MCMC algorithm for the TVP-AR-SVL-t model

To complete our model specification in Section 3.3, we assume the following set of independent priors for the parameters:

$$\begin{aligned}
&\theta_0 \sim \mathcal{N}(\theta_{00}, V_{\theta_0}), && \phi_h \sim \mathcal{N}(\phi_0, V_\phi)1(-1 < \phi_h < 1), && \mu_h \sim \mathcal{N}(\mu_{h0}, V_{\mu_h}),\\
&\rho \sim \mathcal{N}(\rho_0, V_\rho)1(-1 < \rho < 1), && \omega_\theta \sim \mathcal{N}(0, V_\theta), && \omega_h \sim \mathcal{N}(0, V_{\omega_h}),\\
&\nu \sim \mathcal{U}(a_0, b_0).
\end{aligned}$$

We set a relatively non-informative prior for initial values of the time-varying parameters that is centered on the zero vector θ_0, i.e., $\theta_{00} = 0$ and $V_{\theta_0} = 5I_3$, where I_3 is the identity matrix of size 3. We also center the correlation between returns and volatility parameter ρ, on zero and allow for a wide variance $\rho_0 = 0$ and $V_\rho = 5$. For the log-volatility, we set $\phi_0 = 0.97$, $V_\phi = 0.1^2$, $\mu_{h0} = -10$, and $V_{\mu_h} = 5$. For ω_θ and ω_h, we set $V_\theta = 0.1^2 I_3$ and $V_{\omega_h} = 0.1^2$. For the degree of freedom parameter ν, we set $a_0 = 0$ and $b_0 = 50$.

In what follows, we discuss how to estimate the model with the following 5 block Metropolis-within-Gibbs sampler:

1. $\mathrm{p}\Big(\widetilde{\theta}|y, \widetilde{\theta}_0, \omega_\theta, h, \mu_h, \omega_h, \phi_h, \rho, \eta, \nu\Big)$
2. $\mathrm{p}\Big(\widetilde{\theta}_0, \omega_\theta|y, \widetilde{\theta}, h, \mu_h, \omega_h, \phi_h, \rho, \eta, \nu\Big)$
3. $\mathrm{p}\Big(h|y, \widetilde{\theta}, \widetilde{\theta}_0, \omega_\theta, \mu_h, \omega_h, \phi_h, \rho, \eta, \nu\Big)$
4. $\mathrm{p}\Big(\mu_h, \omega_h, \phi_h|y, \widetilde{\theta}, \widetilde{\theta}_0, \omega_\theta, h, \rho, \eta, \nu\Big)$
5. $\mathrm{p}\Big(\rho, \eta, \nu|y, \widetilde{\theta}, \widetilde{\theta}_0, \omega_\theta, h, \mu_h, \omega_h, \phi_h\Big)$

Our empirical estimates are based on a thinned chain of 10000 posterior draws in which we save every fifth draw after a burnin-in period of 5000 draws.

A.1 Sampling $\widetilde{\theta}$

A major difficulty in sampling the states is that correlation in the state and measurement equations renders direct application of the standard Kalman filter infeasible. To overcome this hurdle, we exploit the fact that the precision matrix of the implied posterior is a band matrix and bypass the Kalman filter, instead drawing upon a class of band and sparse operations which have been shown to be more efficient than the Kalman filter in a variety of state space models (Chan and Jeliazkov, 2009; McCausland et al., 2011).

To derive the posterior, first note that the conditional distribution of y_t is given by

$$(y_t|\mu_t, \alpha_t, \eta_t, h_t, h_{t+1}, \phi_h, \rho, \omega_h^2) \sim \mathcal{N}\Big(\mu_t + \alpha_t y_{t-1} + \frac{\rho \eta_t^{\frac{1}{2}} e^{\frac{1}{2}h_t}}{|\omega_h|}(h_{t+1} - \mu_h - \phi_h(h_t - \mu_h)), \\ \eta_t e^{h_t}(1-\rho^2)\Big).$$

To combine this with the prior density, we first define $\widetilde{y}_t = y_t - \frac{\rho \eta_t^{\frac{1}{2}} e^{\frac{1}{2}h_t}}{|\omega_h|}(h_{t+1} - \mu_h - \phi_h(h_t - \mu_h))$ and $x_t = (1, y_{t-1})$. This is useful because we can write the non-centered parameterization of the SVL model as

$$\widetilde{y}_t = x_t\theta_0 + x_t\Omega_\theta^{\frac{1}{2}}\widetilde{\theta}_t + \widetilde{\epsilon}_t^y, \qquad \widetilde{\epsilon}_t^y \sim \mathcal{N}(0, \eta_t e^{h_t}(1-\rho^2)), \tag{A.1}$$

$$\widetilde{\theta}_t = \widetilde{\theta}_{t-1} + \widetilde{\epsilon}_t^\theta, \qquad \widetilde{\epsilon}_t^\theta \sim \mathcal{N}(0, I_2), \tag{A.2}$$

where $\Omega_\theta = \text{diag}(\omega_\mu^2, \omega_\alpha^2)$, and $\theta_0 \sim \mathcal{N}(\theta_{00}, V_{\theta_{00}})$.

Stacking the above equations over $t = 1, \ldots, T$, gives

$$\widetilde{y} = X\widetilde{\theta} + \widetilde{\epsilon}^y, \qquad \widetilde{\epsilon}^y \sim \mathcal{N}(0, \Sigma_y),$$
$$H_\theta\widetilde{\theta} = \widetilde{\epsilon}_t^\theta, \qquad \widetilde{\epsilon}^\theta \sim \mathcal{N}(0, I_{2T}),$$

where $\widetilde{y} = (y_1 - x_1\theta_0, \ldots, y_T - x_T\theta_0)'$, $\widetilde{\theta} = (\widetilde{\theta}_1', \ldots, \widetilde{\theta}_T')'$, $X = \text{diag}(x_1\Omega_\theta^{\frac{1}{2}}, \ldots, x_T\Omega_\theta^{\frac{1}{2}})$, $\Sigma_y = \text{diag}(\eta_1 e^{h_1}(1-\rho^2), \ldots, \eta_T e^{h_T}(1-\rho^2))$ and

$$H_\theta = \begin{pmatrix} I_2 & 0 & \cdots & \cdots & 0 \\ -I_2 & I_2 & \vdots & \cdots & 0 \\ 0 & -I_2 & I_2 & \vdots & 0 \\ 0 & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -I_2 & I_2 \end{pmatrix}.$$

Using the fact that H_θ is non-singular, it is straightforward to show that the full conditional distribution is

$$\tilde{\theta} \sim \mathcal{N}(\widehat{\theta}, \widehat{V}_\theta^{-1}),$$

where $\widehat{V}_\theta = X'\Sigma^{-1}X + H_\theta'H_\theta$ and $\widehat{\theta} = \widehat{V}_\theta^{-1}X'\Sigma^{-1}\widetilde{y}$.

Since the precision matrix $\widehat{V}_\theta$ is a band matrix, sampling from this Gaussian distribution can be efficiently conducted via a precision-based algorithm (see, e.g., Chan and Jeliazkov, 2009). In particular, since $\widehat{V}_\theta$ is a band matrix, its (lower) Cholesky decomposition $C_{\tilde{\theta}}$ can be quickly obtained. Then, forward and backward substitutions on the expression

$\widehat{V}_\theta \widehat{\theta} = X'\Sigma^{-1}\widetilde{y}$ gives

$$\widehat{\theta} = (C'_{\tilde{\theta}})^{-1}(C_{\tilde{\theta}}^{-1} X'\Sigma^{-1}\widetilde{y}).$$

Finally, a draw of $\theta \sim \mathcal{N}(\hat{\theta}, V_{\tilde{\theta}}^{-1})$ can be obtained from

$$\tilde{\theta} = \hat{\theta} + (C_{\tilde{\theta}}^{-1})'Z,$$

where Z is a $T \times 1$ vector of standard normal random variables, i.e., $Z \sim \mathcal{N}(0, I_T)$. It is easy to check that $\mathbb{E}(\tilde{\theta}) = \hat{\theta}$ and $Var(\tilde{\theta}) = (C_{\tilde{\theta}} C'_{\tilde{\theta}})^{-1} = V_{\tilde{\theta}}^{-1}$.

A.2 Sampling $(\theta_0, \omega_\theta)$

We sample $\gamma = (\theta_0, \omega_\theta)$ by adapting the procedure in Frühwirth-Schnatter and Wagner (2010) to our model. To get the likelihood, note that equation (A.1) can be written as

$$\widetilde{y}_t = x_t\theta_0 + x_t \text{diag}(\widetilde{\theta})\omega_\theta + \widetilde{\epsilon}_t^y, \quad \widetilde{\epsilon}_t^y \sim \mathcal{N}(0, \eta_t e^{h_t}(1-\rho^2)),$$

where $\text{diag}(\widetilde{\theta})$ is a diagonal matrix with $(\widetilde{\mu}_t, \widetilde{\alpha}_t)$ on its diagonal and $\omega_\theta = (\omega_\mu, \omega_\alpha)'$. Stacking the equation over $t = 1, \ldots, T$ gives

$$\widetilde{y} = Z\gamma + \widetilde{\epsilon}^y, \quad \widetilde{\epsilon}^y \sim \mathcal{N}(0, \Sigma_y).$$

Next, note that the independent prior $\gamma \sim \mathcal{N}(\gamma_0, V_\gamma)$, with $\gamma_0 = (\theta'_{00}, 0'_{2\times 1})'$ and $V_\gamma = \text{diag}(V_{\theta_0}, V_{\omega_\theta})$. Straightforward algebra shows that the full conditional distribution of γ is

$$\gamma \sim \mathcal{N}(\widehat{\gamma}, \widehat{V}_\gamma^{-1}),$$

where $\widehat{V}_\gamma = Z'\Sigma_y^{-1}Z + V_\gamma$ and $\widehat{\gamma} = \widehat{V}_\gamma^{-1}\left(Z'\Sigma_y^{-1}\widetilde{y} + V_\gamma^{-1}\gamma_0\right)$. Since positive and negative values are equally likely, the posterior distribution is fully explored through a random sign switching for $(\omega_\mu, \widetilde{\mu}_t)$ and $(\omega_\alpha, \widetilde{\alpha}_t)$ after each time we draw γ.

A.3 Sampling h

Since the full conditional distribution for the log-volatility is not standard, direct sampling is infeasible. To overcome this problem, Jacquier et al. (2004) linearize the observation equation and approximate the implied log-normal kernel by an inverse gamma distribution. As discussed in Section 3, however, a major shortcoming of their specification is that they rely on the unappealing assumption of contemporaneous dependence between the returns and volatility equations, i.e. $\text{corr}(\varepsilon_t^y, \varepsilon_t^h) = \rho$, as opposed to an intertemporal dependence, i.e. $\text{corr}(\varepsilon_t^y, \varepsilon_{t+1}^h) = \rho$. From a sampling perspective, this means that their approximation will forego information on the dependence between the two error terms (Harvey and Shephard, 1996; Yu, 2005).

To overcome this concern, Yu (2005) implemented a Metropolis–Hastings updating step in the all-purpose Bayesian software package BUGS (Bayesian inference Using Gibbs Sampling). While extremely flexible, a major computational drawback with BUGS is that the current Metropolis-Hastings algorithm is based on a Gaussian proposal distribution, whose standard deviation is tuned over the first 4000 iterations in order to get an acceptance rate of between 20% and 40%.[14] Since this is computationally inefficient, we instead propose an acceptance-rejection Metropolis-Hastings step that uses the Newton-Raphson method on the log-posterior distribution to locate the Gaussian proposal distribution on the mode and set the precision matrix to be the negative Hessian of the the log-posterior evaluated at the mode.

To this end, we first derive the gradient vector and the negative Hessian matrix for the log-likelihood function. For notational convenience, let $dy_t = y_t - \mu_t - \alpha_t y_{t-1}$, $k_t = h_{t+1} - \mu_h - \phi_h(h_t - \mu_h)$, $a_t = \frac{\rho\eta_t^{\frac{1}{2}}}{|\omega_h|}$, $b_t = \eta_t(1-\rho^2)$ and $p_t = p(y_t|\mu_t, \alpha_t, \eta_t, h_t, h_{t+1}, \phi_h, \rho, \omega_h^2)$. We have

$$\log p_t = c_1 - \frac{1}{2}h_t - \frac{1}{2b_t}e^{-h_t}\left(dy_t - a_t k_t e^{\frac{1}{2}h_t}\right)^2, \tag{A.3}$$

where c_1 is a constant that is independent of h. It is also worth noting that dy_t, a_t and b_t are independent of h, but k_t is dependent of h. The gradiant vector and the negative Hessian of the $\sum_{t=1}^{T} \log p_t$ evaluated at $\widetilde{h}$ are given by

$$f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{T+1} \end{pmatrix}, \quad G = \begin{pmatrix} G_{1,1} & G_{1,2} & 0 & \cdots & 0 \\ G_{2,1} & G_{2,2} & G_{2,3} & \cdots & 0 \\ \vdots & \ddots & \ddots & \cdots & \vdots \\ 0 & \cdots & G_{T-1,T} & G_{T,T} & G_{T,T+1} \\ 0 & \cdots & 0 & G_{T,T+1} & G_{T+1,T+1} \end{pmatrix},$$

where $f_1 = \frac{\partial \log p_1}{\partial h_t}\big|_{h=\widetilde{h}}$ and $G_{1,1} = -\frac{\partial^2 \log p_t}{\partial h_t^2}\big|_{h=\widetilde{h}}$. For $t = 2, \ldots, T+1$,

$$f_t = \frac{\partial}{\partial h_t}(\log p_t + \log p_{t-1})\Big|_{h=\widetilde{h}}, G_{t,t} = -\frac{\partial^2}{\partial h_t^2}(\log p_t + \log p_{t-1})\Big|_{h=\widetilde{h}},$$
$$G_{t-1,t} = -\frac{\partial^2}{\partial h_{t-1}\partial h_t}\log p_{t-1}\Big|_{h=\widetilde{h}}.$$

[14] Further information about the BUGs program can be found here: https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/

It can be shown that

$$
\begin{aligned}
\frac{\partial \log p_t}{\partial h_t} &= -\frac{1}{2} + \frac{dy_t^2}{2b_t} e^{-h_t} + \frac{a_t^2 k_t \phi_h}{b_t} - \frac{dy_t a_t}{b_t} e^{-\frac{1}{2}h_t} \left(\frac{k_t}{2} + \phi_h \right), \\
\frac{\partial^2 \log p_t}{\partial h_t^2} &= -\frac{dy_t^2}{2b_t} e^{-h_t} - \frac{a_t^2 \phi_h^2}{b_t} + \frac{dy_t a_t}{b_t} e^{-\frac{1}{2}h_t} \left(\phi_h + \frac{k_t}{4} \right), \\
\frac{\partial \log p_t}{\partial h_{t+1}} &= -\frac{a_t^2 k_t}{b_t} + \frac{dy_t a_t}{b_t} e^{-\frac{1}{2}h_t}, \\
\frac{\partial^2 \log p_t}{\partial h_{t+1}^2} &= -\frac{a_t^2}{b_t}, \\
\frac{\partial^2 \log p_t}{\partial h_t \partial h_{t+1}} &= \frac{a_t^2 \phi_h}{b_t} - \frac{dy_t a_t}{2b_t} e^{-\frac{1}{2}h_t}.
\end{aligned}
$$

Given the gradient vector and negative Hessian matrix, we can approximate the log likelihood around a given point $\widetilde{h}$ by a second-order Taylor expansion

$$
\begin{aligned}
\log p(y|\theta, h, \eta, \phi_h, \mu_h, \rho, \omega_h^2) &\approx p(y|\theta, \widetilde{h}, \eta, \phi_h, \mu_h, \rho, \omega_h^2) + (h - \widetilde{h})' f - \frac{1}{2}(h - \widetilde{h})' G (h - \widetilde{h}), \\
&= -\frac{1}{2}(h' G h - 2h'(f + G\widetilde{h})) + c_2, \qquad \text{(A.4)}
\end{aligned}
$$

where c_2 is a constant independent of h.

Next, we note that the transition equation for h_t can be written as

$$
H_{\phi_h} h = \delta_h + \epsilon^h, \quad \epsilon^h \sim \mathcal{N}(0, \Sigma_h), \qquad \text{(A.5)}
$$

where $\delta_h = (\mu_h, (1-\phi_h)\mu_h, \ldots, (1-\phi_h)\mu_h)'$, $\Sigma_h = \text{diag}(\omega_h^2/(1-\phi_h^2), \omega_h^2, \ldots, \omega_h^2)$ and

$$
H_{\phi_h} = \begin{pmatrix}
1 & 0 & 0 & \cdots & 0 \\
-\phi_h & 1 & 0 & \cdots & 0 \\
0 & -\phi_h & 1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & -\phi_h & 1
\end{pmatrix}.
$$

The log density function is given by

$$
\log p(h|\mu_h, \phi_h, \omega_h^2) = c_3 - \frac{1}{2}\left(h' H_{\phi_h}' \Sigma_h^{-1} H_{\phi_h} h - 2h' H_{\phi_h} \Sigma_h^{-1} \delta_h \right). \qquad \text{(A.6)}
$$

The log full conditional density function of h can be obtained by combining equation (A.4)

and equation (A.6) as follows

$$\begin{aligned}\log p(h|y,\theta,\eta,\phi_h,\mu_h,\rho,\omega_h^2) &= \log p(y|\theta,h,\eta,\phi_h,\mu_h,\rho,\omega_h^2) + \log p(h|\mu_h,\phi_h,\omega_h^2) + c_4 \\ &= -\frac{1}{2}(h'K_h h - 2h'k_h) + c_5,\end{aligned}$$

where c_4 and c_5 are normalization constants whcih are independent of h, $K_h = H'_{\phi_h}\Sigma_h^{-1}H_{\phi_h} + G$ and $k_h = f + G\widetilde{h} + H'_{\phi_h}\Sigma_h^{-1}\delta_h$. It can be seen that the above equation is the log-kernel of a Gaussian distribution. To obtain a draw of h from the full conditional distribution, we need to select $\widetilde{h}$. To that end, we choose $\widetilde{h}$ to be the mode of the log distribution $\log p(h|y,\theta,\eta,\phi_h,\mu_h,\rho,\omega_h^2)$, and obtain it using the Newton-Raphson method. The Gaussian distribution $\mathcal{N}(\widetilde{h},\widetilde{K}_h^{-1})$ then is used as our proposal distribution in the acceptance-rejection Metropolis-Hastings step, where $\widetilde{K}_h^{-1}$ is the K_h^{-1} evaluated at $\widetilde{h}$.

A.4 Sampling $(\mu_h, \omega_h, \phi_h)$

To further improve the simulation efficiency over that in Yu (2005), we build upon Kastner and Frühwirth-Schnatter (2014) and Hosszejni and Kastner (2019) and implement the Ancillarity Sufficiency Interweaving Strategy (ASIS) for drawing the latent parameters in the stochastic volatility process. Specifically, we draw μ_h, ω_h and ϕ_h as follows

1. draw $(\mu_h, \omega_h, \phi_h)$ using the centered parameterization of the model, then set $\widetilde{h}_t = \frac{h_t - \mu_h}{\omega_h}$;
2. redraw $(\mu_h, \omega_h, \phi_h)$ using the non-centered prarameterization of the model, then set $h_t = \mu_h + \omega_h \widetilde{h}_t$.

For Step (1), we implement a three-block sampler. In the first instance, to draw μ_h, we first let $X_{\mu_h} = (1, (1-\phi_h), \ldots, (1-\phi_h))'$, then equation (A.5) can be re-expressed as

$$H_{\phi_h} h = X_{\mu_h}\mu_h + \epsilon^h, \quad \epsilon^h \sim \mathcal{N}(0, \Sigma_h),$$

then it can be shown that

$$\mu_h \sim \mathcal{N}(\widehat{\mu}_h, \widehat{V}_{\mu_h}),$$

with $\widehat{V}_{\mu_h}^{-1} = V_{\mu_h}^{-1} + X'_{\mu_h}\Sigma_h^{-1}X_{\mu_h}$ and $\widehat{\mu}_h = \widehat{V}_{\mu_h}(X'_{\mu_h}\Sigma_h^{-1}H_{\phi_h}h + V_{\mu_h}^{-1}\mu_{h0})$.

Second, since the conditional distribution of ϕ_h is nonstandard, we implement an independence-chain Metropolis-Hastings step to obtain draw of ϕ_h with a truncated Gaussian proposal $\mathcal{N}(\widehat{\phi}_h, \widehat{V}_{\phi_h})\mathbb{1}(|\phi_h| < 1)$ where $\widehat{V}_{\phi_h}^{-1} = X'_{\phi_h}X_{\phi_h}/\omega_h^2 + V_{\phi_h}^{-1}$ and $\widehat{\phi}_h = \widehat{V}_{\phi_h}(X'_{\phi_h}y_{\phi_h}/\omega_h^2 + V_{\phi_h}^{-1}\phi_{h0})$, with $X_{\phi_h} = (h_1 - \mu_h, \ldots, h_T - \mu_h)'$ and $y_{\phi_h} = (h_2 - \mu_h, \ldots, h_{T+1} - \mu_h)'$.

Third, an independence-chain Metroplis-Hastings step is used to obtain sample of ω_h^2 with a truncated Gaussian poposal $\mathcal{N}(\widehat{\omega}_h^2, \widehat{V}_{\omega_h}^{-1})\mathbb{1}(\omega_h^2 > 0)$, where $\widehat{\omega}_h^2$ is the mode for the log

density function

$$\log(\omega_h^2|y,\theta,h,\eta,\phi_h,\mu_h,\rho) = c_6 + \log p(h|\omega_h^2,\mu_h,\phi_h) + \log p(\omega_h^2) + \sum_{t=1}^{T} \log p_t,$$

where $\log p_t$ is given in equation (A.3), $\log p(h|\omega_h^2,\mu_h,\phi_h)$ can be derived from equation (A.6). By a change of variable, it can be shown that $\omega_h \sim \mathcal{N}(0, V_{\omega_h})$ implies $\omega_h^2 \sim \mathcal{G}(\frac{1}{2}, \frac{1}{2V_{\omega_h}})$. The $\widehat{V}_{\omega_h}$ is selected to be the negative Hessian of the $\log(\omega_h^2|y,\theta,h,\eta,\phi_h,\mu_h,\rho)$.

To implement Step (2), a two-block sampler is proposed. In the first instance, we draw (μ_h,ω_h) using an independent-chain Metroplis-Hastings step. Note that under the non-centered parameterization, it can be shown that the log likelihood is given by

$$\begin{aligned}\log p(y_t|\mu_h,\omega_h,\mu_t,\alpha_t,\eta_t,\widetilde{h}_t,\widetilde{h}_{t+1},\phi_h,\rho) = c_7 &- \frac{1}{2}(\mu_h + \omega_h\widetilde{h}_t) \\ &- \frac{1}{2b_t} e^{-(\mu_h+\omega_h\widetilde{h}_t)} \left(dy_t - \widetilde{k}_t e^{\frac{1}{2}(\mu_h+\omega_h\widetilde{h}_t)}\right)^2,\end{aligned}$$

where $\widetilde{k}_t = \rho\eta_t^{\frac{1}{2}}(\widetilde{h}_{t+1} - \phi_h\widetilde{h}_t)$, dy_t and b_t are defined previously in the step of sampling h. Given $\omega_w \sim \mathcal{N}(0, V_{\omega_h})$ and $\mu_h \sim \mathcal{N}(\mu_{h0}, V_{\mu_h})$, we have

$$\log p(\mu_h,\omega_h) = c_8 - \frac{\omega_h^2}{2V_{\omega_h}} - \frac{(\mu_h-\mu_{h0})^2}{2V_{\mu_h}}.$$

Hence the log density of the full conditional distribution of (μ_h,ω_h) is given by

$$\begin{aligned}\log p(\mu_h,\omega_h|\theta,\widetilde{h}_t,\phi_h,\rho,\eta,y) =& c_9 + \sum_{t=1}^{T}\left(-\frac{1}{2}(\mu_h+\omega_h\widetilde{h}_t) - \frac{1}{2b_t}e^{-(\mu_h+\omega_h\widetilde{h}_t)}\left(dy_t - \widetilde{k}_t e^{\frac{1}{2}(\mu_h+\omega_h\widetilde{h}_t)}\right)^2\right) \\ &- \frac{\omega_h^2}{2V_{\omega_h}} - \frac{(\mu_h-\mu_{h0})^2}{2V_{\mu_h}}.\end{aligned}$$

To sample (μ_h,ω_h), we first find the mode and the negative Hessian evaluated at the mode of $\log p(\mu_h,\omega_h|\theta,\widetilde{h}_t,\phi_h,\rho,\eta,y)$, denoted respectively by $(\widehat{\mu}_h,\widehat{\omega}_h)$ and $K_{(\mu_h,\omega_h)}$. We then use the distribution $\mathcal{N}\left((\widehat{\mu}_h,\widehat{\omega}_h)', K^{-1}_{(\mu_h,\omega_h)}\right)$ as a proposal in the Metroplis-Hastings step.

Second, to draw ϕ_h under the non-centered parameterization we use a similar strategy to that in the Step 1. Specifically, we implement an independence-chain Metropolis-Hastings step to obtain draw of ϕ_h with a truncated Gaussian proposal $\mathcal{N}(\widetilde{\phi}_h, \widetilde{V}_{\phi_h})\mathbb{1}(|\phi_h|<1)$ where $\widetilde{V}_{\phi_h}^{-1} = \widetilde{X}'_{\phi_h}\widetilde{X}_{\phi_h} + V_{\phi_h}^{-1}$ and $\widetilde{\phi}_h = \widetilde{V}_{\phi_h}(\widetilde{X}'_{\phi_h}\widetilde{y}_{\phi_h} + V_{\phi_h}^{-1}\phi_{h0})$, with $\widetilde{X}_{\phi_h} = (h_1-\mu_h,\ldots,h_T-\mu_h)'$ and $\widetilde{y}_{\phi_h} = (\widetilde{h}_2,\ldots,\widetilde{h}_{T+1})'$.

A.5 Sampling (ρ, η, ν)

The remaining parameters are sampled in a standard manner. In the first instance, we note that the log full conditional density of ρ is

$$\log p(\rho|y, \theta, h, \rho, \phi_h, \mu_h, \eta, \omega_h) \propto \log(\rho) - \frac{T}{2}\log(1-\rho^2) - \frac{1}{2}\sum_{t=1}^{T}\left(\frac{1}{2\eta_t(1-\rho^2)}e^{-h_t}\left(dy_t - \rho\kappa_t\right)^2\right),$$

where $\kappa_t = \frac{\eta_t e^{\frac{1}{2}h_t}}{|\omega_h|}(h_{t+1} - \mu_h - \phi_h(h_t - \mu_h))$. As ρ is bounded, the Griddy-Gibbs method can be applied for obtaining draw of ρ.

Second, note that $(\eta_1, \ldots \eta_T)$ are conditionally independent, we sample η_t by using an independence-chain Metropolis-Hastings step with proposal $\mathcal{IG}(\widehat{\alpha}_{\eta_t}, \widehat{\beta}_{\eta_t})$ for drawing η_t, where

$$\widehat{\alpha}_{\eta_t} = \frac{\nu+1}{2}, \quad \widehat{\beta}_{\eta_t} = \frac{1}{2}\left(\nu + \frac{e^{-h_t}}{1-\rho^2}(y_t - \mu_t - \alpha_t y_{t-1})^2\right).$$

It can be shown that given the current draw η_t, a draw from the proposal, i.e., $\eta_t^* \sim \mathcal{IG}(\widehat{\alpha}_{\eta_t}, \widehat{\beta}_{\eta_t})$, will be accepted with probability

$$\min\left(1, \exp\left(\delta_t(\eta_t^{-1/2} - \eta_t^{*-1/2})\right)\right),$$

where $\delta_t = \frac{\rho e^{-\frac{1}{2}h_t}}{(1-\rho^2)|\omega_h|}(y_t - \mu_t - \alpha_t y_{t-1})(h_{t+1} - \mu_h - \phi_h(h_t - \mu_h))$.

Finally, the degree of freedom parameter ν is sampled using an independence-chain Metropolis-Hastings step with proposal distribution $\mathcal{N}(\widehat{\nu}, K_\nu^{-1})$, where $\widehat{\nu}$ is the mode of $\log p(\nu|\eta)$ and K_ν is the negative Hessian evaluated at the mode.

B Estimating the Savage-Dickey density ratio

In this appendix we outline the Monte Carlo method used to compute the Savage-Dickey density ratio. To set the stage, recall that the Savage-Dickey density ratio is given by

$$\text{BF}_{ur} = \frac{p(\Phi = 0)}{p(\Phi = 0|y)}.$$

The numerator can be evaluated analytically, however the denominator has no closed form solution. We therefore utilize Monte Carlo methods to evaluate this term. To that end, let Θ be the vector collecting all model parameters except Φ. Then, the denominator can

be evaluated using the Monte Carlo average

$$\hat{p}(\Phi = 0|y) = \frac{1}{R}\sum_{i=1}^{R} p(\Phi = 0|\Theta^i, y),$$

where $(\Theta^1, \ldots, \Theta^R)$ are posterior draws that can be obtained with the sampler in Appendix A. Since the full conditional posterior distribution $p(\Phi|\Theta, y)$ may not be standard, the normalizing constant of this distribution is unknown. We therefore utilize numerical methods. To be precise, suppose that $p(\Phi|\Theta, y) = c^{-1}\tilde{p}(\Phi|\Theta, y)$, where $\tilde{p}(\Phi|\Theta, y)$ has an analytical expression, but the normalizing constant $c = \int \tilde{p}(\Phi|\Theta, y)$ is unknown. Then, c can be estimated by computing $\int \tilde{p}(\Phi|\Theta, y)$ numerically. Importantly, since Φ has a low dimension (i.e., one or two), the numerical error for computing this constant is negligible.

C Monte Carlo Simulation

In this section, we investigate the performance of our proposed algorithm in estimating the parameters for the TVP-AR-SVL-t through a Monte Carlo simulation exercise.[15] To facilitate further discussion, we re-present the model as follows:

$$y_t = \mu_t + \alpha_t y_{t-1} + \varepsilon_t^y, \tag{C.1}$$

$$\mu_t = \mu_{t-1} + \varepsilon_t^\mu, \qquad \varepsilon_t^\mu \sim \mathcal{N}(0, \omega_\mu^2), \tag{C.2}$$

$$\alpha_t = \alpha_{t-1} + \varepsilon_t^\alpha, \qquad \varepsilon_t^\alpha \sim \mathcal{N}(0, \omega_\alpha^2), \tag{C.3}$$

$$h_{t+1} = \mu_h + \phi_h(h_t - \mu_h) + \varepsilon_{t+1}^h, \tag{C.4}$$

$$\begin{pmatrix} \varepsilon_t^y \\ \varepsilon_{t+1}^h \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \eta_t e^{h_t} & \rho\eta_t^{1/2} e^{h_t/2}\omega_h \\ \rho\eta_t^{1/2} e^{h_t/2}\omega_h & \omega_h^2 \end{pmatrix}\right), \tag{C.5}$$

$$\eta_t \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2}). \tag{C.6}$$

We generate a dataset with $T = 500$ observations using the model described in equations (C.1) - (C.6) with the following true parameter values: $\mu_h = 0$, $\phi_h = 0.97$, $\rho = 0.5$, $\nu = 3$, $\omega_\alpha^2 = \omega_\mu^2 = 0.001$ and $\omega_h^2 = 0.1$. The initial conditions of the state parameters are set as: $\mu_0 = 0, \alpha_0 = 0, h_1 \sim \mathcal{N}(0, \frac{0.1}{1-0.97^2})$.

The model parameters are estimated with 10000 retained posterior draws after a burn-in period of 5000 draws. The posterior means and 95% credible intervals for the time varying coefficients μ_t, α_t and log-volatility h_t are plotted in Figure C.1. Estimation results for the static coefficients are reported in Table C.1. The results show that the true values are generally well estimated by the posterior means and are always within the 95% credible intervals.

[15] We refer readers to Appendix A for more details of the proposed MCMC estimation method.

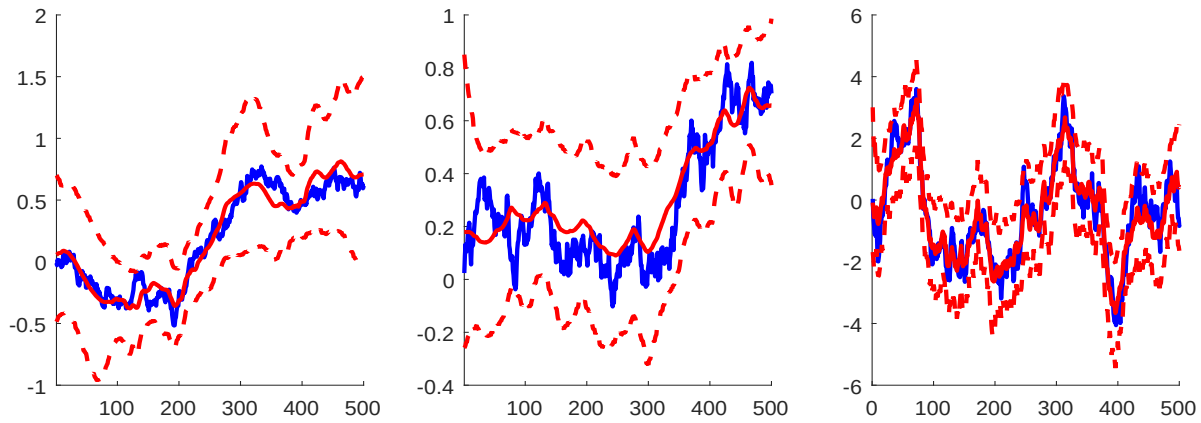

Figure C.1: Estimated posterior means and the 95% credible intervals for μ_t (left), α_t (middle) and h_t (right). The red solid lines are the posterior means,the dotted red lines are the 95% credible intervals and the blue solid lines are the true values.

Table C.1: Estimated posterior means and the 95% Credible intervals.

Parameters	**True value**	**Posterior mean**	**95% Credible interval**
ρ	0.50	0.48	(0.23, 0.66)
μ_h	0.00	-0.04	(-0.36, 0.57)
ϕ_h	0.97	0.96	(0.946, 0.993)
ν	5.00	5.43	(3.87, 6.83)
ω_μ^2	0.001	0.0012	(0.0005, 0.0047)
ω_α^2	0.001	0.0009	(0.0002, 0.0023)
ω_h^2	0.10	0.12	(0.07, 0.16)