



Bayesian nonlinear expectation for time series modelling and its application to Bitcoin

Tak Kuen Siu¹

Received: 21 July 2021 / Accepted: 25 April 2022 / Published online: 25 May 2022
© The Author(s) 2022, corrected publication 2022

Abstract

This paper proposes a two-stage approach to parametric nonlinear time series modelling in discrete time with the objective of incorporating uncertainty or misspecification in the conditional mean and volatility. At the first stage, a reference or approximating time series model is specified and estimated. At the second stage, Bayesian nonlinear expectations are introduced to incorporate model uncertainty or misspecification in prediction via specifying a family of alternative models. The Bayesian nonlinear expectations for prediction are constructed from closed-form Bayesian credible intervals evaluated using conjugate priors and residuals of the estimated approximating model. Using real Bitcoin data including some periods of Covid 19, applications of the proposed method to forecasting and risk evaluation of Bitcoin are discussed via three major parametric nonlinear time series models, namely the self-exciting threshold autoregressive model, the generalized autoregressive conditional heteroscedasticity model and the stochastic volatility model.

Keywords Parametric time series modelling · Nonlinear expectations · Bayesian statistics · Girsanov’s transform · Drift and volatility uncertainties · Bitcoin

JEL Classification C22 · C11 · C58

1 Introduction

Human activities in taking time series observations, such as sunspot numbers, have a long history. An early inquiry for “formally” modelling time series may be tracked back to the seminal works of Yule (1927) and Slutsky (1927), where linear time series models were introduced to describe cyclical behaviour of time series observations.

✉ Tak Kuen Siu
Ken.Siu@mq.edu.au, ktsiu2005@gmail.com

¹ Department of Actuarial Studies and Business Analytics, Macquarie Business School, Macquarie University, Sydney, NSW 2109, Australia

See, for example, Tong (1990, Page 18), and Hansen (2013, Page 400). Owing to the realization of limitations of linear time series models in explaining nonlinear real-world phenomena in the late 1970s, some major parametric nonlinear time series models were introduced in the late 1970s and the 1980s, such as the threshold autoregressive (TAR) model and its sub-class, namely the self-exciting TAR (SETAR) model, pioneered by Tong (1977, 1978, 1983), the autoregressive conditional heteroscedasticity (ARCH) model pioneered by Engle (1982), the stochastic volatility (SV) model of Taylor (1982, 1986), the generalized ARCH (GARCH) model in Bollerslev (1986) and Taylor (1986) and the Markov-switching autoregressive model in Hamilton (1989) (see also Tong (1983)). See Tong (2002) for further discussions.

Though many significant problems in time series modelling have been studied, it seems that model uncertainty may not have received as much attention as those “mainstream” topics in the subject. It is, however, that in certain modern practical applications of time series modelling, such as artificial intelligence and Bitcoin forecasting and risk analysis, model uncertainty may play a significant role. In economics and econometrics, where time series analysis has been widely applied, the significance of incorporating model uncertainty and the “worst-case” scenario approach to model uncertainty were discussed in the Prize Lecture of Nobel Prize in Economic Sciences by Hansen (2013). The key idea of the “worst-case” scenario approach is to introduce a family of alternative models from a reference or approximating model and make decisions in the “worst-case” scenario over the set of alternative models (see Hansen and Sargent (2007)). Studies on model uncertainty, from the economics perspective, may be traced back to the distinction between risk and uncertainty in the classic monograph by Knight (1921), (see, for example, Hansen (2013), Section 6), where risk and uncertainty were referred to the situations where the probability models are known and where they are unknown, respectively.

Nonlinear expectations provide a plausible way to describe model uncertainty. The notion of nonlinear expectations and their stochastic calculus were introduced in the seminal works by Peng (1997, 2004, 2006, 2019), where the g -expectations for incorporating drift uncertainty and the G -expectations for describing volatility uncertainty were introduced. The mathematics for the G -expectations are more complicated than the g -expectations in a continuous-time modelling framework since the former involves singular measure changes, while the latter only involves absolutely continuous measure changes. Discrete-time nonlinear expectations were introduced in Elliott (2017) and Elliott and Siu (2017) with the objective of providing a convenient way to incorporating both drift and volatility uncertainties in normally distributed random shocks using a discrete-time Girsanov’s transformation in, for example, Elliott et al. (1995) and Elliott and Madan (1998). Specifically, both drift and volatility uncertainties are incorporated by discrete-time nonlinear expectations with absolutely continuous measure changes. A key issue for implementing nonlinear expectations is how a set of probability models defining the nonlinear expectations might be chosen. Specifically, from the practical perspective, it seems desirable to specify the set of probability models so that certain probability levels describing credibility of the set of probability models may be assigned. However, this issue has yet not been well-addressed in Elliott (2017), Elliott and Siu (2017), and possibly, in some other literature on nonlinear expectations.

The purpose of this paper is to propose a two-stage approach to parametric nonlinear time series modelling in discrete time with the objective of incorporating model uncertainty or misspecification in the conditional mean and volatility, which are two key components in nonlinear time series modelling. At the first stage, a reference or approximating time series model is specified. Then the reference model is estimated from given data. At the second stage, Bayesian nonlinear expectations are introduced to incorporate model uncertainty or misspecification in prediction via specifying a family of alternative models. Specifically, closed-form Bayesian credible intervals characterizing model uncertainty or misspecification in the conditional mean and volatility are constructed using conjugate priors and residuals of the estimated approximating model from the first stage. It may be noted that the residuals contain information that is left over, which may be attributed to model misspecification in the conditional mean and volatility of the reference or approximating model estimated from the first stage. The misspecifications in the conditional mean and volatility are estimated using the closed-form Bayesian credible intervals, and discrete-time Girsanov's transforms are used to introduce the family of alternative models from the reference or approximating model. Those alternative models have misspecifications in the conditional mean and volatility lying in the ranges specified by the respective Bayesian credible intervals. Using the closed-form Bayesian credible intervals, the Bayesian nonlinear expectations incorporating those alternative models in prediction are then constructed. The closed-form Bayesian credible intervals have the advantage that they provide a convenient way to construct the Bayesian nonlinear expectations so that probability levels are assigned to the nonlinear expectations. It may also be noted that the closed-form Bayesian credible intervals provide a feasible way to incorporate both a model user's subjective view (or expert opinion) and the objective data in estimating misspecifications in the conditional mean and volatility. Indeed, under the Bayesian statistical framework, the "unknown" misspecifications in the conditional mean and volatility are considered random variables, and this may provide a natural way to characterize model uncertainties about the conditional mean and volatility. These uncertainties are called drift and volatility uncertainties. Using real data in a Bitcoin exchange rates series including some volatile periods which may be due to Covid 19, applications of the proposed approach to forecasting and risk evaluation of Bitcoin are discussed. Three major nonlinear time series models, namely the SETAR model, the GARCH model and the SV model, are used for illustrating the applications. Two major tail-based risk metrics, say value at risk (VaR) and expected shortfall (ES), are used for risk evaluation of Bitcoin. Through the lens of Bayesian nonlinear expectations, the impacts of model uncertainty or misspecification in the conditional mean and volatility on forecasting and risk evaluation of Bitcoin are examined. It is hoped that the approach may throw light on exploring the interplay among time series modelling, Bayesian theory, nonlinear expectations, model uncertainty and their applications.

The proposed approach may be related to the Bayesian approach to nonlinear time series and Bayesian econometrics. See, for example, Pole and Smith (1985) for Bayesian analysis of SETAR models and Koop (2003) for an excellent account on Bayesian econometrics. However, the key difference between them is that the proposed approach is a two-stage approach, where the reference model is specified and estimated in the first stage, and the second stage is concerned with prediction in the presence of

model uncertainty or misspecification using the Bayesian nonlinear expectations over a set of alternative models. The worst-case scenario approach to model uncertainty in economic decision-making was introduced in discrete-time economic models in, for example, Cagetti et al. (2002), Hansen et al. (2002) and Hansen and Sargent (2007). However, their focuses, modelling structures and methods are different from those of the current paper. Siu and Yang (1999) and Siu et al. (2001) adopted a Bayesian approach to generating probability scenarios for evaluating coherent risk measures whose representations are related to nonlinear expectations, (see Artzner et al. 1999 and Rosazza Gianin 2006). However, Siu and Yang (1999) and Siu et al. (2001) determined a set of probability models defining the coherent risk measures using a set of prior distributions. Siu (2001) used Bayesian credible intervals for evaluating interval estimates for coherent risk measures for derivatives under a discrete-time binomial tree model. Frazier et al. (2020) considered model misspecification in approximate Bayesian computation. However, the focuses and modelling structures of Siu (2001) and Frazier et al. (2020) are different from those of the current paper.

The rest of the paper is structured as follows. A brief introduction to nonlinear expectations is provided in the next section. Section 3 presents parametric nonlinear time series models with drift and volatility uncertainties using discrete-time Girsanov's transforms, Gaussian uncertain noises and nonlinear expectations. The Bayesian nonlinear expectations are constructed using Bayesian credible intervals in Sect. 4. The estimation, forecasting and risk evaluation procedures are discussed in Sect. 5. The applications of the proposed approach to Bitcoin forecasting and risk evaluation using real data are provided in Sect. 6. The impacts of drift and volatility uncertainties on Bitcoin forecasting and risk evaluation are also examined. Section 7 gives some concluding remarks. Some tables and figures are presented in Appendices I–II, which are included after the bibliography. Online Appendices A–G provide some technical details, derivations, constructions, discussions on potential links to other important Bayesian techniques and potential generalizations of the proposed two-stage approach.

2 g -Expectations, G -expectations and model uncertainty

In this section, a brief review on two types of nonlinear expectations, namely the g -expectation and the G -expectation, and their links with model uncertainty or ambiguity is provided. Some technical details on these are presented in Online Appendix A. The g -expectation was introduced by Peng (1997), where the g -expectation was defined by the solution of a backward stochastic differential equation. The g -expectation describes model uncertainty or ambiguity about the drift of a process, which is also called the drift uncertainty. The G -expectation was also introduced by Peng (2006) and describes model uncertainty or ambiguity about the volatility of a process, which is called the volatility uncertainty.

The notion of nonlinear expectations is closely linked with the concept of coherent risk measures in Artzner et al. (1999). Indeed, as noted by Peng (2004, 2006), the notion of nonlinear expectations are equivalent to the concept of coherent risk measures. Rosazza Gianin (2006) established the link between conditional g -expectations and dynamic risk measures. The concepts of coherent risk measures and nonlinear

expectations also play a fundamental role in the two-price conic finance in, for example, Cherny and Madan (2009) and Madan and Cherny (2010), where market cones generating two-price systems were defined by sets of acceptable risks. A (conditional) nonlinear expectation satisfying the sub-additivity property is called a (conditional) sublinear expectation. A (conditional) expectation satisfying the super-additivity property is called a (conditional) superlinear expectation.

In a continuous-time modelling framework, a conditional g -expectation is defined by the solution of a backward stochastic differential equation. See, for example, Chen and Epstein (2002), where the conditional g -expectation was used to discuss a stochastic differential utility with multiple priors in the presence of model uncertainty or ambiguity in continuous time. The specification of a family of alternative models with the drift uncertainty described by the conditional g -expectation involves absolutely continuous probability measures. Specifically, the family of alternative models can be specified by applying Girsanov's transforms for changing probability measures in continuous time.

The G -expectation was motivated by some works on option valuation with volatility uncertainty. Boyle and Ananthanarayanan (1977) introduced the use of Bayesian statistics to estimate volatility for option valuation, where Bayesian conjugate prior was used to capture the estimation risk of volatility and to construct interval estimates for option prices. Avellaneda et al. (1995) and Lyons (1995) introduced volatility uncertainty to option valuation and hedging, where the "uncertain" volatility was supposed to lie in a bounded interval and a convex region, respectively. Fouque and Ren (2014) considered the use of a conditional G -expectation in continuous time to study option valuation under volatility uncertainty. In fact, in a continuous-time modelling framework, the specification of alternative models with the volatility uncertainty described by the conditional G -expectation involves singular probability measures. Consequently, it is more complicated than the specification of a family of alternative models with the drift uncertainty described by the conditional g -expectation. However, in a discrete-time modelling framework, both the drift and volatility uncertainties can be specified by a discrete-time Girsanov's transform as in Elliott (2017) and Elliott and Siu (2017). Indeed, in the discrete-time situation, only probability measures that are absolutely continuous to the reference probability measure are used to define a family of alternative models in the presence of both the drift and volatility uncertainties.

3 Parametric nonlinear time series models with drift and volatility uncertainties

Two general parametric nonlinear time series models incorporating uncertainties in the conditional mean (or drift) and volatility are introduced using discrete-time Girsanov's transforms, Gaussian uncertain noises and nonlinear expectations. The first model is a nonlinear autoregressive model with conditional heteroscedasticity which includes the SETAR model and the GARCH model. The second model is a general specification which includes the first model and a product process. Since the product process includes the SV model, the second model also includes the SV model. The two models are introduced in accordance with the two-stage approach. At the

first stage, reference or approximating models¹ are specified. At the second stage, discrete-time Girsanov's transforms are used to introduce Gaussian uncertain noises and sets of alternative models from the reference models. Then the (conditional) nonlinear expectations which incorporate the sets of alternative models in prediction are defined, and some motivations for introducing the Bayesian nonlinear expectations for prediction in the second stage are discussed. The idea of using (Gaussian) uncertain noises and nonlinear expectations in filtering and estimation of hidden Markov models appeared in Elliott (2017), where discrete-time Girsanov's transforms in, for example, Elliott et al. (1995) and Elliott and Madan (1998), were used to introduce uncertain noises and the respective family of alternative models. The work of Elliott (2017) was then extended to the filtering and estimation of hidden Markov-modulated stochastic volatility models with drift and volatility uncertainties in Elliott and Siu (2017).

A complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is considered, where \mathbb{P} is a reference probability measure corresponding to an approximating model. A time index set \mathbb{T} is defined as $\{1, 2, \dots, T\}$, where T is the overall number of observations in time series data and future time steps in prediction. Let $\{X_t\}_{t \in \mathbb{T}}$ be a real-valued time series process,² or stochastic process, defined on $(\Omega, \mathcal{F}, \mathbb{P})$. In the sequel, the first and second models for $\{X_t\}_{t \in \mathbb{T}}$ are described in Sects. 3.1 and 3.2, respectively.

3.1 A nonlinear autoregressive model with heteroscedasticity

Let $\{\epsilon_t\}_{t \in \mathbb{T}}$ be a sequence of independent and identically distributed (i.i.d.) standard normal random variables under the reference probability measure \mathbb{P} . That is, under \mathbb{P} , $\epsilon_t \sim N(0, 1)$, for each $t \in \mathbb{T}$. It is supposed that under \mathbb{P} , $\{X_t\}_{t \in \mathbb{T}}$ follows the reference parametric nonlinear time series model:

$$X_t = f(X_{t-1}, X_{t-2}, \dots, X_{t-p}) + g(X_{t-1}, X_{t-2}, \dots, X_{t-q})\epsilon_t, \quad (3.1)$$

where $f : \mathfrak{R}^p \rightarrow \mathfrak{R}$ and $g : \mathfrak{R}^q \rightarrow \mathfrak{R}$ are two given measurable functions; p and q are two positive integers. It is assumed, for simplicity, that $g > 0$. The reference model in Eq. (3.1) is a nonlinear autoregressive model with conditional heteroscedasticity, which incorporates some important parametric nonlinear time series models such as the SETAR model and the GARCH model. It is assumed that the initial observations $\{X_{-\max(p,q)}, X_{-\max(p,q)+1}, \dots, X_0\}$ are given. When f is a linear function and g is a constant function, the model in Eq. (3.1) becomes a linear time series model with homoscedastic errors.

A discrete-time Girsanov's transform is now employed to construct a family of probability measures equivalent to the reference measure \mathbb{P} . The family of equivalent probability measures generates Gaussian uncertain noises with drift and volatility uncertainties and the respective alternative models.

Let $\theta := (\mu, \sigma)' = (\mu(\theta), \sigma(\theta))' \in \mathfrak{R} \times \mathfrak{R}_+$, where μ may be interpreted as misspecification in the conditional mean (or drift) of the reference model and σ may be interpreted as misspecification in the volatility of the reference model. By definition,

¹ The term "a baseline model" was used in Hansen and Sargent (2021).

² In general one may consider a multiple time series process.

μ is considered the first coordinate $\mu(\theta)$ of θ , while σ is regarded as the second coordinate $\sigma(\theta)$ of θ . If $\mu = 0$, there is no misspecification in the drift of the reference model. Likewise, if $\sigma = 1$, there is no misspecification in the volatility of the reference model. In practice, the misspecifications in the drift and volatility may not be known. Consequently, μ and σ may be regarded as “uncertain” parameters. One may describe these “uncertain” parameters by specifying the ranges in which those parameters lie. Mathematically, this may be done by considering the following product intervals:

$$\Theta := [\mu^-, \mu^+] \times [\sigma^-, \sigma^+], \tag{3.2}$$

for some $\mu^-, \mu^+ \in \mathfrak{R}$ and $\sigma^-, \sigma^+ \in \mathfrak{R}_+$ with $\mu^- < \mu^+$ and $\sigma^- < \sigma^+$. Then, the family $\{(\mu(\theta), \sigma(\theta)) | \theta \in \Theta\}$ may be used to describe the “uncertain” parameters (μ, σ) and to characterize drift and volatility uncertainties.

Let $\phi_{\mu, \sigma}(x)$ denote the probability density function (pdf) of a normal distribution $N(\mu, \sigma^2)$ with mean μ and variance σ^2 . Write $\phi(x)$ for $\phi_{0,1}(x)$. Let \mathbb{F}^X be the \mathbb{P} -augmentation of the natural filtration $\{\mathcal{F}_t^X\}_{t \in \mathbb{T}}$ generated by the time series process $\{X_t\}_{t \in \mathbb{T}}$. That is, for each $t \in \mathbb{T}$, \mathcal{F}_t^X is the \mathbb{P} -completed σ -field generated by the observations $\{X_1, X_2, \dots, X_t\}$. For each $\theta \in \Theta$, let $\{\lambda_t(\theta)\}_{t \in \mathbb{T}}$ be an \mathbb{F}^X -adapted process on $(\Omega, \mathcal{F}, \mathbb{P})$ defined by:

$$\lambda_t(\theta) := \frac{\phi_{\mu, \sigma}(\epsilon_t)}{\phi(\epsilon_t)} = \frac{\phi\left(\frac{\epsilon_t - \mu}{\sigma}\right)}{\sigma \phi(\epsilon_t)}. \tag{3.3}$$

Consider, for each $\theta \in \Theta$, the following \mathbb{F}^X -adapted process $\{\Lambda_t(\theta)\}_{t \in \mathbb{T}}$:

$$\Lambda_t(\theta) := \prod_{k=1}^t \lambda_k(\theta). \tag{3.4}$$

It is known that for each $\theta \in \Theta$, $\{\Lambda_t(\theta)\}_{t \in \mathbb{T}}$ is an $(\mathbb{F}^X, \mathbb{P})$ -martingale, and so $E[\Lambda_t(\theta)] = 1$. Consequently, it can be used as a density process for changing probability measures on \mathcal{F}_T^X , and a new probability measure \mathbb{P}^θ equivalent to \mathbb{P} on \mathcal{F}_T^X , for each $\theta \in \Theta$, can be defined as:

$$\left. \frac{d\mathbb{P}^\theta}{d\mathbb{P}} \right|_{\mathcal{F}_T^X} := \Lambda_T(\theta). \tag{3.5}$$

Using a discrete-time Girsanov’s theorem, (see, for example, Elliott et al. (1995) and Elliott and Siu (2017)), for each $\theta \in \Theta$, under a new probability measure \mathbb{P}^θ , a family $\{\epsilon_t(\theta)\}_{t \in \mathbb{T}}$ of random variables defined by putting:

$$\epsilon_t(\theta) := \frac{\epsilon_t - \mu}{\sigma}, \tag{3.6}$$

is a sequence of i.i.d. standard normal random variables. That is, under \mathbb{P}^θ ,

$$\epsilon_1(\theta), \epsilon_2(\theta), \dots, \epsilon_T(\theta) \stackrel{i.i.d.}{\sim} N(0, 1).$$

Recall that the family $\{(\mu(\theta), \sigma(\theta)) | \theta \in \Theta\}$ may be used to describe the “uncertain” parameters (μ, σ) for drift and volatility misspecifications. For each $\theta \in \Theta$, let $\epsilon(\theta) := \{\epsilon_t(\theta)\}_{t \in \mathbb{T}}$, which is a sequence of i.i.d. standard normal random variables under \mathbb{P}^θ . Then the Gaussian uncertain noises are defined by the family $\{\epsilon(\theta) | \theta \in \Theta\}$ of sequences of normal random variables indexed by Θ in Eq. (3.2), say the product intervals in which the “uncertain” parameters (μ, σ) lie. It may be noted that the Gaussian uncertain noises $\{\epsilon(\theta) | \theta \in \Theta\}$ are defined with respect to a family of probability measures $\{\mathbb{P}^\theta | \theta \in \Theta\}$ indexed by Θ .

Furthermore, under \mathbb{P}^θ , $\{X_t\}_{t \in \mathbb{T}}$ follows the parametric nonlinear time series model:

$$\begin{aligned} X_t &= f(X_{t-1}, X_{t-2}, \dots, X_{t-p}) + \mu g(X_{t-1}, X_{t-2}, \dots, X_{t-q}) \\ &\quad + \sigma g(X_{t-1}, X_{t-2}, \dots, X_{t-q}) \epsilon_t(\theta). \end{aligned} \quad (3.7)$$

It may be noted that $\{X_t\}_{t \in \mathbb{T}}$ follows the reference model in Eq. (3.1) under the reference measure \mathbb{P} . Under the reference model in Eq. (3.1), the conditional mean and volatility are given by the functions f and g , respectively. However, $\{X_t\}_{t \in \mathbb{T}}$ follows an alternative model in Eq. (3.7) under the new probability measure \mathbb{P}^θ defined by Eq. (3.5), for each $\theta \in \Theta$. Under the alternative model in Eq. (3.7), the conditional mean and volatility are given by the functions $f + \mu g$ and σg , respectively. That is, the effects of changing the probability measures from \mathbb{P} to \mathbb{P}^θ are to (1) perturb the conditional mean function f by μ in the “direction” of the conditional volatility function g and (2) scale the conditional volatility function g by σ . When $\mu = 0$ and $\sigma = 1$, (i.e. there are no drift and volatility misspecifications), the alternative model in Eq. (3.7) coincides with the reference model in Eq. (3.1). Recall that (μ, σ) are the “uncertain” parameters for drift and volatility misspecifications since the misspecifications may not be known in practice. Consequently, it is uncertain which alternative model may be used. Instead of using a single alternative model, again, a family of alternative models may be used. Recall that the focus in the second stage is prediction. Consequently, a family of alternative models for prediction is specified. With the “uncertain” parameters $\theta := (\mu, \sigma)$ varying in $\Theta := [\mu^-, \mu^+] \times [\sigma^-, \sigma^+]$, a family of alternative models in the form of Eq. (3.7) for prediction is defined with respect to the family of probability measures $\{\mathbb{P}^\theta | \theta \in \Theta\}$. That is, for each $\theta \in \Theta$, an alternative model under the probability measure \mathbb{P}^θ has the conditional mean function $f + \mu g$, the conditional volatility function σg and the sequence of i.i.d. standard normal random shocks $\epsilon(\theta)$. It may be noted that for each $\theta \in \Theta$, the probability measure \mathbb{P}^θ defined by the discrete-time Girsanov’s transform in Eq. (3.5) is equivalent to the reference probability measure \mathbb{P} .

To incorporate the family of alternative models in prediction, (conditional) nonlinear expectations are considered here. For example, similar to the “worst-case” scenario approach to model uncertainty in economic decision-making, one may wish to make a conservative prediction by considering the “worst-case” scenario over the family of

alternative models, say the “worst-case” scenario prediction. This may be achieved by considering a pair of conditional nonlinear expectations, namely a conditional sublinear expectation and a conditional superlinear expectation. Specifically, suppose that one wishes to predict a random variable Y given the information described by a sub- σ -field \mathcal{H} of \mathcal{F} . Then a conditional sublinear expectation and a conditional superlinear expectation for Y given \mathcal{H} with respect to $\{\mathbb{P}^\theta | \theta \in \Theta\}$ defining the family of alternative models may be used. The conditional sublinear expectation and the conditional superlinear expectation for Y given \mathcal{H} with respect to $\{\mathbb{P}^\theta | \theta \in \Theta\}$ are, respectively, defined by:

$$\mathbb{E}_\Theta^S[Y|\mathcal{H}] := \text{ess} - \sup_{\theta \in \Theta} E^\theta[Y|\mathcal{H}], \tag{3.8}$$

and

$$\mathbb{E}_\Theta^I[Y|\mathcal{H}] := \text{ess} - \inf_{\theta \in \Theta} E^\theta[Y|\mathcal{H}], \tag{3.9}$$

where $E^\theta[\cdot|\mathcal{H}]$ is the conditional expectation given \mathcal{H} under \mathbb{P}^θ . $\text{ess} - \sup$ and $\text{ess} - \inf$ are the essential supremum and the essential infimum, respectively. The essential supremum and the essential infimum are taken with respect to the “uncertain” parameters $\theta := (\mu, \sigma)$ varying in the product intervals Θ . Consequently, the family of alternative models defined by $\{\mathbb{P}^\theta | \theta \in \Theta\}$ is incorporated in the definitions of the conditional sublinear expectation in Eq. (3.8) and the conditional superlinear expectation in Eq. (3.9). From the technical perspective, the essential supremum and the essential infimum are used since the conditional expectation $E^\theta[Y|\mathcal{H}]$ is a random variable, for each $\theta \in \Theta$. Also, to simplify the discussion, it is supposed that the random variable Y is integrable with respect to \mathbb{P}^θ , (i.e. $E^\theta[|Y|] < \infty$, where $E^\theta[\cdot]$ is the expectation under \mathbb{P}^θ), for each $\theta \in \Theta$.

Depending on the purposes of applications, either the conditional sublinear expectation or the conditional superlinear expectation may be used as the “worst-case” scenario prediction. Specifically, if one wishes to predict the future profit (loss) described by the random variable Y based on the current and past information described by \mathcal{H} , then the conditional superlinear expectation (the conditional sublinear expectation) may be used as the “worst-case” scenario prediction. Instead of making a point prediction, one may also make an interval prediction based on a pair of conditional superlinear and sublinear expectations, where the former and the latter give the lower and upper limits of the interval prediction, respectively.

Whether one wishes to use the conditional nonlinear (sublinear and/or superlinear) expectations for making a point prediction or an interval prediction, a key issue to be addressed is to determine the upper limits, say μ^+, σ^+ , and the lower limits, say μ^-, σ^- , in the product intervals Θ . In the next section, concepts in Bayesian statistics will be used to inform the determination of the upper and lower limits in Θ . Specifically, Bayesian credible intervals are used to determine the upper and lower limits in Θ to which certain probability levels can be assigned to indicate the degree of credibility that the family of alternative models defined by $\{\mathbb{P}^\theta | \theta \in \Theta\}$ may have. The Bayesian credible intervals also incorporate both a model user’s subjective view

(or expert opinion) and the objective data in determining the upper and lower limits in Θ . The Bayesian nonlinear expectations for prediction in the second stage are then defined by using the conditional nonlinear expectations in Eq. (3.8) and Eq. (3.9) with the upper and lower limits in Θ determined by the Bayesian credible intervals as well as with the given information \mathcal{H} specified by a σ -field generated by the time series observations.

3.2 A product process

Though the reference model in Eq. (3.1) (i.e. a nonlinear autoregressive model with conditional heteroscedasticity) includes the GARCH model, it does not include a SV model. This is because the SV model involves two random shocks per unit of time while the reference model in Eq. (3.1) involves one random shock per unit of time. To include the SV model, a product process is needed. As noted in Sherphard (2005), the product process was introduced in an unpublished paper of Rosenberg (1972) and was studied in Taylor (1980, 1982). It was pointed out in Taylor (1982) that the product process is not included in some nonlinear time series models such as the bilinear model in Granger and Anderson (1978) and the state-dependent model in Priestley (1980) for the product process involves two random shocks per unit of time.

A reference model which includes the first model in Sect. 3.1 and a product process is now specified. Recall from Sect. 3.1 that $\{\epsilon_t\}_{t \in \mathbb{T}}$ is a sequence of i.i.d. standard normal random variables under the reference probability measure \mathbb{P} . Let $\{\eta_t\}_{t \in \mathbb{T}}$ be another sequence of i.i.d. standard normal random variables under \mathbb{P} . Assume that the two sequences $\{\epsilon_t\}_{t \in \mathbb{T}}$ and $\{\eta_t\}_{t \in \mathbb{T}}$ are independent under \mathbb{P} . It is supposed that under \mathbb{P} , $\{X_t\}_{t \in \mathbb{T}}$ follows the reference parametric nonlinear time series model:

$$\begin{aligned} X_t &= f(X_{t-1}, X_{t-2}, \dots, X_{t-p}) + h(X_{t-1}, X_{t-2}, \dots, X_{t-q}, V_t)\epsilon_t, \\ V_t &= f_v(V_{t-1}, V_{t-2}, \dots, V_{t-p_v}) + h_v(V_{t-1}, V_{t-2}, \dots, V_{t-q_v})\eta_t, \end{aligned} \quad (3.10)$$

where $f : \mathbb{R}^p \rightarrow \mathbb{R}$, $h : \mathbb{R}^{q+1} \rightarrow \mathbb{R}$, $f_v : \mathbb{R}^{p_v} \rightarrow \mathbb{R}$ and $h_v : \mathbb{R}^{q_v} \rightarrow \mathbb{R}$ are given measurable functions; $h > 0$ and $h_v > 0$; p, q, p_v and q_v are positive integers. $\{V_t\}_{t \in \mathbb{T}}$ is a latent process, which is a stochastic process defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Defining the time series process $\{X_t\}_{t \in \mathbb{T}}$ and the latent process $\{V_t\}_{t \in \mathbb{T}}$ on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ may avoid the complexity of introducing a product probability space. It is assumed that the initial values $\{V_{-\max(p_v, q_v)}, V_{-\max(p_v, q_v)+1}, \dots, V_0\}$ are also random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. It may be noted that there are two random shocks per unit of time in the reference model in Eq. (3.10). Specifically, at each time $t \in \mathbb{T}$, there are two random shocks ϵ_t and η_t , where ϵ_t is the random shock driving the time series process $\{X_t\}_{t \in \mathbb{T}}$ and η_t is the random shock driving the latent process $\{V_t\}_{t \in \mathbb{T}}$. Consequently, the reference model in Eq. (3.10) includes a product process, which in turn includes the SV model. Also, if the volatility function h in the first equation of Eq. (3.10) becomes the volatility function g in Eq. (3.1), then the reference model in Eq. (3.10), (i.e. the second model), coincides with the reference model in Eq. (3.1), (i.e. the first model).

Suppose that

$$\begin{aligned}
 f(X_{t-1}, X_{t-2}, \dots, X_{t-p}) &= 0, & h(X_{t-1}, X_{t-2}, \dots, X_{t-q}, V_t) &= \exp(V_t/2), \\
 f_v(V_{t-1}, V_{t-2}, \dots, V_{t-p_v}) &= \alpha_v + \delta_v V_{t-1}, & h_v(V_{t-1}, V_{t-2}, \dots, V_{t-q_v}) &= (\delta_v \mathbb{1})
 \end{aligned}$$

Then the reference model in Eq. (3.10) becomes:

$$\begin{aligned}
 X_t &= \exp(V_t/2)\epsilon_t, \\
 V_t &= \alpha_v + \delta_v V_{t-1} + \sigma_v \eta_t.
 \end{aligned}
 \tag{3.12}$$

This is the discrete-time SV model presented in Jacquier et al. (2002). The discrete-time SV model, which is also called a lognormal autoregressive model for a latent volatility process, was first proposed by Taylor (1982). For Bitcoin applications to be presented in Sect. 6, the discrete-time SV model in Eq. (3.12) will be used.

After specifying the reference model in Eq. (3.10) in the first stage, the second stage is considered. Again, a discrete-time Girsanov’s transform is used to construct a family of probability measures equivalent to the reference measure \mathbb{P} . The family of probability measures is used to define Gaussian “uncertain” noises and a family of alternative models from the reference model in Eq. (3.10). The constructions and derivations are presented in Online Appendix B.

4 Bayesian nonlinear expectations

The second stage of the proposed two-stage approach is concerned with prediction, which is an ultimate goal of time series modelling. If an estimated reference model from the first stage is used for prediction, there are two concerns. Firstly, as discussed in Sect. 3, there may be model misspecifications or uncertainties in the conditional mean and volatility of the reference model. Indeed, the “true” underlying data generating process, if exists, is often unknown to model users in practice. Consequently, it appears that model misspecifications or uncertainties are ubiquitous. Secondly, the future may not follow from the present and the past. Consequently, there may be a concern that the future developments of a time series process may not follow what are predicted from the reference model estimated from the present and past data. To address these concerns, a family of alternative models is used to predict the future developments of the time series process in the second stage. Specifically, the discrete-time Bayesian nonlinear expectations incorporating a family of alternative models for prediction are constructed in this section. To construct the Bayesian nonlinear expectations, the first step is to evaluate the residuals of an estimated reference model. The residuals may be thought of as proxies/estimates for the random errors in the reference model which in turns contain information that is left over due to model misspecifications in the conditional mean and volatility of the reference model. Then a Bayesian model for the random errors in the reference model is constructed by treating the “uncertain” parameters for drift and volatility misspecifications as random variables. Using conjugate priors for the “uncertain” parameters and the residuals of the estimated reference model as

proxies/estimates for the random errors, closed-form Bayesian credible intervals for the “uncertain” parameters are obtained. The closed-form Bayesian credible intervals, specifically their lower and upper limits, are then used to determine the lower and upper limits of the product intervals indexing the family of probability measures which defines the family of alternative models to be incorporated in prediction. Effectively, these alternative models are introduced by extracting information that is left over due to model misspecifications in the conditional mean and volatility of the reference model via Bayesian credible intervals. The Bayesian nonlinear expectations are given by the conditional nonlinear expectations with the lower and upper limits of the product intervals determined by the Bayesian credible intervals.

Bayesian statistics, as it may be quite well-known, is an important approach to model uncertainty. Indeed, Bayesian statistics has a solid foundation in decision theory under uncertainty as established in, for example, Bernardo and Smith (2000). As far as prediction is concerned, the Bayesian averaging approach has been adopted to incorporate model uncertainty in evaluating Bayesian predictive distributions. The basic idea of the Bayesian averaging approach is to evaluate the predictive distributions by averaging out full posteriors of the unknown parameters. If the model has many unknown parameters, the evaluation of the predictive distributions using the Bayesian averaging approach will involve computing high-dimensional integrals. Also, if the parametric form of the model is complicated, closed-form solutions to the posteriors may be very difficult, if not impossible, to obtain using conjugate priors. In this case, numerical simulations based on, for example, Markov Chain Monte Carlo (MCMC) methods are required for computing the full posteriors. It appears that the Bayesian averaging approach may be related to the smooth ambiguity approach for decision-making with model uncertainty proposed in Klibanoff et al. (2005), where Bayesian averaging was used in constructing the criterion for preferences. The Bayesian nonlinear expectations to be introduced in this section provide an alternative way to incorporate model uncertainties or misspecifications in prediction. Like the conditional nonlinear expectations as discussed in Sect. 3.1, the Bayesian nonlinear expectations can be used to provide a conservative prediction by considering the “worst-case” scenario over the family of alternative models, say the “worst-case” scenario prediction. It may be thought that the Bayesian nonlinear expectations combine concepts in Bayesian statistics and the “worst-case” scenario approach to model uncertainty for making predictions. Using the closed-form Bayesian credible intervals based on conjugate priors, the Bayesian nonlinear expectations can be easily computed and provide a convenient way to make predictions. Furthermore, the use of the Bayesian credible intervals allows probability levels to be assigned with the objective of indicating credibility of predictions made by Bayesian nonlinear expectations. As pointed out by Hansen (2013), Page 431 therein, under the smooth ambiguity approach, ambiguity was identified as a characteristic of subjective beliefs of a model user and a Bayesian prior was assigned to alternative models. This may also apply to both the Bayesian averaging approach and the Bayesian nonlinear expectations approach. Indeed, both the Bayesian averaging approach and the Bayesian nonlinear expectations approach incorporate a model user’s subjective view (or expert opinion) and the objective data in making predictions. The model user’s subjective view (or expert opinion) may be used to adjust predictions based

solely on the present and the past data in case the future may not follow the present and the past.

In the sequel, the Bayesian credible intervals and the Bayesian nonlinear expectations corresponding to the first model in Sect. 3.1 will be constructed. The construction of the Bayesian credible intervals and the Bayesian nonlinear expectations corresponding to the second model in Sect. 3.2 follows similarly and is presented in Online Appendix C. Some formulas and notations in Bernardo and Smith (2000), (see, for example, Pages 121 and 440 therein), will be used. For an excellent account of Bayesian statistics, one may refer to, for example, Box and Tiao (1992), Bernardo and Smith (2000) and Lee (2012). In Online Appendix D, a (potential) fusion between Bayesian nonlinear expectations and the Bayesian shrinkage and regularization techniques for estimation and variable selection is discussed using the proposed two-stage approach. The variable selection is an important topic in econometrics, statistics and machine learning. As far as the sparsity of a high-dimensional regression model is concerned, the Bayesian shrinkage and regularization techniques in the Bayesian model selection literature are the key techniques for estimation and variable selection, (see, for example, Carvalho et al. 2009; Polson and Scott 2010; Bhadra et al. 2019 and the relevant references therein). It may be of interest to explore the (potential) link between the Bayesian shrinkage and regularization techniques and the Bayesian nonlinear expectations.

The Bayesian nonlinear expectations corresponding to the first model in Sect. 3.1 are now constructed. From Eq. (3.6), for each $\theta \in \Theta$,

$$\epsilon_t = \mu + \sigma \epsilon_t(\theta), \tag{4.1}$$

where (μ, σ) are the ‘‘uncertain’’ parameters for the drift and volatility misspecifications in the reference model in Eq. (3.1). In Bayesian statistics, μ and σ are assumed to be random variables, and prior distributions for μ and σ are assigned. Recall from Sect. 3.1 that $\{\epsilon_t(\theta)\}_{t \in \mathbb{T}}$ is a sequence of i.i.d. standard normal random variables under the probability measure \mathbb{P}^θ , for each $\theta \in \Theta$. Then from Eq. (4.1), under \mathbb{P}^θ , given (μ, σ) , $\{\epsilon_t\}_{t \in \mathbb{T}}$ is a sequence of conditionally i.i.d. normal random variables with mean μ and standard deviation σ . That is, under \mathbb{P}^θ , $\epsilon_1, \epsilon_2, \dots, \epsilon_T | (\mu, \sigma) \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$. It may be noted that under the reference probability measure \mathbb{P} , $\epsilon_1, \epsilon_2, \dots, \epsilon_T \stackrel{i.i.d.}{\sim} N(0, 1)$. Here the precision λ , where $\lambda = \frac{1}{\sigma^2}$, is considered, and a prior distribution is assigned for λ instead of σ . Again, if $\lambda = 1$, then there is no volatility misspecification in a reference model. If $\lambda > 1$, (i.e. $\sigma < 1$), then the volatility specification in the reference model in Eq. (3.1) may be too high. If $\lambda < 1$, (i.e. $\sigma > 1$), then the volatility specification in the reference model in Eq. (3.1) may be too low. Consequently, either λ or σ may be used to describe volatility misspecifications in the reference model in Eq. (3.1).

To assign prior distributions for (μ, λ) , a normal-gamma prior distribution is considered. Specifically,

$$\mu | \lambda \sim N\left(\mu_0, \frac{1}{t_0 \lambda}\right), \quad \lambda \sim Ga(\alpha, \beta), \tag{4.2}$$

where $N(\mu_0, \frac{1}{t_0\lambda})$ is a normal distribution with mean μ_0 and variance $\frac{1}{t_0\lambda}$, and $Ga(\alpha, \beta)$ is a Gamma distribution with shape parameter α and rate parameter β . Consequently, the prior mean of μ is μ_0 and the prior variance of $\mu|\lambda$ is $\frac{1}{t_0\lambda}$. The prior mean μ_0 may be thought of as a prior estimate for μ , and the prior precision $t_0\lambda$ may indicate the strength of belief on the prior estimate μ_0 . Say, for a fixed value of λ , a larger (smaller) prior parameter t_0 may express a stronger (weaker) belief on the prior estimate μ_0 . Furthermore, the prior mean and variance of λ are $\frac{\alpha}{\beta}$ and $\frac{\alpha}{\beta^2}$, respectively. A larger (smaller) prior variance $\frac{\alpha}{\beta^2}$ may express a weaker (stronger) belief on the prior estimate $\frac{\alpha}{\beta}$ for λ . Recall that the random variables μ and λ describe misspecifications in the drift and volatility in the reference model in Eq. (3.1). Consequently, when λ becomes smaller, (i.e. $\lambda < 1$), for a fixed prior parameter t_0 , the belief on the prior estimate μ_0 becomes weaker, (i.e. $t_0\lambda$ becomes smaller). To explain the intuition behind this, when λ becomes smaller, the volatility specification in the reference model in Eq. (3.1) may be too low. In this case, the conditional mean specification in the reference model may dominate, in a relative sense, its volatility specification in prediction. Consequently, misspecifications in the conditional mean may have a more pronounced impact on the prediction than misspecifications in the volatility. To be cautious, one may express a weaker belief on the prior estimate μ_0 for misspecifications in the conditional mean. Likewise, when λ becomes larger, (i.e. $\lambda > 1$), for a fixed prior parameter t_0 , the belief on the prior estimate μ_0 becomes stronger, (i.e. $t_0\lambda$ becomes larger). The intuition behind this can be explained similarly. The prior parameters μ_0, t_0, α and β , which are also called the hyperparameters, are chosen by a model user according to the model user’s prior/subjective belief (or expert opinion) on the drift and volatility misspecifications. They will then be combined with the data on the residuals of the reference model to compute the posterior distributions for the drift and volatility misspecifications rationally by the Bayes’ formula.

Let $\{X_1, X_2, \dots, X_n\}$ be a sequence of observations of a time series process, where n is the number of observations. Write \mathbb{N} for the indices of the observations, where $\mathbb{N} := \{1, 2, \dots, n\}$. After estimating the reference model in Eq. (3.1) using the observations $\{X_1, X_2, \dots, X_n\}$ in the first stage, which will be done by the maximum likelihood estimation (MLE) as will be discussed in Sect. 5.1, the residuals of the reference model $\{e_t\}_{t \in \mathbb{N}}$ can be computed as follows:

$$e_t := \frac{X_t - \hat{f}(X_{t-1}, X_{t-2}, \dots, X_{t-p})}{\hat{g}(X_{t-1}, X_{t-2}, \dots, X_{t-q})}, \tag{4.3}$$

where \hat{f} and \hat{g} are the estimates for the conditional mean f and the volatility g , respectively. For each $t \in \mathbb{N}$, the residual e_t is a proxy (or an estimate) for the random error ϵ_t . Recall that under $\mathbb{P}^\theta, \epsilon_1, \epsilon_2, \dots, \epsilon_n | (\mu, \sigma) \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$. It is also assumed that as an approximation, under $\mathbb{P}^\theta, e_1, e_2, \dots, e_n | (\mu, \sigma) \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$.

Suppose that the observations $\mathbf{e}(n) := (e_1, e_2, \dots, e_n)$ are now given. Let

$$\beta_n := \beta + \frac{1}{2}ns^2 + \frac{t_0n(\mu_0 - \bar{e})^2}{2(t_0 + n)}, \tag{4.4}$$

where $\bar{e} = \frac{1}{n} \sum_{t=1}^n e_t$, (i.e. the sample mean of $\mathbf{e}(n)$), and $s^2 = \frac{1}{n} \sum_{t=1}^n (e_t - \bar{e})^2$, (i.e. the sample variance of $\mathbf{e}(n)$). Then it is known (see Bernardo and Smith 2000, Page 440) that by the Bayes' formula, the posterior distribution of λ given $\mathbf{e}(n)$ is:

$$\lambda | \mathbf{e}(n) \sim Ga\left(\alpha + \frac{n}{2}, \beta_n\right), \tag{4.5}$$

where $\alpha + \frac{n}{2}$ is the posterior shape parameter of λ given $\mathbf{e}(n)$, and β_n , which is given by Eq. (4.4), is the posterior rate parameter of λ given $\mathbf{e}(n)$.

Let

$$\mu_n := \frac{t_0 \mu_0 + n \bar{e}}{t_0 + n}, \tag{4.6}$$

and

$$\sigma_n := \sqrt{\frac{\beta_n}{\alpha + \frac{1}{2}n}}, \tag{4.7}$$

where β_n is given by Eq. (4.4). Then, it is also known (see Bernardo and Smith 2000, Page 440) that by the Bayes' formula, the posterior distribution of μ given $\mathbf{e}(n)$ is:

$$\mu | \mathbf{e}(n) \sim St\left(\mu_n, \frac{\sigma_n^2}{(t_0 + n)}, 2\alpha + n\right), \tag{4.8}$$

Here $St(a, b, c)$ is a Student's- t distribution with a location parameter a , a scale parameter \sqrt{b} , (or the precision b^{-1}), and the degree of freedom c ; μ_n in Eq. (4.6) is the posterior location parameter for μ given $\mathbf{e}(n)$. $\frac{\sigma_n}{\sqrt{t_0+n}}$ is the posterior scale parameter of μ given $\mathbf{e}(n)$, where σ_n is given by Eq. (4.7).

Then a $100(1-\gamma_1)\%$ Bayesian credible interval for μ given $\mathbf{e}(n)$ is $(L_\mu(\gamma_1), U_\mu(\gamma_1))$ such that the lower limit $L_\mu(\gamma_1)$ and the upper limit $U_\mu(\gamma_1)$ are, respectively, given by:

$$L_\mu(\gamma_1) = \mu_n - t_{2\alpha+n}\left(\frac{\gamma_1}{2}\right) \frac{\sigma_n}{\sqrt{n+t_0}}, \tag{4.9}$$

and

$$U_\mu(\gamma_1) = \mu_n + t_{2\alpha+n}\left(\frac{\gamma_1}{2}\right) \frac{\sigma_n}{\sqrt{n+t_0}}. \tag{4.10}$$

Here $t_{2\alpha+n}(\frac{\gamma_1}{2})$ is the critical value of a Student's- t distribution with degree of freedom $2\alpha + n$ such that a Student's- t random variable with the same degree of freedom is less than or equal to $t_{2\alpha+n}(\frac{\gamma_1}{2})$ with probability $1 - \frac{\gamma_1}{2}$. When the sample size n becomes large as in the majority of time series data encountered in practice, the Student's- t distribution with degree of freedom $2\alpha + n$ tends to a normal distribution. In this

case, the normal distribution can be used to construct the Bayesian credible interval for μ given $\mathbf{e}(n)$. However, for the sake of generality, the Student's- t distribution is considered here since it can also capture the situation of short time series data.

Similarly, a $100(1 - \gamma_2)\%$ Bayesian credible interval for λ given $\mathbf{e}(n)$ is $(L_\lambda(\gamma_2), U_\lambda(\gamma_2))$ such that the lower limit $L_\lambda(\gamma_2)$ and the upper limit $U_\lambda(\gamma_2)$ are, respectively, given by:

$$L_\lambda(\gamma_2) = \Gamma_{\alpha+\frac{n}{2}, \beta_n} \left(1 - \frac{\gamma_2}{2} \right), \tag{4.11}$$

and

$$U_\lambda(\gamma_2) = \Gamma_{\alpha+\frac{n}{2}, \beta_n} \left(\frac{\gamma_2}{2} \right), \tag{4.12}$$

where $\Gamma_{\alpha+\frac{n}{2}, \beta_n}(\gamma)$ is the critical value of a Gamma distribution with a shape parameter $\alpha + \frac{n}{2}$ and a rate parameter β_n such that a random variable following this distribution is less than or equal to $\Gamma_{\alpha+\frac{n}{2}, \beta_n}(\gamma)$ with probability $1 - \gamma$.

We now take:

$$\mu^- = \mu^-(\gamma_1) = L_\mu(\gamma_1), \quad \mu^+ = \mu^+(\gamma_1) = U_\mu(\gamma_1), \tag{4.13}$$

and

$$\sigma^- = \sigma^-(\gamma_2) = \sqrt{\frac{1}{U_\lambda(\gamma_2)}}, \quad \sigma^+ = \sigma^+(\gamma_2) = \sqrt{\frac{1}{L_\lambda(\gamma_2)}}. \tag{4.14}$$

Consequently, the product intervals Θ in Eq. (3.2) are taken as the following product intervals $\Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))$ with probability levels γ_1 and γ_2 :

$$\Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n)) := [\mu^-(\gamma_1), \mu^+(\gamma_1)] \times [\sigma^-(\gamma_2), \sigma^+(\gamma_2)], \tag{4.15}$$

and a family of alternative models is defined by the family of probability measures $\{\mathbb{P}(\boldsymbol{\theta}) | \boldsymbol{\theta} \in \Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))\}$ indexed by $\Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))$. To emphasize that the product intervals in Eq. (4.15) depend on the observations $\mathbf{e}(n)$ of the residuals, the notation $\Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))$ is used.

It may be noted that the Bayesian credible intervals are used in the second stage which is concerned with prediction, and they are not considered in the first stage which is concerned with the estimation of a reference model. Intuitively, the Bayesian credible intervals combine a model user's prior/subjective belief (or expert opinion) via choosing the prior parameters (or hyperparameters) and information about the residuals $\mathbf{e}(n)$ to generate the family of alternative models with the objective of incorporating drift and volatility misspecifications described by the "uncertain" parameters (μ, σ) . For example, the 90% Bayesian credible intervals for μ and σ , which can be constructed based on the 5th and 95th percentiles of the posterior distributions for μ and σ , may be used to generate the family of alternative models which have the degree

of credibility for capturing the drift and volatility misspecifications described by the 90% probability level. In line with the Bayesian interpretation, the 90% probability level may be thought of as a subjective probability level.

A Bayesian (posterior) sublinear expectation for an integrable random variable Y given the σ -field \mathcal{F}_n^X generated by the time series observations $\{X_1, X_2, \dots, X_n\}$ with respect to $\{\mathbb{P}(\theta) | \theta \in \Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))\}$ is constructed as follows:

$$\mathbb{CBE}_{\Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))}[Y | \mathcal{F}_n^X] := \text{ess - sup}_{\theta \in \Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))} E^\theta[Y | \mathcal{F}_n^X]. \tag{4.16}$$

Similarly, a Bayesian (posterior) superlinear expectation for an integrable random variable Y given \mathcal{F}_n^X with respect to $\{\mathbb{P}(\theta) | \theta \in \Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))\}$ is defined as:

$$\mathbb{CBE}_{\Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))}[Y | \mathcal{F}_n^X] := \text{ess - inf}_{\theta \in \Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))} E^\theta[Y | \mathcal{F}_n^X]. \tag{4.17}$$

5 Estimation, forecasting and risk evaluation

The estimation, forecasting and risk evaluation procedures are described along the line of the two-stage approach. At the first stage, the estimation of the reference model in either Eq. (3.1) or Eq. (3.10) is considered. At the second stage, to incorporate the effects of drift and volatility misspecifications or uncertainties, in prediction, the Bayesian nonlinear expectations are used for forecasting and risk evaluation. For illustration, one-step-ahead forecasts for time series observations are adopted, and one-step-ahead predictions for VaR and ES are considered for risk evaluation.

5.1 Estimation

The focus of this paper is parametric nonlinear time series modelling. It is supposed that the parametric forms in the reference models in Eq. (3.1) or Eq. (3.10) are given a priori. Specifically, the parametric forms of the drift function f and the volatility function g are given in the reference model in Eq. (3.1), and the parametric forms of the drift functions f, f_v and the volatility functions h and h_v are given in the reference model in Eq. (3.10). In these cases, the estimation of the two reference models becomes the estimation of the parameters in the two reference models. Since the reference model in Eq. (3.1) does not involve latent processes, once the parametric forms of the drift and volatility functions are given, the maximum likelihood estimation (MLE) may be used to estimate its parameters. Since the reference model in Eq. (3.10) involves a latent process, once the parametric forms of the drift and volatility functions are given, one may first need to estimate the latent process using filtering techniques (or nonlinear filtering techniques in the case when either the observation process or the latent process is nonlinear) and then use the quasi-MLE (QMLE) to estimate the parameters. Indeed, one may also use a fully Bayesian approach coupled with Markov Chain Monte Carlo (MCMC) to estimate the two reference models. The fully Bayesian approach coupled with MCMC gives full posterior distributions for unknown parameters, while the MLE

and the QMLE give point estimates. While the fully Bayesian approach coupled with MCMC is flexible, the point estimates for the unknown parameters in a reference model are what we need in the first stage of the two-stage approach. Using these point estimates, the residuals of a reference model are computed, which are then used to construct the Bayesian credible intervals in the second stage. Of course, we may also use the point estimates from the posterior means or modes for the unknown parameters to compute the residuals of the reference model. However, the uses of the MLE and the QMLE for the estimation of the reference models may provide a clear illustration for the difference between the objectives of the first and second stages in the proposed approach.

Furthermore, the MLE and the QMLE with the Kalman filter are easy to implement for the SETAR model, the GARCH model and the SV model. The Bayesian MCMC estimation of these three models involves numerical simulations of full posterior distributions for the unknown parameters/latent variables, which could be more complicated. The three models will be estimated using real Bitcoin data in Sect. 6. Specifically, when the reference model is specified as either the SETAR model or the GARCH model, the MLE is used to estimate the model; when the reference model is specified as the SV model, the QMLE coupled with the Kalman filter are used to estimate the model. Indeed, the estimation of the SETAR model and the GARCH model can be done using the function “tar” in the R package “TSA” and the function “garch” in the R package “tseries”, respectively. The estimation of the SV model using the QMLE with the Kalman filter can be implemented quite conveniently using R following the filtering and estimation equations in Appendix B of Jacquier et al. (2002). The estimation of a SV model using the QMLE with the Kalman filter can also be implemented using Excel spreadsheets, (see Taylor 2005, Chapter 11 therein). Besides the QMLE with the Kalman filter, there are other methods to estimate the SV model. See, for example, Taylor (1982, 1986) for moment estimates, Melino and Turnbull (1990) and Jacquier et al. (2002) for the generalized method of moments, Jacquier et al. (2002) and Nakajima and Omori (2009) for a fully Bayesian MCMC approach, Kim et al. (1998) and Omori et al. (2007) for the likelihood-based approaches.

5.2 Forecasting

An upper point forecast, a lower point forecast and an interval forecast for the daily percentage logarithmic return of a financial asset in the next time period given information about the returns up to and including the current time period are obtained under each of the reference models in Eq. (3.1) and Eq. (3.10). The upper point forecast and the lower point forecast are defined using a Bayesian posterior sublinear and superlinear expectations, respectively. Likewise, the lower and upper limits of the interval forecast are specified by a Bayesian posterior sublinear and superlinear expectations, respectively. The interval forecast considered here may entail a different interpretation from the prediction interval used in forecasting. Specifically, the former indicates model uncertainty while the latter indicates uncertainty due to the variability of a forecast. The distinction between the interval forecast considered here and a prediction interval may be illustrated by considering the distinction between uncertainty and risk in the

classic work by Knight (1921). Specifically, the interval forecast considered here may be related to the notion of uncertainty while a prediction interval may be related to the concept of risk in the sense of Knight (1921).

As a comparison, a Bayesian risk-neutral forecast based on the posterior means of the “uncertain” parameters is also obtained under each of the reference models in Eq. (3.1) and Eq. (3.10). The derivations of the upper, lower and interval forecasts as well as the Bayesian risk-neutral forecasts under the two reference models in Eq. (3.1) and Eq. (3.10) are provided in Online Appendix E.

5.3 Risk measures

Two tail-based risk metrics, namely value at risk (VaR) and expected shortfall (ES), are adopted here to evaluate risk. VaR has been a popular risk metric. However, it has been pointed out that VaR does not, in general, satisfy the sub-additivity property. Intuitively, this means that diversification of risk may be penalized if VaR is used as a risk metric. An alternative risk metric, namely ES, has been proposed in, for example, Artzner et al. (1999), which satisfies the sub-additivity property when the loss distribution is continuous and takes account of the tail risk beyond the VaR threshold. See, for example, Artzner et al. (1999) and Boyle et al. (2002) for related discussions.

Again, one-step-ahead forecasts are considered for illustration. This does not seem to be an unreasonable consideration for, in practice, risk metrics such as VaR are often evaluated or updated for each period, say one day. Since these estimates of the risk metrics are evaluated based on the Bayesian posterior distributions, they combine both expert opinion and objective market data when evaluating risks, and these risk metrics may be related to Bayesian coherent risk measures in, for example, Siu et al. (2001), and Bayesian Value at Risk in, for example, Siu et al. (2004).

A short position for one unit of a Bitcoin is considered for the computation of the estimates of the VaR and ES. This will be used in the real data applications to be presented in Sect. 6. A short position of Bitcoin was also considered for evaluating VaR and ES in Siu (2021). Firstly, the reference model in Eq. (3.1) is considered. Let X_t be the loss random variable of the short position of the Bitcoin in the t th period. Since (percentage) logarithmic returns may be used to approximate percentage changes in the value of a risky portfolio, the loss random variable X_t is equal to the logarithmic return in the t th-period, where X_t is expressed as percentage. See, for example, Tsay (2013), Page 337.

Let $VaR_{1-p}^\theta[X_{n+1}|\mathcal{F}_n^X]$ and $ES_{1-p}^\theta[X_{n+1}|\mathcal{F}_n^X]$ be the VaR and ES of the short position of the Bitcoin in the $(n + 1)^{st}$ -period given \mathcal{F}_n^X with confidence level $1 - p$ evaluated under the probability measure \mathbb{P}^θ , respectively. For example, p could be 0.05. Then applying Eq. (7.2) on Page 332 and Eq. (7.6) on Page 335 in Tsay (2013) and using the notation introduced in Online Appendix E and Eq. (3.7), $VaR_{1-p}^\theta[X_{n+1}|\mathcal{F}_n^X]$ and $ES_{1-p}^\theta[X_{n+1}|\mathcal{F}_n^X]$ have the following analytical expressions:

$$VaR_{1-p}^\theta[X_{n+1}|\mathcal{F}_n^X] = \hat{f} + \mu \hat{g} + \sigma \hat{g} z_{1-p}, \tag{5.1}$$

and

$$ES_{1-p}^\theta[X_{n+1}|\mathcal{F}_n^X] = \hat{f} + \mu\hat{g} + \sigma\hat{g}\frac{\phi(z_{1-p})}{p}, \tag{5.2}$$

where z_{1-p} and $\phi(\cdot)$ denote the $(1 - p)$ th quantile and the pdf of a standard normal distribution $N(0, 1)$, respectively.

The upper-point estimates for the VaR in Eq. (5.1) and the ES in Eq. (5.2) with respect to $\{\mathbb{P}^\theta | \theta \in \Theta_{2\gamma_1, 2\gamma_2}(\mathbf{e}(n))\}$ are, respectively, given by:

$$UVaR_{1-p, \Theta_{2\gamma_1, 2\gamma_2}(\mathbf{e}(n))}[X_{n+1}|\mathcal{F}_n^X] = \text{ess} - \sup_{\theta \in \Theta_{2\gamma_1, 2\gamma_2}(\mathbf{e}(n))} VaR_{1-p}^\theta[X_{n+1}|\mathcal{F}_n^X], \tag{5.3}$$

and

$$UES_{1-p, \Theta_{2\gamma_1, 2\gamma_2}(\mathbf{e}(n))}[X_{n+1}|\mathcal{F}_n^X] = \text{ess} - \sup_{\theta \in \Theta_{2\gamma_1, 2\gamma_2}(\mathbf{e}(n))} ES_{1-p}^\theta[X_{n+1}|\mathcal{F}_n^X]. \tag{5.4}$$

The lower-point estimates for the VaR in Eq. (5.1) and the ES in Eq. (5.2) with respect to $\{\mathbb{P}^\theta | \theta \in \Theta_{2\gamma_1, 2\gamma_2}(\mathbf{e}(n))\}$, denoted by $LVaR_{1-p, \Theta_{2\gamma_1, 2\gamma_2}(\mathbf{e}(n))}[X_{n+1}|\mathcal{F}_n^X]$ and $LES_{1-p, \Theta_{2\gamma_1, 2\gamma_2}(\mathbf{e}(n))}[X_{n+1}|\mathcal{F}_n^X]$, are, respectively, defined by replacing “ess – sup” in Eq. (5.3) and Eq. (5.4) with the “ess – inf”.

An interval estimate for the VaR in Eq. (5.1) with respect to $\{\mathbb{P}^\theta | \theta \in \Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))\}$ is given by:

$$(LVaR_{1-p, \Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))}[X_{n+1}|\mathcal{F}_n^X], UVaR_{1-p, \Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))}[X_{n+1}|\mathcal{F}_n^X]). \tag{5.5}$$

Similarly, an interval estimate for the ES in Eq. (5.2) with respect to $\{\mathbb{P}^\theta | \theta \in \Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))\}$ is evaluated. The Bayesian risk-neutral estimate for the VaR is given by:

$$BRNVaR_{1-p}[X_{n+1}|\mathcal{F}_n^X] = \hat{f} + \mu_n\hat{g} + \sigma_n\hat{g}z_{1-p}, \tag{5.6}$$

where μ_n and σ_n are given by Eq. (4.6) and Eq. (4.7), respectively. Similarly, the Bayesian risk-neutral estimate for the ES is evaluated.

Now the reference model in Eq. (3.10) is considered. Let $VaR_{1-p}^{\theta_h}[X_{n+1}|\mathcal{F}_n^X]$ and $ES_{1-p}^{\theta_h}[X_{n+1}|\mathcal{F}_n^X]$ be the VaR and ES of the short position of the Bitcoin in the $(n + 1)^{st}$ -period given \mathcal{F}_n^X with confidence level $1 - p$ evaluated under the probability measure \mathbb{P}^{θ_h} , respectively. Using the notation in Online Appendix E and Eq. (B.7) in Online Appendix B, $VaR_{1-p}^{\theta_h}[X_{n+1}|\mathcal{F}_n^X]$ and $ES_{1-p}^{\theta_h}[X_{n+1}|\mathcal{F}_n^X]$ have the following analytical expressions:

$$VaR_{1-p}^{\theta_h}[X_{n+1}|\mathcal{F}_n^X] = \hat{f} + \mu\hat{h}(\mu_h) + \sigma\hat{h}(\mu_h)z_{1-p}, \tag{5.7}$$

and

$$ES_{1-p}^{\theta_h}[X_{n+1}|\mathcal{F}_n^X] = \hat{f} + \mu\hat{h}(\mu_h) + \sigma\hat{h}(\mu_h)\frac{\phi(z_{1-p})}{p}, \tag{5.8}$$

where $\hat{h}(\mu_h) := \hat{h}(X_n, X_{n-1}, \dots, X_{n+1-q}, \hat{V}_{n+1}(\mu_h))$ as defined in Online Appendix E; $\hat{V}_{n+1}(\mu_h)$ is given by Eq. (E.12) in Online Appendix E.

When the reference model is the first-order SV model in Eq. (3.12), $VaR_{1-p}^{\theta_h}[X_{n+1}|\mathcal{F}_n^X]$ in Eq. (5.7) and $ES_{1-p}^{\theta_h}[X_{n+1}|\mathcal{F}_n^X]$ Eq. (5.8) become:

$$VaR_{1-p}^{\theta_h}[X_{n+1}|\mathcal{F}_n^X] = \mu \exp(\hat{V}_{n+1}(\mu_h)/2) + \sigma \exp(\hat{V}_{n+1}(\mu_h)/2)z_{1-p}, \tag{5.9}$$

and

$$ES_{1-p}^{\theta_h}[X_{n+1}|\mathcal{F}_n^X] = \mu \exp(\hat{V}_{n+1}(\mu_h)/2) + \sigma \exp(\hat{V}_{n+1}(\mu_h)/2)\frac{\phi(z_{1-p})}{p}, \tag{5.10}$$

where $\hat{V}_{n+1}(\mu_h)$ is given by Eq. (E.12) in Online Appendix E.

The upper-point estimates, the lower-point estimates and the interval estimates for the VaR and ES under the reference model in Eq. (3.10) are then evaluated in the same fashion as those under the reference model in Eq. (3.1) as described above, with θ replaced by θ_h , $\Theta_{2\gamma_1, 2\gamma_2}(\mathbf{e}(n))$ replaced by $\Theta_{2\gamma_1, 2\gamma_2, 2\gamma_{h1}, 2\gamma_{h2}}(\mathbf{e}(n), \mathbf{v}(n))$ and $\Theta_{\gamma_1, \gamma_2}(\mathbf{e}(n))$ replaced by $\Theta_{\gamma_1, \gamma_2, \gamma_{h1}, \gamma_{h2}}(\mathbf{e}(n), \mathbf{v}(n))$.

The Bayesian risk-neutral estimate for the VaR is given by:

$$BRNVaR_{1-p}^h[X_{n+1}|\mathcal{F}_n^X] = \hat{f} + \mu_n\hat{h}(\mu_{hn}) + \sigma_n\hat{h}(\mu_{hn})z_{1-p}, \tag{5.11}$$

where μ_n , μ_{hn} and σ_n are given by Eq. (4.6), Eq. (C.8) in Online Appendix C and Eq. (4.7), respectively. Under the first-order SV model in Eq. (3.12), the Bayesian risk-neutral estimate in Eq. (5.11) becomes:

$$BRNVaR_{1-p}^h[X_{n+1}|\mathcal{F}_n^X] = \mu_n \exp(\hat{V}_{n+1}(\mu_{hn})/2) + \sigma_n \exp(\hat{V}_{n+1}(\mu_{hn})/2)z_{1-p}. \tag{5.12}$$

The Bayesian risk-neutral estimates for the ES are evaluated similarly.

Note that the estimates for the VaR and ES described as above represent losses in percentage. As in Tsay (2013), Page 337, the estimates of the VaR and ES in dollar amounts can be approximated by first multiplying the estimates for the VaR and ES described as above with the current dollar amount of one unit of the Bitcoin and then dividing the multiplication by 100. This method will be used to compute the estimates for the VaR and ES of the short position of the Bitcoin in dollar amounts in Sect. 6.

Though the focus of the current paper is parametric nonlinear time series modelling, it may be noted that the proposed two-stage approach may be extended to nonparametric time series modelling. In Online Appendix F, some preliminary discussions on some potential extensions along this line are provided at an intuitive level.

6 Applications to forecasting and risk evaluation of Bitcoin

In this section, applications of the proposed two-stage approach for estimation and prediction to forecasting and risk evaluation of Bitcoin are provided using real Bitcoin data including some periods of Covid 19. Three major parametric nonlinear time series models, namely the SETAR model, the GARCH model and the SV model, are used as the reference models in the first stage and are estimated from the Bitcoin data using the methods described in Sect. 5.1. Then the one-step-ahead point and interval forecasts for the Bitcoin return based on the Bayesian nonlinear expectations are computed using the formulas in Online Appendix E. Lastly, one-step-ahead point and interval predictions of VaR and ES based on the Bayesian credible intervals are computed using the formulas in Sect. 5.3. When nonlinear time series models are considered, one may use numerical simulations for computing multiple-step-ahead forecasts for the return as well as multiple-step-ahead predictions of VaR and ES. The formulas presented in Online Appendix E and Sect. 5.3 provide a convenient way to numerically compute the one-step-ahead forecasts and predictions. To provide an illustration of the intuition and basic idea of the proposed two-stage approach, one-step-ahead forecasts and predictions are considered here for comparing the three models. As noted in Sect. 5.3, risk metrics such as VaR are often computed for each period, say one day, in practice. To study the impacts of the prior parameters (hyperparameters) for the “uncertain” parameters for misspecifications in the drift and volatility of the reference models on forecasting and risk evaluation of Bitcoin, three sets of prior parameters are used. The computations were done using the statistical package R.

Bitcoin is a digital currency, or cryptocurrency. Cryptocurrencies, together with a related technology called Blockchain, seem to be thought of as important innovations in the FinTech world. The term “cryptocurrency” appears to be related to an important branch in mathematics for coding, namely cryptography, which may have been used in mining Bitcoins. Recently, there have been growing in trading activities of Bitcoins around the world. The growth has accelerated in some periods of Covid 19. Nevertheless, it appears that trading Bitcoins may be risky. Empirical studies reveal that the volatility level of Bitcoins is high. See, for example, Baur and Dimpfl (2021) and the relevant literature therein. Consequently, risk measurement and management for trading Bitcoins could be practically relevant issues. Some recent studies employed GARCH-type models to describe the daily volatility of a Bitcoin series. See, for example, Bouoiyour and Selmi (2016), Dyhrberg (2016), Bouri et al. (2017), Katsiampa (2017) and Baur and Dimpfl (2021). The uses of the GARCH-type models and some of their variants to study risk metrics, such as VaR and ES, for Bitcoins or other cryptocurrencies were considered in the literature. See, for example, Chan et al. (2017), Chu et al. (2017), Osterrieder and Lorenz (2017), Stavroyiannis (2017), Colucci (2018), Aslanidis et al. (2019), Caporale and Zekokh (2019), Trucíos (2019) and Siu (2021). Forecasting and risk evaluation of Bitcoin using Bayesian econometrics were also considered in the literature. See, for example, Hotz-Behofsits et al. (2018), Catania et al. (2019) and Philip et al. (2020). In a recent paper by Siu and Elliott (2021), the pricing of Bitcoin options under a SETAR-GARCH model was considered.

Daily adjusted close prices of Bitcoin in USD from 1 January 2018 to 21 January 2022 with 1,482 observations are used in the empirical study of this paper. The Bitcoin

dataset covers some periods of Covid 19. The data were downloaded from Yahoo Finance. To visualize some patterns of the Bitcoin data, we can look at Fig. 1, which gives the time series plots for the daily adjusted close prices, the daily logarithmic returns in percent, the sample ACF and the sample PACF.

From Fig. 1 (Panel A), it is clear that there were sharp upward trends in the daily adjusted close prices in some periods of Covid 19, (i.e. some periods after the 1000th observation. Figure 1 (Panel B) shows that the volatility of the daily logarithmic returns appears to change over time. From Fig. 1 (Panels C and D), the daily logarithmic returns look stationary.

The summary statistics of the daily percentage logarithmic returns from the Bitcoin, which were computed using “basicStats” in the R package “fBasics” (Wuertz et al. 2020), are displayed in Table 1.

From Table 1, it may be seen that the Bitcoin returns data are negatively skewed, (i.e. skewness -1.141107) and heavy-tailed, (i.e. kurtosis 14.086406). The range of fluctuations in the returns data are large, (i.e. minimum -46.473018 and maximum 17.182056).

For illustration, a SETAR(2,1,1) model with $d = 1$, a GARCH(1,1) model and a first-order SV model are used as three reference models for the daily percentage logarithmic returns of the Bitcoin. The SETAR(2,1,1) model with $d = 1$ is given by:

$$X_t = (\alpha_0^{(1)} + \alpha_1^{(1)} X_{t-1} + \epsilon_t)I_{\{X_{t-1} \leq r\}} + (\alpha_0^{(2)} + \alpha_1^{(2)} X_{t-1} + \epsilon_t)I_{\{X_{t-1} > r\}}, \quad (6.1)$$

where I_E is the indicator function of an event E ; r is the threshold parameter. The SETAR(2,1,1) model in Eq. (6.1) is the reference model in Eq. (3.1) with the conditional mean function $f = (\alpha_0^{(1)} + \alpha_1^{(1)} X_{t-1} + \epsilon_t)I_{\{X_{t-1} \leq r\}} + (\alpha_0^{(2)} + \alpha_1^{(2)} X_{t-1} + \epsilon_t)I_{\{X_{t-1} > r\}}$ and the volatility function $g = 1$.

The GARCH(1,1) model is given by:

$$\begin{aligned} X_t &= \sqrt{h_t} \epsilon_t, \\ h_t &= \gamma_0 + \gamma_1 X_{t-1}^2 + \delta h_{t-1}. \end{aligned} \quad (6.2)$$

The GARCH(1,1) model in Eq. (6.2) is the reference model in Eq. (3.1) with the conditional mean function $f = 0$ and the conditional volatility function $g = \sqrt{h_t}$. The first-order SV model is given by Eq. (3.12), which is included the reference model in Eq. (3.10), as described in Sect. 3.2.

The estimation results for the SETAR(2,1,1) model, the GARCH(1,1) model and the first-order SV model are presented in Tables 2, 3 and 4, respectively. The estimation of the SETAR(2,1,1) model was done by the function “tar” in the R package “TSA” (Chan and Ripley 2020). The testing of the SETAR(2,1,1) model based on the Threshold nonlinearity test in Tsay (1989) was done using the function “thr.test” in the R package “NTS” (Tsay et al. 2020). The estimation and testing of the GARCH(1,1) model were done using the function “garch” in the R package “tseries” (Trapletti et al. 2021), where the testing of the GARCH(1,1) model is based on the Box–Ljung test for the squared residuals. The R codes were written to estimate the first-order SV model using the QMLE with the Kalman filter (see Appendix B of Jacquier et al. (2002)). The R

function “nlminb” was used in the optimization of the QMLE with the initial values of the parameters: $\alpha_v = 0.05$, $\delta_v = 0.9$ and $\sigma_v = 0.2$.

From Table 2, the result of the Threshold nonlinearity test provides evidence to support the SETAR(2,1,1) model with $d = 1$, (i.e. p -value = 0.01489898), at 5% significance level. From Table 3, the result of the Box–Ljung test for the squared residuals indicates the presence of the ARCH effect, (i.e. p -value = 0.6965). Comparing the log-likelihood values of the three models, the first-order SV model provides the best fit to the Bitcoin data, which gives the largest log-likelihood value of -3443.831 .

Now three sets of prior parameters (hyperparameters) are considered. The first set of prior parameters (Set 1) is: $\mu_0 = 0.05$, $t_0 = 100$, $\mu_{h0} = 0.05$, $t_{h0} = 100$, $\alpha = 2$, $\beta = 1$, $\alpha_h = 2$ and $\beta_h = 1$. The second set of prior parameters (Set 2) is: $\mu_0 = 0$, $t_0 = 10$, $\mu_{h0} = 0.05$, $t_{h0} = 100$, $\alpha = 1$, $\beta = 1$, $\alpha_h = 2$ and $\beta_h = 1$. Under Set 2, the prior means for μ and σ are 0 and 1, respectively. This indicates that according to the prior estimates given by the prior means for μ and σ , there are no misspecifications in the conditional mean and volatility of a reference model given by either the SETAR(2,1,1) model in Eq. (6.1) or the GARCH(1,1) model in Eq. (6.2), and there are no misspecifications in the conditional mean and volatility of the observation process in a reference model given by the first-order SV model in Eq. (3.12). According to the prior precisions for μ and σ , there is less (higher) confidence on the prior mean for μ (σ) under Set 2 than Set 1. The third set of prior parameters (Set 3) is: $\mu_0 = 0.5$, $t_0 = 100$, $\mu_{h0} = 0.5$, $t_{h0} = 100$, $\alpha = 1$, $\beta = 2$, $\alpha_h = 1$ and $\beta_h = 1$. Under Set 3, the prior mean for σ_h is 1. This indicates that according to the prior estimate given by the prior mean for σ_h , there is no misspecification in the volatility of the latent process in a reference model given by the first-order SV model in Eq. (3.12). According to the prior estimates given by the prior means for μ and μ_h , there are higher levels of misspecifications in the conditional means μ and μ_h of the respective reference models under Set 3 than Set 1. According to the prior precisions for σ and σ_h , there is higher confidence on the prior means for σ and σ_h under Set 3 than Set 1.

Table 5 presents the 95% Bayesian credible intervals (BCIs) and the Bayesian risk-neutral estimates (BRNEs) for the “uncertain” parameters (μ , σ) of the SETAR(2,1,1) model in Eq. (6.1) and the GARCH(1,1) model in Eq. (6.2) under the prior parameters in Set 1. Table 6 gives the 95% BCIs and the BRNEs for the “uncertain” parameters (μ , σ , μ_h , σ_h) of the SV model in Eq. (3.12) under the prior parameters in Set 1. All the BCIs and BRNEs are expressed as percentage.

From Table 5, according to the BRNEs, the misspecification in the volatility σ under the GARCH(1,1) model is less severe than the SETAR(2,1,1) model. Also, the misspecification in the conditional mean μ under the SETAR(2,1,1) model is less severe than the GARCH(1,1) model. These results are consistent with intuition because the SETAR(2,1,1) model incorporates nonlinearity such as regime switching in the conditional mean while the GARCH(1,1) model captures changing volatility, say the conditional heteroscedasticity. According to the BCIs, the interval estimates for the misspecifications in the volatilities of the GARCH(1,1) model and the SETAR(2,1,1) model are precise. However, the interval estimate for the misspecifications in the conditional means of the SETAR(2, 1, 1) model and the GARCH(1,1) model are not precise. In view of the Bitcoin data, there were sharp upward trends in the daily adjusted close

prices in some periods of Covid 19 as depicted in the time series plot in Fig. 1 (Panel A). Consequently, modelling and predicting the conditional mean of the Bitcoin return data become challenging. It is difficult to give a precise estimation for the misspecification in the conditional mean, which is reflected in the wider interval estimates for the misspecifications in the conditional means by the BCIs. This appears to be in line with Merton (1980), where it was noted that the drift of financial returns was uneasy to be precisely estimated. From Table 6, according to the BRNEs, the misspecifications in the conditional mean and volatility of the observation process under the first-order SV model are slightly more severe than the GARCH(1,1) model. However, the misspecifications on the conditional mean and volatility of the observation process are more severe than the latent process under the SV model. This is consistent with the intuition because it may be more difficult to specify a latent process, which is unobservable, correctly than an observable process. Furthermore, according to the BCIs, the interval estimates for the misspecifications in the conditional means of both the observation and latent processes are less precise than the interval estimates for the misspecifications in the volatility of the two processes under the SV model. This is consistent with the respective results for the SETAR(2,1,1) model and the GARCH(1,1) model.

Table 7 gives the one-step-ahead forecasts, say an upper-point forecast (UPF), a lower-point forecast (LPF), an interval forecast (IF), a Bayesian risk-neutral forecast (BRNF) and a classical forecast (CF) based on a reference model only, of the daily percentage logarithmic returns from the SETAR(2,1,1) model, the GARCH(1,1) model and the first-order SV model estimated from the Bitcoin data under the prior parameters in Set 1. All the forecasts using the Bayesian nonlinear expectations, namely UPF, LPF and IF, were evaluated at the 95% probability level for the “uncertain” parameters $(\mu, \sigma, \mu_h, \sigma_h)$. The one-step-ahead forecasts were expressed as percentage and were computed based on the Bitcoin data in the whole sampling period from 1 January 2018 to 21 January 2022 with 1,482 observations.

From Table 7, the one-step-ahead forecasts from the GARCH(1,1) model and the SV model are generally lower than those from the SETAR(2,1,1) model. This is consistent with the assumption that the conditional means of the reference models given by the GARCH(1,1) model and the SV model are zero. From the LPF, UPF and IF from the GARCH(1,1) model and the SV model, it may be seen that the drift and volatility misspecifications have a more pronounced impact on the one-step-ahead forecasts from the SV model than those from the GARCH(1,1) model. This is also evidenced by comparing the BRNF and CF from each of the GARCH(1,1) model and the SV model. To explain this, it may be noted that the drift and volatility misspecifications are incorporated in both the observation and latent processes under the SV model, while the drift and volatility misspecifications are captured in the observation process only under the GARCH(1,1) model. Lastly, the difference between the BRNE and the CF under the SETAR(2,1,1) model is small, while the difference between the UPF and the LPF under the SETAR(2,1,1) model is large. This is consistent with the results in Table 5 that the BRNE for μ is small and the 95% BCI for μ is wide under the SETAR(2,1,1) model.

Table 8 gives the one-step-ahead point and interval predictions for the two risk metrics, namely VaR and ES, for the short position of one unit of the Bitcoin from the SETAR(2,1,1) model, the GARCH(1,1) model and the SV model estimated from

the Bitcoin data under the prior parameters in Set 1. Specifically, an upper-point VaR prediction (UVaR), an upper-point ES prediction (UES), a lower-point VaR prediction (LVaR), a lower-point ES prediction (LES), an interval VaR prediction (IVaR), an interval ES prediction (IES), a Bayesian risk-neutral VaR prediction (BRNVaR), a Bayesian risk-neutral ES prediction (BRNES), a classical VaR prediction (VaR) and a classical ES prediction (ES) based on the respective reference model only are presented. All the estimates of the two risk metrics based on the Bayesian nonlinear expectations, namely the UVaR, UES, LVaR, LES, IVaR and IES, were evaluated at the 95% probability level for the “uncertain” parameters $(\mu, \sigma, \mu_h, \sigma_h)$. The confidence levels for the VaR and ES are both equal to 95%, (i.e. $p = 0.05$). All the estimates of the two risk metrics are expressed in USD. The one-step-ahead VaR and ES predictions were computed based on the Bitcoin data in the whole sampling period from 1 January 2018 to 21 January 2022 with 1,482 observations. The adjusted close price of one unit of Bitcoin in USD on 21 January 2022 was \$38889.68. To numerically compute the UVaR, UES, LVaR, LES, IVaR and IES, a set of 11 grid points is used to equally partition the 95% BCI for each of the “uncertain” parameters $(\mu, \sigma, \mu_h, \sigma_h)$.

From Table 8, with the drift and volatility misspecifications incorporated, the SETAR(2,1,1) model gives the most conservative one-step-ahead VaR and ES predictions among the three models, while the GARCH(1,1) model gives the least conservative one-step-ahead VaR and ES predictions. However, without incorporating the drift and volatility misspecifications, the SV model gives the most conservative one-step-ahead VaR and ES predictions among the three models, while the SETAR(2,1,1) model gives the least conservative one-step-ahead VaR and ES predictions. Unlike the one-step-ahead forecasts for the Bitcoin return, where the conditional mean of the observation process is the key factor, the volatility of the observation process is the key factor for the one-step-ahead VaR and ES predictions. From Tables 5-6, the SETAR(2,1,1) model has the most severe misspecifications in the volatility of the observation process among the three model, while the GARCH(1,1) model has the least severe misspecification in the volatility of the observation process. Consequently, the results from Table 8 indicate that the more severe the misspecification in the volatility of the observation process is, the more conservative the one-step-ahead risk prediction is.

The results based on different sets of prior parameters (hyperparameters), say Set 2 and Set 3, are presented in Online Appendix G. It may be seen that the results under the prior parameters in Set 2 and Set 3 are qualitatively similar to those under the prior parameters in Set 1. This indicates that the observations and conclusions from the results under the prior parameters in Set 1 may be quite robust.

7 Conclusions

A two-stage approach was proposed to parametric time series modelling with the drift and volatility misspecifications or uncertainties. A reference model is specified and estimated in the first stage. At the second stage, the Bayesian nonlinear expectations were used to incorporate a family of alternative models for prediction, where the family alternative models were introduced using discrete-time Girsanov’s transforms.

The idea is to make use of the information that is left over in the residuals of the estimated reference model to construct Bayesian credible intervals for the “uncertain” parameters describing the drift and volatility misspecifications. Using conjugate priors, closed-form Bayesian credible intervals are obtained, which provide a convenient way to construct the Bayesian nonlinear expectations used for prediction so that probability levels are assigned to the nonlinear expectations. The closed-form Bayesian credible intervals combine both a modeller’s prior/subjective belief (or expert opinion) and the objective data in estimating the drift and volatility misspecifications. The prior/subjective belief may be useful to make adjustments for prediction in case the future developments may not follow the current and past data. To introduce the two-stage approach, two general classes of parametric nonlinear time series models, namely a nonlinear autoregressive model with conditional heteroscedasticity and a product process, are considered. The former includes the SETAR model and the GARCH model, while the latter includes the SV model. The link between the Bayesian shrinkage and regularization techniques for estimation and variable selection and the Bayesian nonlinear expectations was discussed in Online Appendix D. Some potential generalizations based on both classical and Bayesian nonparametric approaches were outlined in Online Appendix F. Applications of the proposed two-stage approach to forecasting and risk evaluation of Bitcoin were illustrated empirically using real Bitcoin data covering some periods of Covid 19 and the three models, namely the SETAR model, the GARCH model and the SV model. The results highlight the importance of incorporating misspecifications in the conditional mean and volatility for making one-step-ahead forecasts for the Bitcoin return and one-step-ahead VaR and ES predictions for the short position of the Bitcoin, respectively.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00181-022-02255-z>.

Acknowledgements I would like to sincerely thank an Associate Editor and the reviewers for insightful and helpful comments. The author wishes to acknowledge the support from the Discovery Grant from the Australian Research Council (ARC), (Project Number: DP190102674).

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. Discovery Grant from the Australian Research Council (ARC), (Project Number: DP190102674)

Declarations

Conflict of interest Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Appendix

See Tables 1, 2, 3, 4, 5, 6, 7 and 8.

Table 1 Summary statistics of the daily percentage log returns

nobs	Mean	Stdev	Skewness	Kurtosis
1481	0.070659	4.005866	-1.141107	14.086406
Minimum	1 Quartile	Median	3 Quartile	Maximum
-46.473018	-1.572926	0.136811	1.836281	17.182056

Table 2 Estimation and testing results of the SETAR(2,1,1) model with ($d = 1$)

$\alpha_0^{(1)}$	$\alpha_1^{(1)}$	$\alpha_0^{(2)}$	$\alpha_1^{(2)}$
-0.2141628	-0.1214365	0.01773623	0
r	log-like	F -ratio	p -value
0.02149885	-4141.243	4.218792	0.01489898

Table 3 Estimation and testing results of the GARCH(1,1) model

γ_0	γ_1	δ	log-like	X -squared	p -value
0.97348	0.08517	0.85911	-4087.579	0.15217	0.6965

Table 4 Estimation results of the SV model

α_v	δ_v	σ_v	log-like
0.05157367	0.97363784	0.23617299	-3443.831

Table 5 95% BCIs and BRNEs for μ and σ (Set 1 priors)

	95% BCI (SETAR)	BRNE (SETAR)	95% BCI (GARCH)	BRNE (GARCH)
μ	(-0.1617407, 0.1680658)	0.003162555	(-0.02291122, 0.05989249)	0.01849063
σ	(3.867361, 4.108072)	3.983795	(0.9709686, 1.031403)	1.000201

Table 6 95% BCIs and BRNEs for μ , σ , μ_h and σ_h (Set 1 priors)

	95% BCI (SV)	BRNE (SV)
μ	(-0.04263042, 0.08687772)	0.02212365
σ	(1.518632, 1.613154)	1.564353
μ_h	(0.001556987, 0.004768124)	0.003162555
σ_h	(0.03765428, 0.03999794)	0.03878793

Table 7 One-step-ahead forecasts (Set 1 priors)

	SETAR	GARCH	SV
LPF	0.1707797	-0.03527248	-0.09918374
UPF	0.5005863	0.1007441	0.2022059
IF	(0.139151, 0.532215)	(-0.04831657, 0.1137882)	(-0.1280752, 0.2311215)
BRNF	0.335683	0.03273581	0.05148254
CF	0.3325205	0	0

Table 8 Estimates of VaR and ES (Set 1 priors)

	SETAR	GARCH	SV
UVaR	2822.521	1123.269	2480.36
UES	3490.099	1398.671	3090.495
LVaR	2540.282	1006.871	2221.565
LES	3168.744	1266.142	2795.732
IVaR	(2514.025, 2850.405)	(996.0422, 1134.768)	(2197.497, 2505.936)
IES	(3138.941, 3521.943)	(1253.851, 1411.803)	(2768.403, 3119.712)
BRNVaR	2678.893	1064.036	2348.646
BRNES	3326.276	1331.109	2940.211
VaR	768.9944	1049.765	1487.999
ES	931.4985	1316.447	1866.011

B Appendix

See Fig. 1.

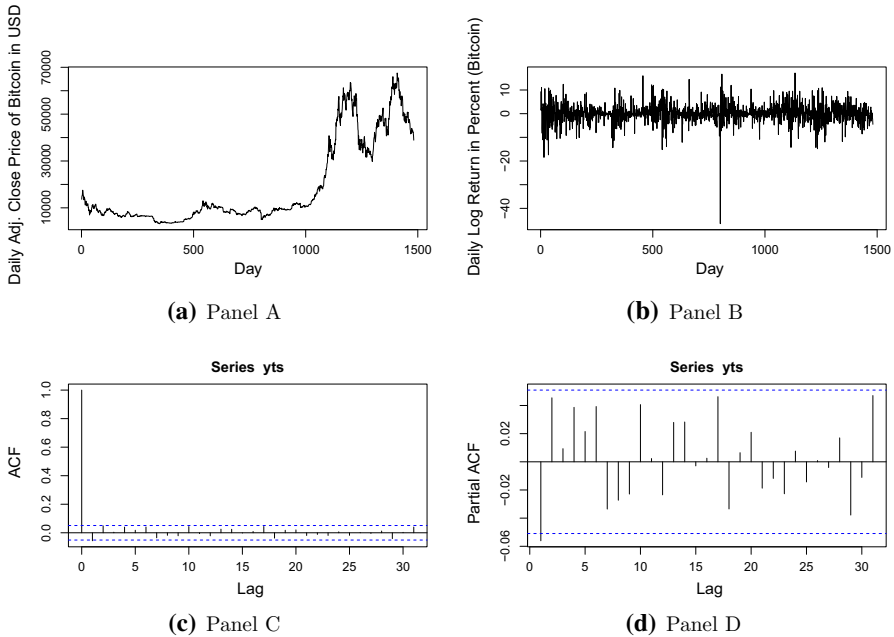


Fig. 1 Panel A plots the time series plot of the daily adj. close prices of Bitcoin against US dollars. Panel B depicts the time series plot of the daily percentage logarithmic returns of Bitcoin. Panels C and D plot the sample ACF and sample PACF of the daily percentage logarithmic returns of Bitcoin, respectively

References

- Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent measures of risk. *Math Finance* 9(3):203–228
- Aslanidis N, Bariviera AF, Martinez-Ibañez O (2019) An analysis of cryptocurrencies conditional cross correlations. *Finance Res Lett* 31:130–137. <https://doi.org/10.1016/J.FRL.2019.04.019>
- Avellaneda M, Levy A, Par AS (1995) A Pricing and hedging derivative securities in markets with uncertain volatilities. *Appl Math Finance* 2(2):73–88
- Baur DG, Dimpfl T (2021) The volatility of Bitcoin and its role as a medium of exchange and a store of value. *Empir Econ* 61:2663–2683. <https://doi.org/10.1007/s00181-020-01990-5>
- Bernardo JM, Smith AMF (2000) Bayesian theory. Wiley, New York
- Bhadra A, Datta J, Polson NG, Willard B (2019) Lasso meets horseshoe: a survey. *Stat Sci* 34(3):405–427
- Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *J Econom* 31(3):307–327
- Bouoiyour J, Selmi R (2016) Bitcoin: a beginning of a new phase? *Econ Bull* 36:1–11
- Bouri E, Azzi G, Dyhrberg AH (2017) On the return-volatility relationship in the Bitcoin market around the price crash of 2013. *Econ Open Access Open Assess E-J* 11:1–16
- Box GEP, Tiao GC (1992) Bayesian inference in statistical analysis. Wiley, New York
- Boyle PP, Ananthanarayanan A (1977) The impact of variance estimation in option valuation models. *J Financ Econ* 5:375–387
- Boyle PP, Siu TK, Yang H (2002) Risk and probability measures. *Risk* 15(7):53–57
- Cagetti M, Hansen LP, Sargent T, Williams N (2002) Robustness and pricing with uncertain growth. *Rev Financ Stud* 15(2):363–404
- Caporale GM, Zekokh T (2019) Modelling volatility of cryptocurrencies using Markov-switching GARCH models. *Res Int Bus Financ* 48:143–155
- Carvalho CM, Polson NG, Scott JG (2009) Handling sparsity via the horseshoe. *J Mach Learn Res* 5:73–80

- Catania L, Grassi S, Ravazzolo F (2019) Forecasting cryptocurrencies under model and parameter instability. *Int J Forecast* 35(2):485–501
- Chan KS, Ripley B (2020) Time series analysis. R package ‘TSA’. Version 1.3. Date 2020-9-11
- Chan S, Chu J, Nadarajah S, Osterrieder J (2017) A statistical analysis of cryptocurrencies. *J Risk Financ Manag* 10(2):12. <https://doi.org/10.3390/jrfm10020012>
- Chen Z, Epstein L (2002) Ambiguity, risk, and asset returns in continuous time. *Econometrica* 70(4):1403–1443
- Cherny A, Madan DB (2009) New measures for performance evaluation. *Rev Financ Stud* 22(7):2571–2606
- Chu J, Chan S, Nadarajah S, Osterrieder J (2017) GARCH modelling of cryptocurrencies. *J Risk Financ Manag* 10(4):17. <https://doi.org/10.3390/jrfm10040017>
- Colucci S (2018) On estimating Bitcoin value at risk: a comparative analysis (August 22, 2018). SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3236813
- Dyrhberg AH (2016) Bitcoin, gold and the dollar—a GARCH volatility analysis. *Finance Res Lett* 16:85–92
- Elliott RJ (2017) Filtering with uncertain noise. *IEEE Trans Autom Control* 62(2):876–881
- Elliott RJ, Madan DB (1998) A discrete time equivalent martingale measure. *Math Finance* 8(2):127–152
- Elliott RJ, Siu TK (2017) Stochastic volatility with regime switching and uncertain noise: filtering with sub-linear expectations. *Discrete Contin Dyn Syst Ser B* 22(1):59–81
- Elliott RJ, Aggoun L, Moore JB (1995) Hidden Markov models: estimation and control, 1st edn. Springer, New York
- Engle RF (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50:987–1008
- Fouque J-P, Ren B (2014) Approximation for option prices under uncertain volatility. *SIAM J Financ Math* 5(1):360–383
- Frazier DT, Robert CP, Rousseau J (2020) Model misspecification in approximate Bayesian computation: consequences and diagnostics. *J R Stat Soc Ser B* 82(2):421–444
- Granger CW, Anderson AP (1978) An introduction to bilinear time series model. Vandenhoeck and Ruprecht, Göttingen
- Hamilton JD (1989) A new approach to economic analysis of nonstationary time series and the business cycle. *Econometrica* 57:357–384
- Hansen LP (2013) Prize lecture: uncertainty outside and inside economic models. Nobelprize.org. Nobel Media AB 2014. <https://www.nobelprize.org/prizes/economic-sciences/2013/hansen/lecture/>
- Hansen LP, Sargent TJ (2007) Robustness. Princeton University Press, Princeton
- Hansen LP, Sargent TJ (2021) Macroeconomic uncertainty prices when beliefs are tenuous. *J Econom* 223(1):222–250
- Hansen LP, Sargent TJ, Wang NE (2002) Robust permanent income and pricing with filtering. *Macroecon Dyn* 6(1):40–84
- Hotz-Behofsits C, Huber F, Zörner TO (2018) Predicting crypto-currencies using sparse non-Gaussian state space models. *J Forecast* 37(6):627–640
- Jacquier E, Polson NG, Rossi PE (2002) Bayesian analysis of stochastic volatility models. *J Bus Econ Stat* 20(1):69–87
- Katsiampa P (2017) Volatility estimation for Bitcoin: a comparison of GARCH models. *Econ Lett* 158:3–6
- Kim S, Shephard N, Chib S (1998) Stochastic volatility: likelihood inference and comparison with ARCH Models. *Rev Econ Stud* 65(3):361–393
- Klibanoff P, Marinacci M, Mukerji S (2005) A smooth model of decision making under ambiguity. *Econometrica* 73(6):1849–1892
- Knight FH (1921) Risk, uncertainty, and profit. Houghton, Mifflin
- Koop G (2003) Bayesian econometrics. Wiley, New York
- Lee PM (2012) Bayesian statistics: an introduction, 4th edn. Wiley, New York
- Lyons T (1995) Uncertain volatility and the risk-free synthesis of derivatives. *Appl Math Finance* 2(2):117–133
- Madan DB, Cherny A (2010) Markets as a counterparty: an introduction to conic finance. *Int J Theor Appl Finance* 13(8):1149–1177
- Melino A, Turnbull SM (1990) Pricing foreign currency options with stochastic volatility. *J Econom* 45(1–2):239–265
- Merton RC (1980) On estimating the expected return on the market: an exploratory investigation. *J Financ Econ* 8(4):323–361

- Nakajima J, Omori Y (2009) Leverage, heavy-tails and correlated jumps in stochastic volatility models. *Comput Stat Data Anal* 53:2335–2353
- Omori Y, Chib S, Shephard N, Nakajima J (2007) Stochastic volatility with leverage: fast and efficient likelihood inference. *J Econom* 140(2):425–449
- Osterrieder J, Lorenz J (2017) A statistical risk assessment of Bitcoin and its extreme tail behavior. *Ann Financ Econ* 12(1):1750003
- Peng S (2006) G -expectation, G -Brownian motion and related stochastic calculus of Itô's type. In: Benth FE, Di Nunno G, Lindstrom T, Øksendal B, Zhang T (eds) *The Abel symposium 2005*, Abel symposia 2. Springer, Berlin, pp 541–567
- Peng S (1997) Backward SDE and related g -expectation. In: El Karoui N, Mazliak L (eds) *Backward stochastic differential equations*. Pitman research notes in mathematics series 364. Longman Scientific & Technical, London, pp 141–159
- Peng S (2004) Nonlinear expectations, nonlinear evaluations and risk measures. In: Frittelli M, Runggaldier WJ (eds) *Stochastic methods in finance*. Lecture notes in mathematics 1856. Springer, New York, pp 165–168
- Peng S (2019) *Nonlinear expectations and stochastic calculus under uncertainty with robust CLT and G-Brownian motion*. Springer, New York
- Philip A, Chan JSK, Peiris S (2020) On generalized bivariate student- t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin. *Econom Stat* 16:69–90. <https://doi.org/10.1016/j.ecosta.2018.10.003>
- Pole AM, Smith AFM (1985) A Bayesian analysis of some threshold switching models. *J Econom* 29(1–2):97–119
- Polson NG, Scott JG (2010) Shrink globally, act locally: sparse Bayesian regularization and prediction. *Bayesian Stat* 9:501–538
- Priestley MB (1980) State-dependent models: a general approach to nonlinear time series analysis. *J Time Ser Anal* 1(1):47–71
- Rosazza Gianin E (2006) Risk measures via g -expectations. *Insur Math Econom* 39(1):19–34
- Rosenberg B (1972) The behavior of random variables with nonstationary variance and the distribution of security prices. Reprinted in *Stochastic volatility: selected readings*, Shephard N (ed). Oxford University Press, Oxford
- Shephard N (2005) *Stochastic volatility: selected readings*. Oxford University Press, Oxford
- Siu TK (2001) Risk measures in finance and insurance. Ph.D. Thesis. The University of Hong Kong, Hong Kong
- Siu TK (2021) The risks of cryptocurrencies with long memory in volatility, non-normality and behavioural insights. *Appl Econ* 53(17):1991–2014. <https://doi.org/10.1080/00036846.2020.1854669>
- Siu TK, Elliott RJ (2021) Bitcoin option pricing with a SETAR-GARCH model. *Eur J Finance* 27(6):564–595. <https://doi.org/10.1080/1351847X.2020.1828962>
- Siu TK, Yang H (1999) Subjective risk measures: Bayesian predictive scenarios analysis. *Insur Math Econom* 25:157–169
- Siu TK, Tong H, Yang H (2001) Bayesian risk measures for derivatives via random Esscher transform. *N Am Actuar J* 5(3):78–91
- Siu TK, Tong H, Yang H (2004) On Bayesian value at risk: from linear to nonlinear portfolios. *Asia-Pacific Financ Mark* 11(2):161–184
- Slutsky E (1927) The summation of random causes as the source of cyclic processes. In: *Problems of economic conditions*, vol 3, no 1. The Conjunction Institute, Moscow
- Stavroyiannis S (2017) Value-at-risk and related measures for the Bitcoin. *J Risk Finance* 19(2):127–136
- Taylor SJ (1980) Conjectured models for trends in financial prices, tests and forecasts. *J R Stat Soc Ser A* 143:338–362
- Taylor SJ (1982) Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices, 1961–79. In: Anderson OD (ed) *Time series analysis: theory and practice 1*. North Holland, Amsterdam, pp 203–226
- Taylor SJ (1986) *Modelling financial time series*. Wiley, Chichester
- Taylor SJ (2005) *Asset price dynamics, volatility, and prediction*. Princeton University Press, Princeton
- Tong H (1977) Discussion of a paper by A. J. Lawrance and N. T. Kottegoda. *J R Stat Soc Ser A* 140:34–35
- Tong H (1978) On a threshold model. In: Chen CH (ed) *Pattern recognition and signal processing*. NATO ASI series E: applied sciences, vol 29. Sijthoff & Noordhoff, Rockville, pp 575–586
- Tong H (1983) *Threshold models in nonlinear time series analysis*. Springer, Berlin

- Tong H (1990) *Nonlinear time series: a dynamical system approach*. Oxford University Press, Oxford
- Tong H (2002) Nonlinear time series analysis since 1990: some personal reflections. *Acta Math Appl Sin Engl Ser* 18(2):177–184
- Trapletti A, Hornik K, LeBaron B (2021) *Time series analysis and computational finance*. R package ‘tseries’. Version 0.10-49
- Trucíos C (2019) Forecasting Bitcoin risk measures: a robust approach. *Int J Forecast* 35:836–847
- Tsay RS (1989) Testing and modelling threshold autoregressive processes. *J Am Stat Assoc* 84(405):231–240
- Tsay RS (2013) *An introduction to analysis of financial data with R*. Wiley, Hoboken
- Tsay RS, Chen R, Liu X (2020) *Nonlinear time series analysis*. R package ‘NTS’. Version 1.1.2
- Wuertz D, Setz T, Chalabi Y, Maechler M (2020) *Rmetrics—markets and basic statistics*. R package ‘fBasics’. Version 3042.89.1
- Yule U (1927) On a method of investigating periodicities in disturbed series, with special reference to Wolfer’s sunspot numbers. *Philos Trans R Soc (Lond) Ser A* 226:267–298

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.