

A Bayesian quantile time series model for asset returns

J. E. Griffin and G. Mitrodima

Abstract

We consider jointly modelling a finite collection of quantiles over time. Formal Bayesian inference on quantiles is challenging since we need access to both the quantile function and the likelihood. We propose a flexible Bayesian time-varying transformation model, which allows the likelihood and the quantile function to be directly calculated. We derive conditions for stationarity, discuss suitable priors and describe a Markov chain Monte Carlo algorithm for inference. We illustrate the usefulness of the model for estimation and forecasting on stock, index, and commodity returns.

Keywords: Bayesian nonparametrics; Transformation models; Stationarity; Predictive density

1 Introduction

The modelling of the conditional distribution of asset returns plays an important role in understanding the risk of financial decisions (for example, through measures such as Value at Risk (VaR) or expected shortfall). A large literature has been developed around this problem (see Andersen, Bollerslev, and Diebold 2009 and Andersen et al. 2009 for reviews) which has generally concentrated on estimating the time-varying conditional volatility us-

ing either stochastic volatility or GARCH-type models in order to recover the conditional distribution. These models usually assume that other aspects of the conditional distribution are time-invariant, but empirical evidence shows that there is variation beyond volatility. For example, Hansen (1994) and Harvey and Siddique (1999) studied variation in skewness and kurtosis using time-varying skewed t distributions or non-central t distributions respectively.

Semiparametric methods can also be used to extend models in which specific higher-order moments are allowed to be time-varying. One example is to use quantile based estimators, which are found to be more robust to outliers compared to the empirical moment based estimators of higher moments (see *e.g.* Kim and White 2004). Many authors have used this approach including White, Kim, and Manganelli (2010) who proposed a joint model of conditional quantiles to obtain robust estimates of conditional skewness and kurtosis for the S&P500 daily returns. However, multiple quantiles cannot be satisfactorily estimated using separate application of single quantile models at different probability levels. This is because these estimates do not guarantee monotonicity of the quantile function. Further work has considered combining quantile and volatility estimates to gain a better understanding of the conditional distribution (Taylor 2005, Xiao and Koenker 2009).

In the Bayesian literature, there have been several attempts to allow time-varying conditional distributions of returns. Markov-switching or change point models are natural approaches which allow the parameters of the conditional distribution to change over time (see *e.g.* So, Lam, and Li 1998; Chen, So, and Liu 2011; Bauwens, Dufays, and Rombouts 2014; Song 2014). These approaches usually assume that the parameters of the model only change a few points in time and so can struggle to capture small or short-term changes. Rodriguez and Horst (2008) and Griffin and Steel (2011) introduced time-varying nonparametric priors which allow inference about a time-varying conditional return distribution. These approaches are based on infinite mixture models and so inference can be challenging for long time series.

In this paper, we consider jointly modelling a finite collection of quantiles over time. For-

mal Bayesian inference on quantiles is challenging since we need access to both the quantile function and its inverse (the distribution function). Yu and Moyeed (2001) use the check loss function as a log-likelihood (which is equivalent to an asymmetric Laplace distribution) leading to a posterior distribution for the parameters of the quantile regression. See Cai, Stander, and Davies (2012), Liu (2016) and Gerlach, Chen, and Chan (2011) for some application of this approach in time series analysis. This approach has been criticised since it does not use the correct likelihood for regressions at multiple quantiles. Several authors have discussed this concern, for example Dunson and Taylor (2005) and Tokdar and Kadane (2012), but there has been no work in time series to address this to our knowledge.

Conditional transformation models offer an alternative approach to flexibly model quantiles. In this approach, an underlying statistical model is applied to transformed data. The parameters of the transformation are inferred with the parameters of the underlying model. The intention is to extend the data for which a particular parametric model is suitable. Hothorn, Kneib, and Bühlmann (2014) review previous work in this area and develop an approach to modelling transformations of both response and explanatory variables using basis functions in regression models. A similar idea underlies inversion copulas (see *e.g.* Smith and Ma-neesoonthorn (2018)) for stationary time series. A stationary parametric model is assumed with stationary distribution F_Z . The transformed random variables $U_t = F_Z(Z_t)$ form a copula which inherits the time series structure of Z_t . The copula is estimated by fitting the model for Z_t to pseudo-data $F_Z^{-1}(\hat{F}_Y(Y_1)), \dots, F_Z^{-1}(\hat{F}_Y(Y_T))$ where \hat{F} is the empirical distribution of the time series. This can be seen as a conditional time series model with a time-invariant transformation $G(\cdot) = F_Z^{-1}(\hat{F}_Y(\cdot))$. A drawback of this approach is that the successful modelling of the dynamics of Y_t depends on an appropriate choice of the parametric model for Z_t . In this paper, we will consider models which allow time-variation in the transformation.

The paper is organised in the following way. Section 2 describes our general model for

time-varying quantiles. Section 3 derives conditions under which the model is stationary. Section 4 shows how these models can be “centred” over a parametric model. Section 5 describes the computational methods needed for Bayesian inference. Section 6 illustrates the use of these methods on simulated data and some asset returns. Section 7 concludes the paper. Proofs, further details of the computational methods, and additional results are presented in the appendices.

2 Bayesian Joint Quantile Time Series (B-JQTS) model

Before constructing time series models for quantiles, we build a flexible distribution for which the quantiles can be easily calculated using a parameterised transformation of a parametric distribution. Hothorn, Möst, and Bühlmann (2018) showed that this is an extremely flexible framework since any continuous distribution can be represented as a transformation of a random variable with an arbitrary continuous distribution.

Definition 2.1. *F is a Linearised Transformation (LIT) distribution if $F(y) = F_0(G(y))$, where F_0 is a univariate continuous distribution with median m and quantile function $Q_0(\cdot)$, and $G : \mathbb{R} \rightarrow \mathbb{R}$. The transformation G has parameters $0 = a_0 < a_1 < \dots < a_K = 0.5$, $\theta_1^- > 0, \dots, \theta_K^- > 0$, and $\theta_1^+ > 0, \dots, \theta_K^+ > 0$ and the piecewise linear form*

$$G(x) = \begin{cases} m, & x=m \\ G(x_{i-1}^-) + \frac{1}{\theta_i^-}(x - x_{i-1}^-), & x_i^- \leq x < x_{i-1}^- \\ G(x_{i-1}^+) + \frac{1}{\theta_i^+}(x - x_{i-1}^+), & x_{i-1}^+ \leq x < x_i^+, \end{cases}$$

where $x_0^+ = x_0^- = m$,

$$x_i^- = x_{i-1}^- - \theta_i^- (Q_0(0.5 - a_i) - Q_0(0.5 - a_{i-1})) = \sum_{j=1}^i \theta_j^- (Q_0(0.5 - a_j) - Q_0(0.5 - a_{j-1}))$$

and

$$x_i^+ = x_{i-1}^+ + \theta_i^+ (Q_0(0.5 + a_i) - Q_0(0.5 + a_{i-1})) = \sum_{j=1}^i \theta_j^+ (Q_0(0.5 + a_j) - Q_0(0.5 + a_{j-1})).$$

If Y follows an LIT distribution we write $Y \sim \text{LIT}(\theta^-, \theta^+, F_0, a)$, where $\theta^- = (\theta_1^-, \dots, \theta_K^-)^T$, $\theta^+ = (\theta_1^+, \dots, \theta_K^+)^T$ and $a = (a_0, a_1, \dots, a_K)^T$. It is sometimes convenient to write $\theta = (\theta_K^-, \dots, \theta_1^-, \theta_1^+, \dots, \theta_K^+)^T$.

An LIT distribution with the same quantiles at probability levels $\tau = (0.5 - a_K, \dots, 0.5 - a_1, 0.5 + a_1, \dots, 0.5 + a_K)$ as a continuous distribution F^* is defined by choosing

$$\theta_i^- = \frac{Q^*(0.5 - a_{i+1}) - Q^*(0.5 - a_i)}{Q_0(0.5 - a_{i+1}) - Q_0(0.5 - a_i)} \text{ and } \theta_i^+ = \frac{Q^*(0.5 + a_i) - Q^*(0.5 + a_{i-1})}{Q_0(0.5 + a_i) - Q_0(0.5 + a_{i-1})}.$$

where $Q^*(\tau)$ is the quantile function of F^* . Figure 1 shows plots of θ against τ for different

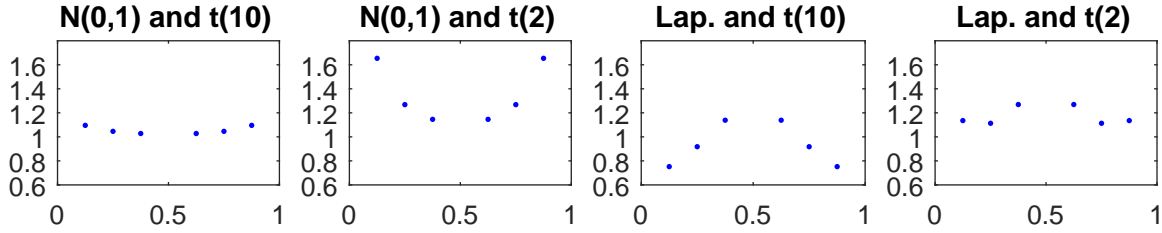


Figure 1: Plots of the value of θ against τ for different F_0 and F^* (F_0 is named first and F^* is named second) and $\tau = (0.125, 0.25, 0.375, 0.625, 0.75, 0.875)$.

choices of F_0 and F^* . If F_0 is a standard normal then a heavier tailed F^* is associated with values of θ_i^- and θ_i^+ that increase with i . If F_0 is a Laplace then the relationship is less straightforward. The values of θ_i^- and θ_i^+ decrease with i when F^* is a t distribution with 10 degrees of freedom, whereas the values of θ_i^- and θ_i^+ do not decrease or increase monotonically with i when F^* is a t distribution with 2 degrees of freedom.

The LIT distribution has some useful properties. Firstly, the density function f of an LIT

distribution is

$$f(x) = \begin{cases} \frac{1}{\theta_i^-} f_0 \left(G(x_{i-1}^-) + \frac{1}{\theta_i^-} (x - x_{i-1}^-) \right), & x_i^- \leq x < x_{i-1}^- \\ \frac{1}{\theta_i^+} f_0 \left(G(x_{i-1}^+) + \frac{1}{\theta_i^+} (x - x_{i-1}^+) \right), & x_{i-1}^+ \leq x < x_i^+, \end{cases} \quad (2.1)$$

where $f_0(x)$ is the density of $F_0(x)$. The density $f(x)$ will have the shape of f_0 with scale parameter θ_i^- for $x_i^- \leq x < x_{i-1}^-$ and θ_i^+ for $x_{i-1}^+ \leq x < x_i^+$. we refer to θ_i^- and θ_i^+ as local scales and F_0 as the centring distribution. Secondly, the quantile function, is available analytically as $Q(\tau) = G^{-1}(Q_0^{-1}(\tau))$ and has the form

$$Q(\tau) = \begin{cases} m, & \tau = 0.5 \\ x_{i-1}^- - \theta_i^- (Q_0(\tau) - Q_0(0.5 + a_{i-1})), & 0.5 - a_i \leq \tau < 0.5 - a_{i-1} \\ x_{i-1}^+ + \theta_i^+ (Q_0(\tau) - Q_0(0.5 + a_{i-1})), & 0.5 + a_{i-1} \leq \tau < 0.5 + a_i. \end{cases} \quad (2.2)$$

Clearly, x_i^- and x_i^+ are the quantiles of F at probability levels $(0.5 - a_i)$ and $(0.5 + a_i)$ respectively and the support points of the transformation G are distributed according to F .

The moments of an LIT distribution and its absolute values can be expressed in terms of the moments of truncated versions of F_0 .

Theorem 2.1. *The ℓ -th moment of an LIT(θ^- , θ^+ , F_0 , a) distribution is*

$$\sum_{i=1}^K \sum_{j=0}^{\ell} \binom{\ell}{j} \left[(\alpha_i^-)^{\ell-j} (\theta_i^-)^j I_{i,j}^- + (\alpha_i^+)^{\ell-j} (\theta_i^+)^j I_{i,j}^+ \right],$$

where $\alpha_i^- = x_{i-1}^- - \theta_i^- Q_0(1/2 - a_{i-1})$, $\alpha_i^+ = x_{i-1}^+ - \theta_i^- Q_0(1/2 + a_{i-1})$, $I_{i,j}^- = \int_{Q_0(1/2 - a_i)}^{Q_0(1/2 - a_{i-1})} z^j f_0(z) dz$ and $I_{i,j}^+ = \int_{Q_0(1/2 + a_{i-1})}^{Q_0(1/2 + a_i)} z^j f_0(z) dz$.

Theorem 2.2. *If $Y \sim \text{LIT}(\theta^-, \theta^+, F_0, a)$ then $E[|y|] = \Phi\theta$, where Φ is a $(1 \times 2K)$ -*

dimensional vector with terms

$$\Phi_i = \begin{cases} -(1/2 - a_i)Q_0(1/2 - a_i) + (1/2 - a_{i-1})Q_0(1/2 - a_{i-1}) - I_{i,1}^-, & 1 \leq i \leq K \\ (1/2 - a_i)Q_0(1/2 + a_i) - (1/2 - a_{i-1})Q_0(1/2 + a_{i-1}) + I_{i,1}^+, & K + 1 \leq i \leq 2K. \end{cases}$$

The LIT distribution construction allows us to build time-varying models for quantiles. We assume that the centring distribution F_0 is parameterised by ψ and define a model with $\theta_1^-, \dots, \theta_K^-, \theta_1^+, \dots, \theta_K^+$ and ψ as time-varying parameters.

Definition 2.2 (B-JQTS model). *Let Y_1, \dots, Y_T be a time series of univariate observations and $0 = a_0 < a_1 < \dots < a_K = 1/2$, the time series follows a Bayesian Joint Quantile Time Series (B-JQTS) model if $Y_t \stackrel{ind.}{\sim} F_t$, where $F_t = LIT(\theta_t^-, \theta_t^+, F_0(\cdot; \psi_t), a)$,*

$$\begin{aligned} \theta_{i,t}^- &= H(\theta_{i,t-N}^-, \dots, \theta_{i,t-1}^-, y_{t-L}, \dots, y_{t-1}; \lambda_i^-), & Q_0(0.5 - a_i) < y < Q_0(0.5 - a_{i-1}), \\ \theta_{i,t}^+ &= H(\theta_{i,t-N}^+, \dots, \theta_{i,t-1}^+, y_{t-L}, \dots, y_{t-1}; \lambda_i^+), & Q_0(0.5 + a_{i-1}) < y < Q_0(0.5 + a_i), \end{aligned}$$

$H(\cdot, \cdot; \lambda)$ is a function parameterised by λ , and N and L are the orders of the model.

The processes θ_t^- and θ_t^+ are specified directly as functions of the previous lags of the parameters and the data (rather than as stochastic processes) to allow faster posterior inference. The model can be interpreted as inducing a time-varying transformation G_t defined using the piecewise linear from in Definition 3.1 and allows departures from the parametric model $F_0(\cdot; \psi_t)$. The B-JQTS model is very general and requires further restrictions for identifiability (for example, fixing the stationary mean of ψ_t).

We first consider models where $\psi_t = 1$ and both $\theta_{i,t}^-$ and $\theta_{i,t}^+$ follow GARCH-like processes. In this case, the model simplifies if $\lambda_1^- = \dots = \lambda_K^- = \lambda_1^+ = \dots = \lambda_K^+ = \lambda$ and $\theta_{1,1}^- = \dots = \theta_{K,1}^- = \theta_{1,1}^+ = \dots = \theta_{K,1}^+ = \theta_1$, to a time series model $Y_t/\theta_t \sim F_0(\cdot; \psi_t)$ and $\theta_t = H(\theta_{t-N}, \dots, \theta_{t-1}, y_{t-L}, \dots, y_{t-1}; \lambda)$. Therefore, the B-JQTS model with $\psi_t = 1$ can be understood as a relaxation of this simpler model, where the values of λ_i^- and λ_i^+ allow

different time series properties for different parts of the return distribution. Some specific examples of B-JQTS models (with $\psi_t = 1$) are given in the following definition.

Definition 2.3. A B-JQTS model is called:

1. *Bayesian Joint Absolute Value Leverage (B-JAVL) model* if

$$H(\theta, y; (\mu, \beta, \gamma, \delta)) = \mu + \sum_{j=1}^N \beta_j \theta_{t-j} + \sum_{k=1}^L \gamma_k |y_{t-k}| + \sum_{k=1}^L I(y_{t-k} < 0) \delta_k |y_{t-k}| \quad (2.3)$$

2. *Bayesian Joint GJR (B-JGJR) model* if

$$H(\theta, y; (\mu, \beta, \gamma, \delta)) = \sqrt{\mu + \sum_{j=1}^N \beta_j \theta_{t-j}^2 + \sum_{k=1}^L \gamma_k y_{t-k}^2 + \sum_{k=1}^L I(y_{t-k} < 0) \delta_k y_{t-k}^2} \quad (2.4)$$

Both processes have standard recursions of GARCH-type models (the Taylor-Schwert model and the GJR model). They allow a leverage effect if any $\delta_{i,k}^-$'s or $\delta_{i,k}^+$'s are non-zero. If all $\delta_{i,k}^-$'s and $\delta_{i,k}^+$'s are zero, these models simplify and will be called the Bayesian Joint Symmetric Absolute Value (B-JSAV) model and the Bayesian Joint Symmetric Squared Value (B-JSSV) model respectively.

Figure 2 shows realisations of 4000 time points of a B-JSAV(1,1) model and a B-JSSV(1,1) model with $L = N = 1$ and a normal centring distribution. The parameter values are given in the appendix. The graphs show the 1%, 2.5%, 5%, 12.5%, 25%, 37.5%, 62.5%, 75%, 87.5%, 95%, 97.5% and 99% quantiles as well as $\theta_{i,t}^-$, $\theta_{i,t}^+$ and robust measures of scale, skewness and kurtosis defined in Section 6.2.3. The realisation of the B-JSAV(1,1) model is shown in panels (a) and (b). The process generates a symmetric conditional return distribution with considerable variation in both the scale and the kurtosis. Symmetry arises from choosing $\beta_i^+ = \beta_i^-$, $\mu_i^- = \mu_i^+$ and $\gamma_i^- = \gamma_i^+$ and $\theta_{i,0}^- = \theta_{i,0}^+$. The heavy tails arise from μ_i^- increasing with i . There is less dependence in the tails of the distribution since the value of β_i^+ is decreasing with i . The realisation of the B-JSSV(1,1) model is shown in panel (c)

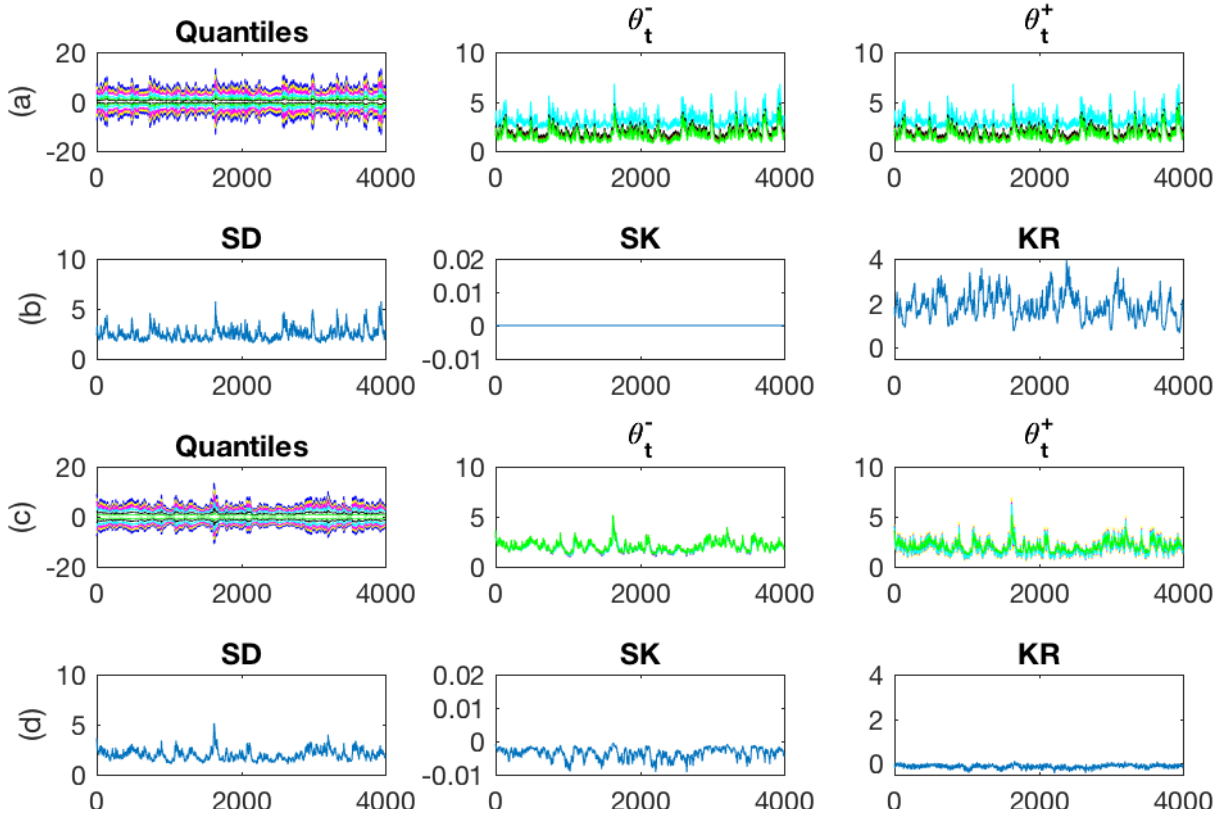


Figure 2: Simulations of a B-JSAV(1,1) process (panels (a) and (b)) and a B-JSSV(1,1) process (panels (c) and (d)). The quantiles are simulated at 1% and 99% (blue), 2.5% and 97.5% (yellow), 5% and 95% (pink), 12.5% and 87.5% (light blue), 25% and 75% (green) and 37.5% and 62.5 (black) probability levels. The $\theta_{i,t}^-$ and $\theta_{i,t}^+$ are shown as: $i = 1$ (blue), $i = 2$ (yellow), $i = 3$ (pink), $i = 4$ (light blue), $i = 5$ (black), $i = 6$ (red) and $i = 7$ (green).

and (d). This process generates substantial variation in the scale and some variation in the skewness and the kurtosis with lighter tails than the B-JSAV(1,1) realisation. The process shows greater dependence in the left-hand tail of the distribution (since $\beta_i^- \geq \beta_i^+$).

The B-JQTS model leads to analytic expressions for the important financial concepts of VaR and expected shortfall. VaR is defined as the conditional quantile usually at level 1%, 2.5%, or 5%. This can be calculated as a function of θ_t^- and θ_t^+ using the expression in Equation (2.2). The expected shortfall is the (negative) expectation of the returns below a specific quantile. Using an obvious extension of the notation in Theorem 2.1, the expected

shortfall for the B-JQTS model (for $\tau < 0.5$) is

$$\begin{aligned} \text{ES}_{t,\tau} &= -\text{E}(Y_t | Y_t \leq Q_t(\tau)) \\ &= -\frac{1}{\tau} \left[\sum_{i=\iota(\tau)+1}^K [\alpha_{i,t}^- I_{i,0}^- + \theta_{i,t}^- I_{i,1}^-] + \theta_{\iota(\tau),t}^- \int_{Q_0(1/2-a_{\iota(\tau)})}^{Q_0(\tau)} z f_0(z) dz + \right. \\ &\quad \left. (x_{\iota(\tau)-1,t}^- - \theta_{\iota(\tau),t}^- Q_0(1/2 - a_{\iota(\tau)-1}))(\tau - 1/2 + a_{\iota(\tau)}) \right] \end{aligned}$$

where $\iota(\tau) = \inf\{j | a_j \geq 1/2 - \tau\}$.

3 Properties of some B-JQTS models

The B-JAVL can be expressed as a generalised autoregressive model with i.i.d. coefficients (Bougerol and Picard 1992) and this can be used to understand the properties of the process. More general models, such as the B-JGJR and B-JSSV, cannot be expressed in this way but the approach of Cline (2007) can often be used for models with $L = N = 1$. We will demonstrate this approach by showing that the first order B-JGJR model can be stationary.

We define $\mathbf{0}_{a \times b}$ to represent an $(a \times b)$ -dimensional matrix of 0's. It is convenient to express the B-JAVL model in terms of a vector $\theta_t = (\theta_{K,t}^-, \dots, \theta_{1,t}^-, \theta_{1,t}^+, \dots, \theta_{K,t}^+)^T$, which leads to the updating equation

$$\theta_t = \mu + \sum_{j=1}^N B_j \theta_{t-j} + \sum_{k=1}^L \Gamma_k |y_{t-k}| + \sum_{k=1}^L \Delta_k \mathbf{I}(y_{t-k} < 0) |y_{t-k}|, \quad (3.1)$$

where $\mu = (\mu_{K,1}^-, \dots, \mu_{1,1}^-, \mu_{1,1}^+, \dots, \mu_{K,1}^+)^T$, $B_j = \begin{pmatrix} \text{diag}(\beta_{K,j}^-, \dots, \beta_{1,j}^-) & \mathbf{0}_{K \times K} \\ \mathbf{0}_{K \times K} & \text{diag}(\beta_{1,j}^+, \dots, \beta_{K,j}^+) \end{pmatrix}$, $\Gamma_k = (\gamma_{k,K}^-, \dots, \gamma_{k,1}^-, \gamma_{k,1}^+, \dots, \gamma_{k,K}^+)^T$ and $\Delta_k = (\delta_{k,K}^-, \dots, \delta_{k,1}^-, \delta_{k,1}^+, \dots, \delta_{k,K}^+)^T$. The following theorem establishes conditions for weak stationarity.

Theorem 3.1. *Suppose that Y_t follows a B-JAVL process and define $(s_1, c_1), \dots, (s_T, c_T)$ to*

be i.i.d. with $p(s_t = k) = |a_{|k-K-1}| - a_{|k-K|}$, $1 \leq k \leq 2K$ and

$$p(c_t | s_t) \propto \begin{cases} f_0(c_t), & Q_0(1/2 - a_{K-s_t}) < c_t < Q_0(1/2 - a_{K+1-s_t}), s_t \leq K \\ f_0(c_t), & Q_0(1/2 + a_{s_t-K-1}) < c_t < Q_0(1/2 + a_{s_t-K}), s_t > K \end{cases}$$

Let D_t and \tilde{D}_t be $(1 \times 2K)$ -dimensional vectors for which

$$(D_t)_i = \begin{cases} c_t & i = s_t \\ 0 & i \neq s_t \end{cases} \quad \text{and} \quad (\tilde{D}_t)_i = \begin{cases} 1 & i = s_t, i \leq K \\ 0 & \text{otherwise} \end{cases}$$

If $L = 1$, define

$$A_t = \begin{pmatrix} B_1 + \Gamma_1 D_t + \Delta_1 \tilde{D}_t & B^* & B_N \\ I_{2K} & \mathbf{0}_{2K \times 2K(N-2)} & \mathbf{0}_{2K \times 2K} \\ \mathbf{0}_{2K(N-2) \times 2K} & I_{2K(N-2) \times 2K(N-2)} & \mathbf{0}_{2K(N-2) \times 2K} \end{pmatrix}$$

where $B^* = (B_2 \ \cdots \ B_{N-1})$ or, if $L > 1$, define

$$A_t = \begin{pmatrix} B_1 + \Gamma_1 D_t + \Delta_1 \tilde{D}_t & B^* & B_N & C^* & \Gamma_L + \Delta_L \tilde{D}_{t-L+1} \\ I_{2K \times 2K} & \mathbf{0}_{2K \times 2K(N-2)} & \mathbf{0}_{2K \times 2K} & \mathbf{0}_{2K \times (L-2)} & \mathbf{0}_{2K \times 1} \\ \mathbf{0}_{2K(N-2) \times 2K} & I_{2K(N-2) \times 2K(N-2)} & \mathbf{0}_{2K(N-2) \times 2K} & \mathbf{0}_{2K(N-2) \times (L-2)} & \mathbf{0}_{2K(N-2) \times 1} \\ D_t & \mathbf{0}_{1 \times 2K(N-2)} & \mathbf{0}_{1 \times 2K} & \mathbf{0}_{1 \times (L-2)} & 0 \\ \mathbf{0}_{(L-2) \times 2K} & \mathbf{0}_{(L-2) \times 2K(N-2)} & \mathbf{0}_{(L-2) \times 2K} & I_{(L-2) \times (L-2)} & \mathbf{0}_{(L-2) \times 1} \end{pmatrix}$$

where $C^* = (\Gamma_2 + \Delta_2 \tilde{D}_{t-1} \ \cdots \ \Gamma_{L-1} + \Delta_{L-1} \tilde{D}_{t-L+2})$. The process Y_t is weakly stationary if the eigenvalues of $E[A_t]$ are all less than 1 and the stationary mean of θ_t is

$\xi = L(I - E[A_t])^{-1}\mu$ where

$$L = \begin{pmatrix} I_{2K \times 2K} & \mathbf{0}_{2K \times 2K(N+L-1)} \\ \mathbf{0}_{2K(N+L-1) \times 2K} & \mathbf{0}_{2K(N+L-1) \times 2K(N+L-1)} \end{pmatrix}.$$

This result can also be used to derive the stationary mean of the quantiles at time t as these are a linear function of θ_t . Simpler sufficient conditions can be derived for the B-JAVL(1, 1) model.

Corollary 3.1. *Let $\beta_{\max} = \max\{\beta_1^-, \dots, \beta_K^-, \beta_1^+, \dots, \beta_K^+\}$, then a B-JAVL(1, 1) model is weakly stationary if $\beta_{\max} + E[D_t]\Gamma_1 + E[\tilde{D}_t]\Delta_1 < 1$.*

More generally, the characteristic function of a B-JAVL($N, 1$) can be represented as the determinant of a matrix polynomial.

Corollary 3.2. *The characteristic function of a B-JAVL($N, 1$) is*

$$\det(B_N + \lambda B_{N-1} + \dots + \lambda^{N-1} (B_1 + \Gamma_1 E[D_t] + \Delta_1 E[\tilde{D}_t])) - \lambda^N I = 0$$

The previous results concern weak stationarity but the results of Bougerol and Picard (1992) can be used to find conditions for the strong stationarity of B-JAVL processes.

The following result establishes condition for the stationarity of the B-JGJR model.

Theorem 3.2. *Suppose that F_0 has finite ν -th moment for some $\nu > 0$ and $\beta_j + \gamma_j + \frac{1}{2}\delta_j < 1$ for $1 \leq j \leq 2K$, then the B-JGJR(1, 1) process is stationary.*

4 Models with a global scale

Many models for financial time series have the form $Y_t/\psi_t \sim F$, where F is a distribution without parameters (such as a standard normal or t distribution) and ψ_t is a time-varying scale. This suggests considering models of this form, where a time-varying distribution F_t

replaces F . This separates changes in the shape of the conditional distribution (modelled through the dynamics in F_t) from the global scale ψ_t and is a natural competitor to the models developed previously. Generalised Autoregressive Score (GAS) model (Creal, Koopman, and Lucas 2013) (see also the Dynamic Conditional Score (DCS) models of Harvey and Chakravarty 2008) is a simple way to define the recursion for ψ_t . If the transformation G_t is known, the GAS model for ψ_t in the B-JQTS framework is given by the recursion $\psi_{t+1} = \omega + \phi\psi_t + \eta\tilde{S}(\psi_t) \left[\frac{\partial \log f_0(G_t(y_t); \psi_t)}{\partial \psi_t} \right]$, since the score function is $\frac{\partial \log p(y_t | \psi_t)}{\partial \psi_t} = \frac{\partial \log f_0(G_t(y_t); \psi_t)}{\partial \psi_t}$. Therefore, the GAS model for the parameters of F_0 in the B-JQTS model is just the GAS model for the parameters of F_0 applied to the transformed data $G_t(y_t)$. For example, if $F_0(\cdot; \psi_t) = N(\cdot | 0, \psi_t^2)$, the GAS model is a GARCH(1, 1) model applied to the transformed $G_t(y_t)$ rather than y_t . This suggests using either GAS models for the time-varying parameters of F_0 or more general parametric models (such as EGARCH or t-GJR) applied to the transformed data.

The transformation G_t is unknown and we suggest the following scheme for updating ψ_t and θ_t . Firstly, calculate ψ_{t+1} using a GARCH-type model with the transformed data $G_t(y_t)$ and, secondly, calculate θ_{t+1} using the recursion of the B-JQTS model with y_t replaced by y_t/ψ_t . We will refer to these models as a B-JQTS models with time-varying global scale (B-JQTS with GS). We identify this model by fixing the stationary mean of ψ_t to 1.

5 Bayesian inference

The B-JQTS model has $2K$ sets of λ_i^- or λ_i^+ parameters and so is richly parameterised. As with any flexible model, there is potential of overfitting data which can lead to poor inference and out-of-sample predictive performance. The parameters λ_i^- and λ_{i+1}^- refer to continuous parts of the distribution (and similarly for λ_i^+ and λ_{i+1}^+) which suggests shrinking λ_i^- towards λ_{i-1}^- and λ_{i+1}^- (and similarly for λ_i^+ , λ_{i-1}^+ and λ_{i+1}^+). A similar argument can be made for θ_i^- and θ_i^+ . As we have discussed if all λ_i^- and λ_i^+ take the same val-

ues and all θ_i^- and θ_i^+ take the same value then the model reduces to a simpler time series model defined by the recursion H . This suggests also shrinking λ_i^- and λ_i^+ towards a constant. We will use a Bayesian approach with a Gaussian process priors (Rasmussen and Williams 2005) to smooth each dimension of λ_i^- and λ_i^+ , and θ_i^- and θ_i^+ . If we consider the B-JSAV model, the priors are $\log \mu \sim \mathbf{N}(\mu_0 \mathbf{1}_{2K}, \Sigma^\mu)$, $\log \beta \sim \mathbf{N}(\beta_0 \mathbf{1}_{2K}, \Sigma^\beta)$, and $\log \gamma \sim \mathbf{N}(\gamma_0 \mathbf{1}_{2K}, \Sigma^\gamma)$ where $\mathbf{1}_{2K}$ represents a $2K$ -dimensional column vector of 1's and, for $\nu \in \{\mu, \beta, \gamma\}$, $\Sigma_{i,j}^\nu = \sigma_\nu^2 \exp\{-(m_i - m_j)^2 / l_\nu^2\}$, σ_ν^2 is the signal variance, l_ν is the length scale, $m_i = (a_i^* + a_{i+1}^*)/2$ for $i = 1, \dots, 2K$ and $a^* = (0.5 - a_K, 0.5 - a_{K-1}, \dots, 0.5 - a_1, 0.5, 0.5 + a_1, \dots, 0.5 + a_{K-1}, 0.5 + a_K)$. The length scale controls the correlation between different elements of the vector with a larger length scale implying weaker correlation between consecutive elements. The correlation is controlled by the differences in the probability levels which are independent of the data being analysed. This allows us to define a default choice and we use $l_\nu = 0.1$.

The prior for the model is completed as follows. The hyperparameters $\mu_0, \beta_0, \gamma_0, \sigma_\mu^2, \sigma_\beta^2$ and σ_γ^2 are given hyperpriors. Again, for $\nu \in \{\mu, \beta, \gamma\}$, $\nu_0 \sim \mathbf{N}(0, \sigma_0^2)$ where σ_0^2 is given a large value to define a vague, proper prior distribution and $\sigma_\nu^2 \sim \text{Half-Cauchy}(1)$ where the density of a half-Cauchy distribution is proportional to $x^{-1/2}(1+x)^{-1}$. If we also consider a GARCH(1, 1) model for the global scale, the joint prior distribution of ϕ and η is chosen to be a uniform distribution restricted so that $0 < \phi < 1$ and $0 < \eta < 1 - \phi$ (which ensure the stationarity of the ψ_t process). Since the stationary mean of ψ_t is 1, we assume that the initial value has the prior $\log \psi_0 \sim \mathbf{N}(0, 1)$, which implies that the prior median of ψ_0 is 1. The initial values of the local and global scales are given the priors: $\log \theta_{j,0}^- \stackrel{i.i.d.}{\sim} \text{Half-Cauchy}(1)$ for $j = 1, \dots, K$ and $\log \theta_{j,0}^+ \stackrel{i.i.d.}{\sim} \text{Half-Cauchy}(1)$ for $j = 1, \dots, K$.

The likelihood function of all parameters (which are denoted $\tilde{\nu}$) in the B-JQTS model is

$$L(\tilde{\nu}) = f(y|\tilde{\nu}) = \prod_{t=1}^T f_t(y_t)$$

using the density in Equation (2.1). The posterior distribution of $\tilde{\nu}$ can be simulated using a Markov chain Monte Carlo (MCMC) scheme with adaptive Metropolis-Hastings updating, adaptive parallel tempering (Miasojedow, Moulines, and Vihola 2013) and interweaving (Yu and Meng 2011). Adaptive MCMC methods are reviewed by Griffin and Stephens (2013). Adaptive Metropolis-Hastings updating is necessary since the full conditional distributions of the posterior are non-standard. The adaptive parallel tempering is useful to move around the whole parameter space (rather than becoming trapped in local modes). A full description of the algorithm is given in Appendix B for a B-JSAV model with a GARCH(1, 1) global scale. This approach can be easily extended to other B-JQTS models (both with and without global time-varying scale) with larger values of L and/or N .

6 Applications

6.1 Simulated data example

We illustrate the ability of B-JQTS models to capture time-varying conditional distributions by fitting B-JSAV(1,1), B-JSSV(1,1) and B-JGJR(1,1) models to three simulated data sets. Each B-JQTS model is fitted with $K = 10$, $a = (0, 0.05, \dots, 0.45, 0.5)$ and a normal centring distribution. We use a GJR-GARCH(1,1) model with a t error distribution (which we will call t-GJR(1,1)) estimated using maximum likelihood as a parametric benchmark. The models are fitted to the two data sets generated from B-JQTS processes and presented in Figure 2. Data were also generated from the parametric t-GJR(1,1) model which is specified as $y_t = \sigma_t \epsilon_t$ where $\sigma_t^2 = 0.995 + 0.91\sigma_{t-1}^2 + 0.061y_{t-1}^2 + 0.032 \mathbf{I}(y_{t-1} < 0)y_{t-1}^2$ and $\epsilon_t \stackrel{i.i.d.}{\sim} t_6$. The MCMC algorithms were run for 20000 iterations, where the first 10000 iterations were discarded as the burnin-in sample. We used a thinning of 5 which resulted in a total of 2000 samples. The mixing and convergence of the posterior parameter draws were checked via trace and autocorrelation plots. Both indicated that the algorithm converged before the end

of the burn-in period, and mixed quickly.

Conditional moments are a useful way to summarise the changing shape of the conditional return distribution. Since our analysis allows for a flexible shape for this distribution, we use robust quantile-based measures of the scale, skewness and kurtosis, we adopt the Pearson-Tukey measure for the standard deviation $SD = \frac{(q_t(0.95) - q_t(0.05))}{(q_0(0.95) - q_0(0.05))}$, the Bowley coefficient of skewness, $SK = \frac{(q_t(0.75) + q_t(0.25) - 2q_t(0.50))}{(q_t(0.75) - q_t(0.25))}$, and the Crow-Siddiqui coefficient of kurtosis $KR = \frac{q_t(0.95) - q_t(0.05)}{q_t(0.75) - q_t(0.25)} - \frac{q_0(0.95) - q_0(0.05)}{q_0(0.75) - q_0(0.25)}$, where q_0 is the quantile of a standard normal distribution.

In Figure 3, we present the robust measures along with the error which we define as the difference between the posterior median and the quantile-based measure of the scale for the simulated data. Clearly, the B-JQTS models are able to capture the dynamics of the series

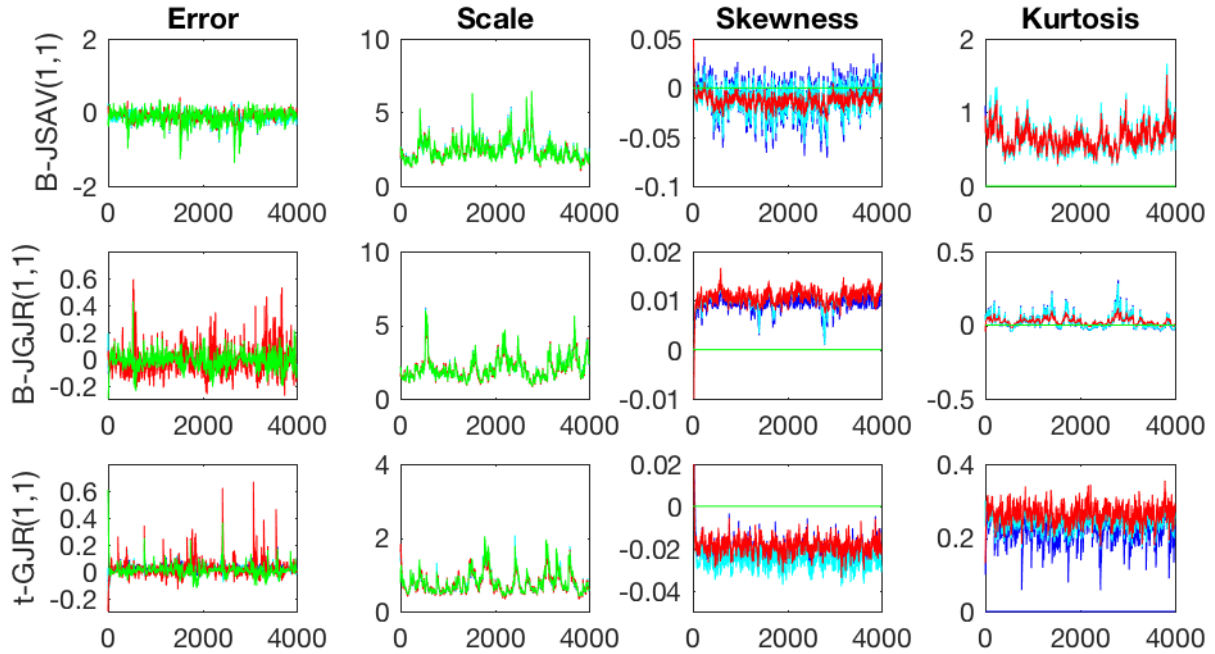


Figure 3: Inference from the B-JQTS (with $K = 10$ and $a = (0, 0.05, \dots, 0.45, 0.5)$) and GJR-GARCH models (red: B-JSAV(1,1), blue: B-JGR(1,1), cyan: B-JSSV(1,1), green: t-GJR(1,1)) using simulated data from B-JQTS and GJR-GARCH processes (shown in each row).

generated using B-JQTS models and give accurate and precise inference on the scale, skewness and kurtosis. They also provide good fits to the data generated from the t-GJR(1,1) model with a fairly constant estimate of the kurtosis around the true value (0.27). To better

Simulation Model	Fitted Model			
	B-JSAV(1,1)	B-JSSV(1,1)	B-JGJR(1,1)	t-GJR(1,1)
B-JSAV(1,1)	0.182	0.195	0.202	0.238
B-JSSV(1,1)	0.098	0.031	0.027	0.064
t-GJR(1,1)	0.065	0.040	0.039	0.042

Table 1: RMSE of SD for the B-JSAV(1, 1), B-JSSV(1, 1), B-JGJR(1, 1) and t-GJR(1, 1) fitted to data simulated using a B-JSAV(1, 1), B-JSSV(1, 1) or t-GJR(1, 1) model. The B-JQTS models were fitted using $K = 10$ and $a = (0, 0.05, \dots, 0.45, 0.5)$.

understand the performance of these models, we calculated the Root Mean Squared Errors (RMSEs) of the estimated SD for the various models and simulated data sets and we presented them in Table 1. The SD estimates from the B-JQTS models are generally better or similar to the estimates from the t-GJR(1,1) in the presence of both time-varying and time-invariant kurtosis (the only exception is the B-JSAV(1,1) model applied to the data simulated from the B-JSSV(1,1) and t-GJR(1,1) models). The B-JSAV(1,1) model gives the best performance when the data is generated from a B-JSAV(1,1) model but performs poorly when the data is generated from a model whose recursion involves squared returns (B-JSSV(1,1) and t-GJR(1,1)). This suggests that B-JQTS models can result in good performance if the observables which enter into the B-JQTS model are carefully chosen.

6.2 Real data examples

The fit and the predictive performance of various B-JQTS models with and without GS is assessed using three long time series of returns. Our data sets consist of the daily log-returns of IBM (1986/01/07–2015/12/31), S&P500 (1975/12/01–2015/12/31) and the West Texas Intermediate (WTI) price (per Barrel, in Dollars per Gallon) of crude oil (1986/01/03–

2015/12/31). The IBM and S&P500 data were obtained from the Center for Research in Security Prices (CRSP), while the WTI data were obtained from the FRED data base (DCOILWTICO series). Figure 4 (panel (a)) shows the evolution of each time series.

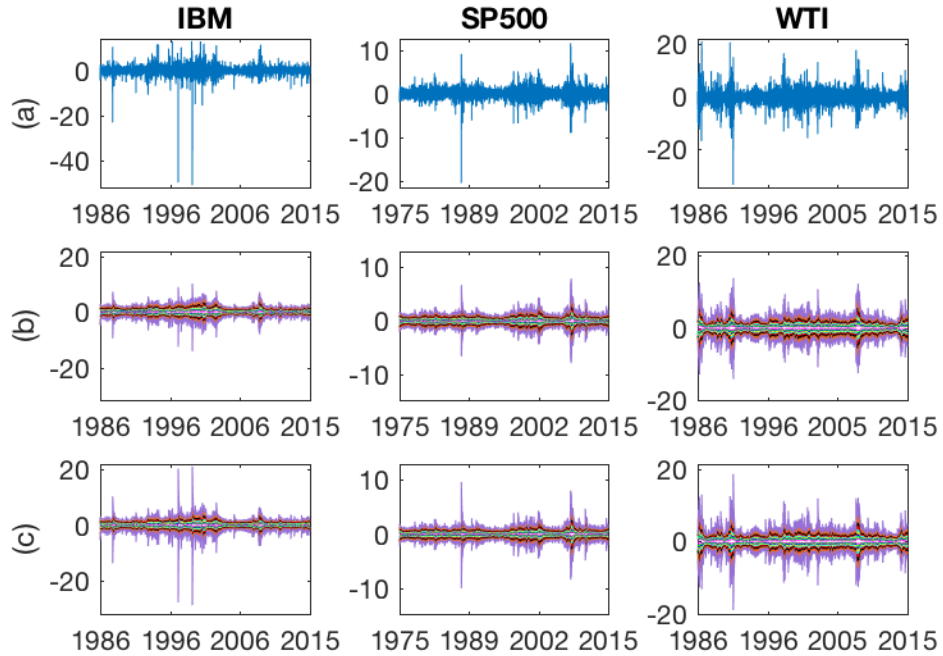


Figure 4: (a) Daily equity returns and posterior median of the conditional quantiles for (b) B-JSAV(1,1), and (c) B-JSSV(1,1) for IBM, S&P500, and WTI. We use different colours to depict the various quantiles for $K = 10$ and $a = (0, 0.05, \dots, 0.45, 0.5)$.

6.2.1 Fitting exercise

We fitted the B-JSAV(1,1), B-JSSV(1,1), B-JGJR(1,1), and B-JAVL(1,1) models with a normal centring distribution to the three time series for $K = 10$ and $a = (0, 0.05, \dots, 0.45, 0.5)$. The posterior medians of the conditional quantiles for the B-JSAV(1,1) and B-JSSV(1,1) models are shown in Figure 4. The results for B-JGJR(1,1) and B-JAVL(1,1) models are shown in Appendix D. The B-JQTS model is able to describe the changing conditional distribution of returns (by construction, the model leads to no quantile crossings). We find that

the inner quantiles (25% and 75%) are less volatile over time than the outer quantiles (5%, 10%, 90% and 95%) for all assets. This is in line with the stylised facts of asset returns. The largest differences in the fitted quantiles between different models occur after large shocks. The models where the dynamics are driven by the squared returns (B-JSSV and B-JGJR) have larger responses to the most extreme returns than the models which consider absolute returns (B-JSAV and B-JAVL).

6.2.2 Comparing forecasting performance

We compare the forecasting performance of the different versions of the B-JQTS model using $K = 4, 10$ and 20 (with $a_i = 0.5i/K$ for $i = 0, 1, \dots, K$). We included the GARCH(1, 1) and GJR-GARCH(1, 1) (with t errors) estimated using maximum likelihood as parametric benchmarks, and two more flexible specifications: a Dirichlet process mixture GARCH(1,1) (DPM-GARCH) model and a Most Likely Transformations model (MLT1) (Hothorn, Kneib, and Bühlmann 2014). The DPM-GARCH model assumes that $y_t/\sigma_t \sim F$, where F is a Dirichlet process mixture of normals and σ_t^2 follows a GARCH(1,1) process. This model allows for time-varying volatility and unknown (time-invariant) return distribution. The MLT1 model uses the approach of Hothorn, Kneib, and Bühlmann (2014) to nonparametrically estimate the conditional distribution of $y_t|y_{t-1}$ using $f_1(y_t) \sim f_2(y_{t-1}) + \epsilon_t$, where f_1 and f_2 are defined using Bernstein polynomials and ϵ_t are independent logistic distributed random variables. The number of knots are estimated using cross-validation on the training sample.

The density forecasting performance of the models was assessed using the log predictive scores (LPS) (Kim, Shephard, and Chib 1998). We use the following version of the LPS, $LPS = -\frac{1}{T-T_0} \sum_{t=T_0+1}^T \log f_t(y_{t+1}|\mathcal{F}_t^Y, \hat{\theta}, f_0(\cdot; \hat{\psi}), a)$, where $\hat{\theta}$ and $\hat{\psi}$ are parameter estimates calculated using the first T_0 observations y_1, \dots, y_{T_0} and \mathcal{F}_s^Y represents the history of the times series Y_t up to (and including) time s . The posterior mean is used as the parameter estimate in the Bayesian models. Under this criterion, models provid-

ing better forecasts have smaller LPS values. In financial time series data, it is common to find that many models give similar LPS values, but show differences in their ability to forecast extreme returns (which, by definition, are relatively rare and only make a small contribution to the LPS). In these cases, the log predictive tail score (LPTS) for one-step ahead predictions (Delatola and Griffin 2011) can be used to distinguish models with good predictive performance for extreme events. The LPTS at τ probability level is calculated as
$$\text{LPTS}_\tau = -\frac{1}{\sum_{t=T_0+1}^T I(y_{t+1} > z_\tau)} \sum_{t=T_0+1}^T I(y_{t+1} > z_\tau) \log f_t(y_{t+1} | \mathcal{F}_t^Y, \hat{\theta}, f_0(\cdot; \hat{\psi}), a),$$
 where z_τ represents the upper 100 τ % point of the empirical distribution of the data and I is an indicator function. It is important to note that this measure can only offer an insight in the differences between various models and cannot be used for formal model selection.

The findings are presented in Table 2. Overall, we find that for most specifications, predictive performance improves as K increases with only a small improvement for $K = 20$ over $K = 10$. Models without a global scale have similar or better LPS than the corresponding models with a global scale. It is striking that the models without global scale give typically better forecasting performance than the models with global scale. This is due to the extra parameters in the models with global scale leading to overfitting of the data.

More specifically, models without global scale tend to perform better in terms of the LPS in the S&P500 and to perform similarly for the IBM and WTI data. Models without global scale tend to perform better in terms of the LPTS in the S&P500, WTI data, and IBM data with $K = 4$, but the advantages for the IBM data is not clear when $K = 10$. We also find that models allowing for the leverage effect can often provide better predictive performance. The models which account for the leverage effect tend to be better than the non-leverage models (particularly in the tails) with the S&P500 data. The B-JAVL(1,1) is found to perform well when $K = 10$ with the WTI data. B-JAVL(1,1) outperforms others in LPS and in the tails for the IBM data.

The forecasting performance of the B-JQTS model is found to be superior to that of the

non-Bayesian parametric GARCH-type models and the DPM-GARCH as we would expect given the properties of B-JQTS. More specifically, B-JQTS allows for a time-varying scale as well as a time-varying shape of the conditional distribution. This extra flexibility offers better out-of-sample forecasts as the model predicts extreme returns more accurately. As for the performance of the different B-JQTS specifications it seems that this varies among the assets. MLT1 outperforms our model only in LPS for IBM but our model performs much better in the tails.

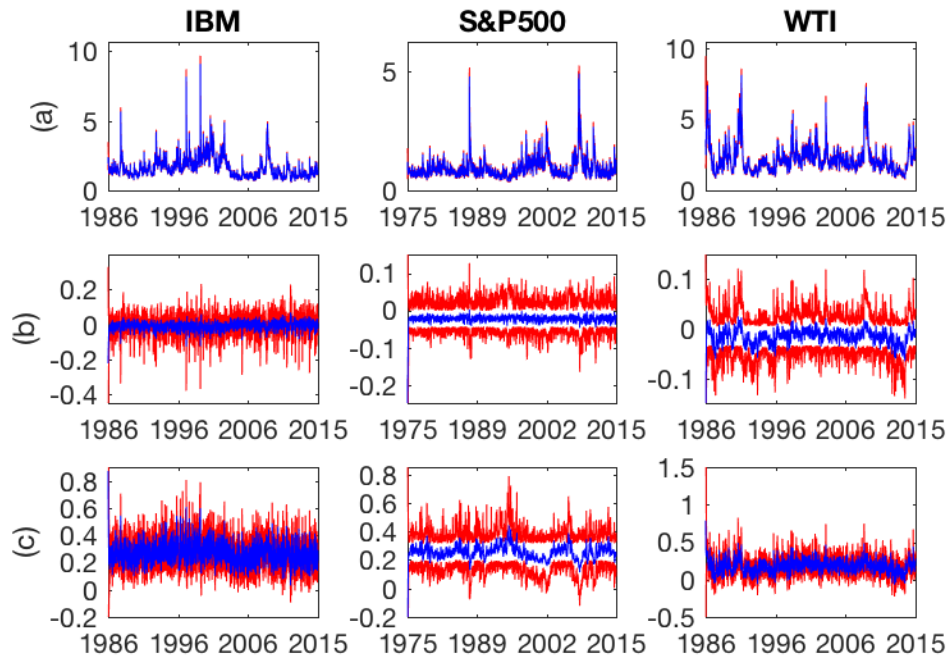


Figure 5: B-JAVL(1,1) robust quantile-based measures of the (a) scale, (b) skewness, and (c) kurtosis for IBM for $K = 10$, S&P500 for $K = 4$, and WTI for $K = 20$. We colour with red the 95% credible intervals.

To better understand the effect of the centring distribution on predictive performance, we have looked at the predictive performance of some B-JQTS models with a t centring distribution. For each value of K , we chose the B-JQTS model (with a normal centring distribution) which gave the best predictive performance (under the LPS and LPTS criteria). The results

are presented in Table 3. We find that the specification with the t distribution outperforms the normal distribution for IBM and WTI under LPS, while the normal distribution outperforms the t distribution for IBM and S&P500 under LPTS. However, there is evidence that the difference becomes smaller as K becomes larger. This is not surprising since the time-varying distribution can adapt more easily to a heavier tailed distribution for larger values of K . This suggests that B-JQTS models with a sufficiently large value of K are not sensitive to the choice of the centring distribution.

6.2.3 Time-varying conditional moments

The B-JQTS model provides a description of the time-varying distribution of returns. In this section, we look in more detail at the results for each asset using the B-JQTS model with the best performance under LPS and LPTS criteria (combined) in each case (detailed results for the B-JSAV(1, 1) model are presented in Appendix C and for other B-JQTS models in Appendix D). The best specification is B-JAVL(1,1) with $K = 10$ for IBM, $K = 4$ for S&P500 and $K = 20$ for WTI.

Figure 5 shows plots of the robust quantile-based measures of the scale (panel (a)), skewness (panel (b)), and kurtosis (panel (c)). The model is able to capture the time-varying volatility with some short periods in which volatility increases rapidly. The volatility of the S&P500 index is more stable when compared to the other two assets. The skewness and the kurtosis are relatively constant for all the assets with some evidence of more time-varying skewness for WTI and kurtosis for S&P500.

Figure 6 shows the inference about the conditional moments in 1996, during the Great Moderation, and 2009, following the global financial crisis. These are periods with very different financial market conditions. Here, we present results for IBM and S&P500, but we did the exercise for WTI too.

We find that the results for 2009 were characterised by larger and more rapid short-term

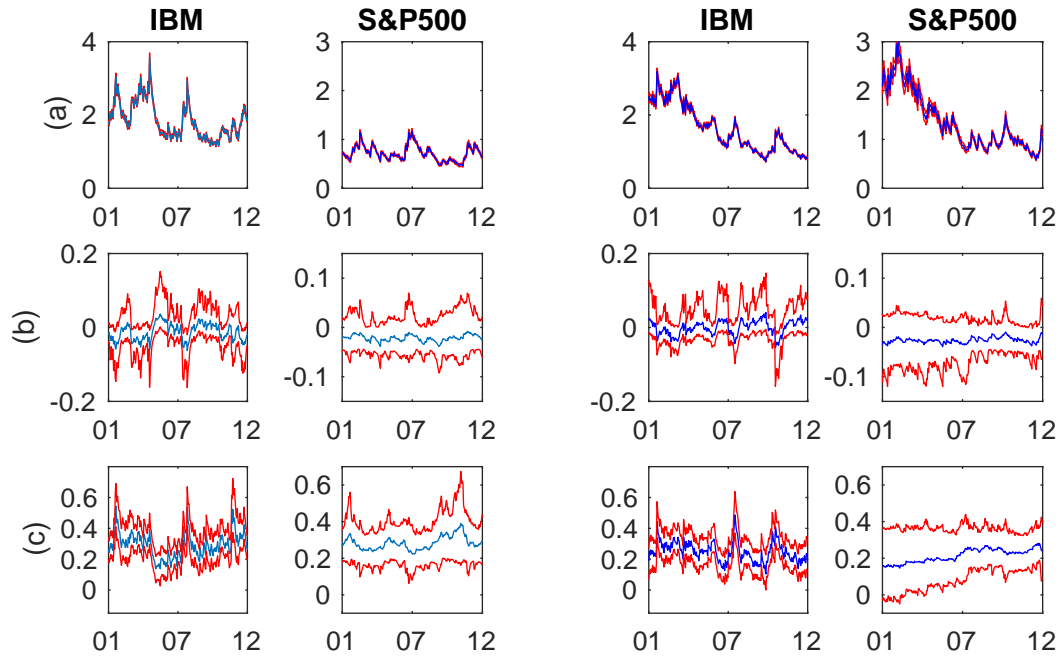


Figure 6: Robust quantile-based measures of the (a) scale, (b) skewness, and (c) kurtosis for B-JAVL(1,1) at $K = 10$ for IBM, $K = 4$ for S&P500 and $K = 20$ for WTI. We colour with red the 95% credible intervals.

changes in the higher moments compared to that of 1996. More specifically, we find that there is strong evidence of time-varying volatility for all 3 assets during 2009. The skewness (panel (b)) for both years is relatively constant. As for the kurtosis (panel (c)), we find that IBM and WTI show more changes than S&P500 for both periods under study.

6.2.4 VaR

As we discuss in Section 2, the B-JQTS model can be used to forecast the VaR. We consider forecasting VaR at the 1%, 2.5%, and 5% probability to measure its out-of-sample performance for estimating quantile in the tail of the conditional return distribution. We look at the three assets under study and compare the forecasting performance of our B-JQTS models

against the non-Bayesian parametric GARCH(1,1) and GJR-GARCH(1,1) with t distribution, for which we ran a rolling out-of-sample exercise.

In Table 4, we examine the out-of-sample violation ratios (the proportion of actual VaR exceedances relative to the expected proportion) along with the Dynamic Quantile (DQ) test (Engle and Manganelli 2004). We find that our model provides better out-of-sample violation ratios for all three data sets. The quantile score provides a measure of performance at a particular probability level. The quantile score is given by $QS(q_t, y_t) = (y_t - q_t)(\tau - I(y_t \leq q_t))$, and it is a strictly consistent measure where a lower value indicates superior accuracy. In Table 5, we present the the sum of the quantile score and we find that our models perform better when compared against the benchmark parametric GARCH-type models for all three data sets.

7 Discussion

The B-JQTS model is a flexible model for time-varying conditional distributions. The model is very general and the structure allows us to derive sufficient conditions for stationarity of some important sub-models. The application of the model to simulated and real data shows that the B-JQTS model (with suitable regularisation) can outperform state-of-the-art methods for density forecasts and VaR evaluation.

The model is defined for a finite number of quantiles but it is interesting to think about how the model limits as $K \rightarrow \infty$. Intuitively, the limit will exist if the Gaussian process priors for the parameters generate continuous sample paths, which illustrates the importance of using a smoothing prior for these parameters. The existence of a limiting process would define a Bayesian nonparametric time series model and would allow us to choose K , the number of quantile, to be “sufficiently large” to adequately approximate the infinite dimensional process rather than choosing K using model selection methods. However, a more formal study of these properties is beyond the scope of this paper.

Supplementary Materials

The online supplementary materials include proofs, detailed description of the MCMC algorithm and additional results.

References

- Andersen, T. G., T. Bollerslev, and F. X. Diebold. 2009. “Parametric and Nonparametric Volatility Measurement”. In *Handbook of Financial Econometrics, Vol. 1: Tools and Techniques*, ed. by Y. Aït-Sahalia and L. P. Hansen. North Holland.
- Andersen, T. G., R. A. Davis, J.-P. Kreiss, and T. V. Mikosch. 2009. *Handbook of Financial Time Series*. Springer-Verlag.
- Bauwens, L., A. Dufays, and J. V. K. Rombouts. 2014. “Marginal likelihood for Markov-switching and change-point GARCH models”. *Journal of Econometrics* 178:508–522.
- Bougerol, P., and N. Picard. 1992. “Strict stationarity of Generalized Autoregressive Processes”. *Annals of Probability* 20:1714–1730.
- Cai, Y., J. Stander, and N. Davies. 2012. “A new Bayesian approach to quantile autoregressive time series model estimation and forecasting”. *Journal of Time Series Analysis* 33:684–698.
- Chen, C. W. S., M. K. P. So, and F.-C. Liu. 2011. “A review of threshold time series models in finance”. *Statistics and Its Interface* 4:167–181.
- Cline, D. B. H. 2007. “Stability of Nonlinear Stochastic Recursions with Application to Nonlinear AR-GARCH Models”. *Advances in Applied Probability* 39 (2): 462–491.

- Creal, D., S. J. Koopman, and A. Lucas. 2013. "Generalized Autoregressive Score Models With Applications". *Journal of Applied Econometrics* 28:777–795.
- Delatola, E.-I., and J. E. Griffin. 2011. "Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility". *Bayesian Analysis* 6:901–926.
- Dunson, D. B., and J. W. Taylor. 2005. "Approximate Bayesian inference for quantiles". *Journal of Nonparametric Statistics* 17:385–400.
- Engle, R., and S. Manganelli. 2004. "CAViaR: Conditional Autoregressive Value-at-Risk by regression quantiles". *Journal of Business and Economic Statistics* 22:367–381.
- Gerlach, R. H, C. W. S. Chen, and N. Y. C. Chan. 2011. "Bayesian time-varying quantile forecasting for Value-At-Risk in financial markets". *Journal of Business & Economic Statistics* 29:481–492.
- Griffin, J. E., and M. F. J. Steel. 2011. "Stick-breaking autoregressive processes". *Journal of Econometrics* 162:383–396.
- Griffin, J. E., and D. A. Stephens. 2013. "Advances in Markov chain Monte Carlo". In *Bayesian Theory and Applications*, ed. by P. Damien, P. Dellaportas, N. G. Polson, and D. A. Stephens, 104–142. Oxford University Press.
- Hansen, B. E. 1994. "Autoregressive conditional density estimation". *International Economic Review* 35:705–730.
- Harvey, A., and T. Chakravarty. 2008. *Beta-t-EGARCH*. Cambridge Working Papers in Economics. Faculty of Economics, University of Cambridge.
- Harvey, C. R., and A. Siddique. 1999. "Autoregressive Conditional Skewness". *Journal of Financial and Quantitative Analysis* 34 (4): 465–487.

- Hothorn, T., T. Kneib, and P. Bühlmann. 2014. “Conditional transformation models”. *Journal of the Royal Statistical Society, Series B* 76:3–27.
- Hothorn, Y., L. Möst, and P. Bühlmann. 2018. “Most Likely Transformations”. *Scandinavian Journal of Statistics* 45:110–134.
- Kim, S., N. Shephard, and S. Chib. 1998. “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models”. *The Review of Economic Studies* 65:361–393.
- Kim, T.-H., and H. White. 2004. “On more robust estimation of skewness and kurtosis”. *Finance Research Letters* 1:56–73.
- Liu, X. 2016. “Markov switching quantile autoregression”. *Statistica Neerlandica* 70:356–395.
- Miasojedow, B., E. Moulines, and M. Vihola. 2013. “An adaptive parallel tempering algorithm”. *Journal of Computational and Graphical Statistics* 22:649–664.
- Rasmussen, C. E., and C. K. I. Williams. 2005. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.
- Rodriguez, Abel, and Enrique ter Horst. 2008. “Bayesian dynamic density estimation”. *Bayesian Anal.* 3:339–365.
- Smith, M. S., and W. Maneesoonthorn. 2018. “Inversion copulas from nonlinear state space models with an application to inflation forecasting”. *International Journal of Forecasting* 34:389–407.
- So, M. K. P., K. Lam, and W. K. Li. 1998. “A stochastic volatility model with Markov switching”. *Journal of Business and Economic Statistics* 16:244–253.

- Song, Y. 2014. “Modelling regime switching and structural breaks with an infinite hidden Markov model”. *Journal of Applied Econometrics* 29:825–842.
- Taylor, J. W. 2005. “Generating volatility forecasts from Value At Risk estimates”. *Management Science* 51:712–725.
- Tokdar, Surya T., and Joseph B. Kadane. 2012. “Simultaneous Linear Quantile Regression: A Semiparametric Bayesian Approach”. *Bayesian Anal.* 7:51–72.
- White, H., T.-H. Kim, and S. Manganelli. 2010. “Modeling Autoregressive Conditional Skewness and Kurtosis with Multi-Quantile CAViaR”. In *Volatility and Time Series Econometrics: Essays in Honour of Robert Engle*. Ed. by M. Watson, T. Bollerslev, and J. Russell, 231–256. Oxford University Press.
- Xiao, Z., and R. Koenker. 2009. “Conditional Quantile Estimation for Generalized Autoregressive Conditional Heteroscedasticity Models”. *Journal of the American Statistical Association* 104:1696–1712.
- Yu, K., and R. A. Moyeed. 2001. “Bayesian quantile regression”. *Statistics & Probability Letters* 54:437–447.
- Yu, Y., and X.-L. Meng. 2011. “To center or not to center: that is not the question – an ancillarity-sufficiency interweaving strategy (ASIS) for Boosting MCMC efficiency”. *Journal of Computational and Graphical Statistics* 20:531–570.

	IBM			S&P500			WTI		
	LPS	LPTS	LPTS	LPS	LPTS	LPTS	LPS	LPTS	LPTS
		(0.01)	(0.05)		(0.01)	(0.05)		(0.01)	(0.05)
	$K = 4$								
B-JSAV(1,1)	1.701	5.463	4.200	1.428	5.067	3.936	2.318	5.909	4.549
B-JSSV(1,1)	1.731	5.446	4.182	1.430	4.993	3.935	2.173	5.870	4.577
B-JGJR(1,1)	1.744	5.297	4.092	1.420	4.868	3.892	2.204	5.854	4.575
B-JAVL(1,1)	1.700	5.401	4.199	1.417	4.902	3.853	2.171	5.937	4.562
With GS									
B-JSAV(1,1)	1.707	5.677	4.251	1.435	5.780	4.329	2.165	5.986	4.565
B-JSSV(1,1)	1.717	5.686	4.205	1.449	6.716	4.692	2.233	5.781	4.550
B-JGJR(1,1)	1.771	5.932	4.373	1.451	6.331	4.589	2.601	6.328	5.188
B-JAVL(1,1)	1.772	5.926	4.421	1.445	6.410	4.602	2.342	5.860	4.538
	$K = 10$								
B-JSAV(1,1)	1.686	5.541	4.345	1.427	5.035	3.987	2.176	5.842	4.579
B-JSSV(1,1)	1.716	5.385	4.261	1.426	4.895	3.903	2.237	5.799	4.574
B-JGJR(1,1)	1.725	5.278	4.212	1.423	4.816	3.913	2.177	5.806	4.589
B-JAVL(1,1)	1.689	5.249	4.197	1.419	4.914	3.943	2.168	5.877	4.610
With GS									
B-JSAV(1,1)	1.711	5.351	4.280	1.457	6.685	4.780	2.163	5.899	4.631
B-JSSV(1,1)	1.707	5.414	4.272	1.457	6.744	4.790	2.220	6.001	4.743
B-JGJR(1,1)	1.717	5.505	4.366	1.450	6.335	4.672	2.348	5.752	4.614
B-JAVL(1,1)	1.703	5.481	4.324	1.446	6.379	4.642	2.243	5.788	4.552
	$K = 20$								
B-JSAV(1,1)	1.680	5.743	4.419	1.429	5.022	3.994	2.172	5.866	4.632
B-JSSV(1,1)	1.720	5.692	4.461	1.427	4.912	3.938	2.195	5.815	4.630
B-JGJR(1,1)	1.719	5.312	4.294	1.425	4.739	3.911	2.166	5.830	4.636
B-JAVL(1,1)	1.677	5.291	4.239	1.421	4.809	3.921	2.159	5.773	4.600
With GS									
B-JSAV(1,1)	1.711	5.584	4.387	1.463	6.818	4.883	2.170	6.076	4.805
B-JSSV(1,1)	1.713	5.593	4.371	1.460	6.729	4.854	2.176	5.887	4.701
B-JGJR(1,1)	1.709	5.602	4.420	1.455	6.654	4.784	2.175	6.353	4.869
B-JAVL(1,1)	1.695	5.527	4.392	1.449	6.342	4.689	2.166	5.743	4.606
DPM-GARCH	1.831	8.353	6.604	1.531	7.578	4.939	2.554	11.423	8.107
MLT1	1.595	12.923	8.175	1.667	11.661	7.646	3.061	18.428	12.759
GARCH(1,1)	8.876	8.936	8.933	8.787	8.876	8.867	8.975	9.010	9.009
GJR GARCH(1,1)	8.902	8.961	8.958	8.781	8.861	8.852	8.974	9.010	9.009

Table 2: The LPS and LPTS for different models applied to the three real data sets (bold type indicates the best performing model on each criterion per asset).

	IBM			S&P500			WTI		
	LPS	LPTS	LPTS	LPS	LPTS	LPTS	LPS	LPTS	LPTS
	(0.01)	(0.05)		(0.01)	(0.05)		(0.01)	(0.05)	
	$K = 4$								
	B-JGJR(1,1)			B-JAVL(1,1)			B-JSSV(1,1) with GS		
t	1.715	5.566	4.303	1.424	5.034	3.982	2.156	5.650	4.625
normal	1.744	5.297	4.092	1.417	4.902	3.853	2.233	5.781	4.550
	$K = 10$								
	B-JAVL(1,1)			B-JGJR(1,1)			B-JAVL(1,1) with GS		
t	1.675	5.678	4.367	1.421	4.825	3.946	2.161	5.959	4.791
normal	1.689	5.249	4.197	1.423	4.816	3.913	2.243	5.788	4.552
	$K = 20$								
	B-JAVL(1,1)			B-JGJR(1,1)			B-JAVL(1,1)		
t	1.665	5.710	4.421	1.427	4.858	3.950	2.155	5.665	4.597
normal	1.677	5.291	4.239	1.425	4.739	3.911	2.159	5.773	4.600

Table 3: The LPS and LPTS for different models applied to the three data sets when F_0 is a t or a normal distribution (bold type indicates the best performing model on each criterion per asset).

	IBM			S&P500			COIL		
	5%	2.5%	1%	5%	2.5%	1%	5%	2.5%	1%
B-JSAV(1,1)	0.039*	0.021*	0.006	0.066*	0.038*	0.014	0.049	0.019	0.007
B-JGJR(1,1)	0.038*	0.018	0.005*	0.066*	0.039*	0.011	0.051	0.021*	0.007
GARCH(1,1)	0.020*	0.010*	0.003*	0.031*	0.012*	0.003*	0.023*	0.009*	0.003*
GJR GARCH(1,1)	0.017*	0.010*	0.004*	0.032*	0.011*	0.003*	0.021*	0.009*	0.003*

Table 4: Out-of-sample violation ratios for B-JSAV(1,1), B-JGJR(1,1), GARCH(1,1), and GJR GARCH(1,1). The B-JQTS models were fitted using $K = 6$ and $a = (0, 0.25, 0.4, 0.45, 0.475, 0.49, 0.5)$. * denotes rejection from the DQ test (Engle and Manganelli 2004) at 1% significance level.

	IBM			S&P500			COIL		
	5%	2.5%	1%	5%	2.5%	1%	5%	2.5%	1%
B-JSAV(1,1)	284	240	224	525	410	306	578	459	387
B-JGJR(1,1)	296	251	238	534	427	327	588	468	396
GARCH(1,1)	662	777	975	661	761	956	940	1104	1396
GJR GARCH(1,1)	629	745	937	650	753	947	939	1105	1399

Table 5: Out-of-sample QS for B-JSAV(1,1), B-JGJR(1,1), GARCH(1,1), and GJR GARCH(1,1). The B-JQTS models were fitted using $K = 6$ and $a = (0, 0.25, 0.4, 0.45, 0.475, 0.49, 0.5)$. Bold type indicates the best performing model at each probability level per asset.