Supplementary Material for "A Bayesian quantile time series model for asset returns"

J. E. Griffin and G. Mitrodima

Abstract

Proofs, detailed description of the MCMC algorithm and additional results for the paper "A Bayesian quantile time series model for asset returns"

A Parameter values for simulations

	(0, 0.01)	(0.01,0.025)	(0.025, 0.05)	(0.05, 0.125)	(0.125, 0.25)	(0.25, 0.375)	(0.375, 0.5)
μ^-	0.27	0.27	0.27	0.27	0.06	0.03	0.015
β^-	0.845	0.845	0.845	0.845	0.9	0.9	0.9
γ^-	0.14	0.14	0.14	0.14	0.09	0.09	0.095

	(0.5, 0.625)	(0.625, 0.75)	(0.75, 0.875)	(0.875, 0.95)	(0.95, 0.975)	(0.975, 0.99)	(0.99, 1)
μ^+	0.015	0.03	0.06	0.27	0.27	0.27	0.27
β^+	0.9	0.9	0.9	0.845	0.845	0.845	0.845
γ^+	0.095	0.09	0.09	0.14	0.14	0.14	0.14

Table 1: Parameter values for the simulated B-JSAV(1, 1) process

	(0, 0.01)	(0.01,0.025)	(0.025, 0.05)	(0.05, 0.125)	(0.125, 0.25)	(0.25, 0.375)	(0.375, 0.5)
μ^-	0	0.0125	0.01	0.0075	0.08	0.08	0.075
β^-	0.94	0.945	0.955	0.965	0.9	0.9	0.9
γ^-	0.06	0.05	0.04	0.03	0.09	0.09	0.09

	(0.5, 0.625)	(0.625, 0.75)	(0.75, 0.875)	(0.875, 0.95)	(0.95, 0.975)	(0.975, 0.99)	(0.99, 1)
μ^+	0.075	0.07	0.07	0.0023	0.0025	0.0028	0.0324
β^+	0.9	0.9	0.9	0.82	0.8	0.75	0.7
γ^+	0.09	0.09	0.09	0.175	0.195	0.245	0.246

Table 2: Parameter values for the simulated B-JSSV(1, 1) process

B Proofs

B.1 Proof of Theorem 3.1

Let $Y \sim F$, where $F = \text{LIT}(\theta^-, \theta^+, F_0, a)$ then

$$
\begin{aligned}
\mathrm{E}[y^\ell] &= \int y^\ell f(y)\, dy \\
&= \sum_{i=1}^{K} \int_{x_i^-}^{x_{i-1}^-} y^\ell g(y) f_0(G(y))\, dy + \sum_{i=1}^{K} \int_{x_{i-1}^+}^{x_i^+} y^\ell g(y) f_0(G(y))\, dy \\
&= \sum_{i=1}^{K} \int_{G(x_i^-)}^{G(x_{i-1}^-)} (G^{-1}(z))^\ell f_0(z)\, dz + \sum_{i=1}^{K} \int_{G(x_{i-1}^+)}^{G(x_i^+)} (G^{-1}(z))^\ell f_0(z)\, dz \\
&= \sum_{i=1}^{K} \int_{Q_0(1/2-a_i)}^{Q_0(1/2-a_{i-1})} (x_{i-1}^- + \theta_i^-(z - Q_0(1/2 - a_{i-1})))^\ell f_0(z)\, dz \\
&\quad + \sum_{i=1}^{K} \int_{Q_0(1/2+a_{i-1})}^{Q_0(1/2+a_i)} (x_{i-1}^+ + \theta_i^+(z - Q_0(1/2 + a_{i-1}))^\ell f_0(z)\, dz \\
&= \sum_{i=1}^{K} \int_{Q_0(1/2-a_i)}^{Q_0(1/2-a_{i-1})} (x_{i-1}^- - \theta_i^- Q_0(1/2 - a_{i-1}) + \theta_i^- z)^\ell f_0(z)\, dz \\
&\quad + \sum_{i=1}^{K} \int_{Q_0(1/2+a_{i-1})}^{Q_0(1/2+a_i)} (x_{i-1}^+ - \theta_i^+ Q_0(1/2 + a_{i-1}) + \theta_i^+ z)^\ell f_0(z)\, dz \\
&= \sum_{i=1}^{K} \int_{Q_0(1/2-a_i)}^{Q_0(1/2-a_{i-1})} \sum_{j=0}^{\ell} \begin{pmatrix} \ell \\ j \end{pmatrix} (x_{i-1}^- - \theta_i^- Q_0(1/2 - a_{i-1}))^{\ell-j} (\theta_i^-)^j z^j f_0(z)\, dz \\
&\quad + \sum_{i=1}^{K} \int_{Q_0(1/2+a_{i-1})}^{Q_0(1/2+a_i)} \sum_{j=0}^{\ell} \begin{pmatrix} \ell \\ j \end{pmatrix} (x_{i-1}^+ - \theta_i^+ Q_0(1/2 + a_{i-1}))^{\ell-j} (\theta_i^+)^j z^j f_0(z)\, dz \\
&= \sum_{i=1}^{K} \sum_{j=0}^{\ell} \begin{pmatrix} \ell \\ j \end{pmatrix} (x_{i-1}^- - \theta_i^- Q_0(1/2 - a_{i-1}))^{\ell-j} (\theta_i^-)^j \int_{Q_0(1/2-a_i)}^{Q_0(1/2-a_{i-1})} z^j f_0(z)\, dz \\
&\quad + \sum_{i=1}^{K} \sum_{j=0}^{\ell} \begin{pmatrix} \ell \\ j \end{pmatrix} (x_{i-1}^+ - \theta_i^+ Q_0(1/2 + a_{i-1}))^{\ell-j} (\theta_i^+)^j \int_{Q_0(1/2+a_{i-1})}^{Q_0(1/2+a_i)} z^j f_0(z)\, dz
\end{aligned}
$$

B.2 Proof of Theorem 3.2

$$
\begin{aligned}
\mathrm{E}[|Y|] &= \int |y| f(y)\, dy = \int |y| g(y) f_0(G(y))\, dy \\
&= \sum_{i=1}^{K} \int_{x_i^-}^{x_{i-1}^-} |y| g(y) f_0(G(y))\, dy + \sum_{i=1}^{K} \int_{x_{i-1}^+}^{x_i^+} |y| g(y) f_0(G(y))\, dy \\
&= \sum_{i=1}^{K} \int_{G(x_i^-)}^{G(x_{i-1}^-)} |G^{-1}(z)| f_0(z)\, dz + \sum_{i=1}^{K} \int_{G(x_{i-1}^+)}^{G(x_i^+)} |G^{-1}(z))| f_0(z)\, dz \\
&= -\sum_{i=1}^{K} \int_{Q_0(1/2-a_i)}^{Q_0(1/2-a_{i-1})} (x_{i-1}^- + \theta_i^-(z - Q_0(1/2 - a_{i-1}))) f_0(z)\, dz \\
&\quad + \sum_{i=1}^{K} \int_{Q_0(1/2+a_{i-1})}^{Q_0(1/2+a_i)} (x_{i-1}^+ + \theta_i^+(z - Q_0(1/2 + a_{i-1})) f_0(z)\, dz \\
&= -\sum_{i=1}^{K} (a_i - a_{i-1})(x_{i-1}^- - \theta_i^- Q_0(1/2 - a_{i-1})) - \theta_i^- \sum_{i=1}^{K} I_{i,1}^- \\
&\quad + \sum_{i=1}^{K} (a_i - a_{i-1})(x_{i-1}^- + \theta_i^+ Q_0(1/2 + a_{i-1})) + \theta_i^+ \sum_{i=1}^{K} I_{i,1}^+ \\
&= -\sum_{i=1}^{K} \theta_i^- \Big((1/2 - a_i)\, Q_0(1/2 - a_i) - (1/2 - a_{i-1})\, Q_0(1/2 - a_{i-1}) + I_{i,1}^- \Big) \\
&\quad + \sum_{i=1}^{K} \theta_i^+ \Big((1/2 - a_i)\, Q_0(1/2 + a_i) - (1/2 + a_{i-1})\, Q_0(1/2 + a_{i-1}) + I_{i,1}^+ \Big) \\
&= \Phi\theta,
\end{aligned}
$$

where Φ is a $(1 \times 2K)$-dimensional vector with terms

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

and

$$
\theta = \left(\theta_K^-, \ldots, \theta_1^-, \theta_1^+, \ldots, \theta_K^+ \right).
$$

B.3 Proof of Theorem 4.1

Define $\theta_t^\star = (\theta_{t+1}, \ldots, \theta_{t-N+2})$ if $L = 1$ or $\theta_t^\star = (\theta_{t+1}, \ldots, \theta_{t-N+2}, |y_t|, \ldots, |y_{t-L+2}|)^T$ if $L > 1$. Then

$$\theta_t^\star = \mu^\star + A_t \theta_{t-1}^\star$$

where $\mu^{\star T} = (\mu \mathbf{0}_{1\times(2K(L-1)+N)})$.

This is a *generalised autoregressive model with i.i.d. coefficients* (Bougerol and Picard 1992) which implies that

$$\theta_t^\star = \mu^\star + \sum_{i=1}^{k-1} \prod_{j=1}^{i} A_{t-j}\mu^\star + \prod_{j=1}^{k} A_{t-j}\theta_{t-k}^\star$$

and taking expectations leads to

$$\mathrm{E}\left[\theta_t^\star\right] = \mu^\star + \sum_{i=1}^{k-1} \prod_{j=1}^{i} \mathrm{E}\left[A_{t-j}\right]\mu^\star + \prod_{j=1}^{k} \mathrm{E}\left[A_{t-j}\right]\mathrm{E}\left[\theta_{t-k}^\star\right] = \mu^\star + \sum_{i=1}^{k-1} A^i \mu^\star + A^k \mathrm{E}\left[\theta_{t-k}^\star\right]$$

where $A = \mathrm{E}\left[A_t\right]$ and

$$\mathrm{E}\left[\theta_t^\star\right] = \sum_{i=0}^{\infty} A^i \mu^\star = (I - A)^{-1}\mu^\star$$

if the absolute value of the eigenvalues of A are all less than 1.

B.4 Proof of Corollary 4.1

Theorem 4.1 states that a B-JSAV(1,1) is weakly stationary if the eigenvalues of $\mathrm{E}[A_t] = B_1 + \Gamma\mathrm{E}[D_t] + \Delta_1\mathrm{E}[\tilde{D}_t]$ are less than 1. Noticing that Γ, $\mathrm{E}[D_t]$, Δ_1 and $\mathrm{E}[\tilde{D}_t]$ are vectors and applying Theorem 2.1 of Ding and Zhou (2007) twice gives the result.

B.5 Proof of Corollary 4.2

Row reduction of $\det(\mathrm{E}[A_t] - \lambda I)$ leads to the result.

B.6 Proof of Theorem 4.2

Cline (2007) considers a time series $\theta_t \in \Theta$ with the structure

$$\theta_t = F(\theta_{t-1}, \epsilon_t),$$

where $F : \mathbb{R}^K \times \mathbb{E} \to \Theta \subset \mathbb{R}^K$ and ϵ_t are i.i.d. errors with $\epsilon_t \in \mathbb{E}$ for some Euclidean space $\mathbb{E}$ and provides conditions for the process to be stationary. We then write

$$\theta_t = B\left(\frac{\theta_{t-1}}{\|\theta_{t-1}\|}, \epsilon_t\right) \|\theta_{t-1}\| + C(\theta_{t-1}, \epsilon_t).$$

It is convenient to define the following notation. Let $\tilde{\psi}_t = \frac{\theta_t}{\|\theta_t\|}$ and $\Psi = \{\theta \in \Theta | \|\theta\| = 1\}$ and define $w(\tilde{\psi}, u) = \|B(\tilde{\psi}, u)\|$ and $\tilde{\eta}(\tilde{\psi}, u) = \frac{B(\tilde{\psi},u)}{\|B(\tilde{\psi},u)\|}$ for $\tilde{\psi} \in \Psi, u \in \mathbb{E}$. Let $\theta^\star_t = B\left(\frac{\theta^\star_{t-1}}{\|\theta^\star_{t-1}\|}, \epsilon_t\right) \|\theta^\star_{t-1}\|$ and define $\tilde{\psi}^\star_t = \frac{\theta^\star_t}{\|\theta^\star_t\|} = \tilde{\eta}\left(\tilde{\psi}^\star_{t-1}, \epsilon_t\right)$ for $\tilde{\psi}^\star_t, \tilde{\psi}^\star_t \in \Psi$.

Theorem 3.4 of Cline (2007) establishes conditions for stationarity of θ_t under the following assumptions:

Assumption 1 The error sequence $\{\epsilon_t\}$ are i.i.d. with $\mathrm{E}(|\epsilon_t|^\beta) < \infty$ for some $\beta > 0$.

Assumption 2
- (i) There exist $\underline{b}_1 > 0$, $\underline{b}_2 \leq 0$, $\tilde{b} < \infty$ and $\tilde{c}(x) = o(\|x\|)$, as $\|x\|$ such that $\max(\underline{b}_1|u| - \underline{b}_2, 0) \leq \|B(\tilde{\psi}, u)\| \leq \tilde{b}(1 + |u|)$, for all $u \in \mathbb{E}$ and $\tilde{\psi} \in \Psi$ and $\|C(\theta, u)\| \leq \tilde{c}(\theta)(1 + |u|)$, for all $u \in \mathbb{E}$ and $\theta \in \Theta$.
- (ii) For some finite K and $a \in (0, 1]$, $P(\|B(\theta/\|\theta\|, \epsilon_1)\|\theta\| + C(\theta, \epsilon_1)\| < \delta\|\theta\|) < K\delta^a$ for all $\theta \in \Theta$ such that $\|\Theta\| > 1$, and for all $\delta \in (0, 1]$.

Assumption 3 $\{X_t\}$ and $\{X^\star_t\}$ are each aperiodic r-irreducible Markov chains, on $\mathbb{X}$ and $\mathbb{X}^\star$, respectively. Furthermore, bounded subsets of $\mathbb{X}$ are small for $\{X_t\}$, and subsets of $\mathbb{X}^\star$ that are bounded and bounded away from $\{0\}$ are small for $\{X^\star_t\}$.

Assumption 4 There exists a set $\Psi_\#$, open in $\Psi = \{\theta \in \Theta : \|\theta\| = 1\}$, such that

- (i) $\{B(\cdot, u)\}_{|u| \leq M}$ is equicontinuous on $\Psi_\#$ for all finite M. That is, for each $\epsilon > 0$

and $M < \infty$, there exists $\delta > 0$, such that $|\tilde{\psi} - \tilde{\psi}'| < \delta$, $\tilde{\psi}, \tilde{\psi}' \in \Psi_{\#}$ implies $\|B(\tilde{\psi}, u) - B(\tilde{\psi}', u)\| < \epsilon$, for all $|u| \leq M$.

(a) for each $\epsilon > 0$, there exists $L < \infty$ such that $P(\tilde{\psi}_1 \in \Psi_{\#}, \tilde{\psi}_1^\star \in \Psi_{\#}|\theta_0 = \theta) > 1 - \epsilon$, for all $\theta \in \Theta$ with $\theta/\|\theta\| \in \Psi_{\#}$ and $\|\theta\| > L$, and

(b) for every $\epsilon > 0$, there exists $n \geq 1$ and $L < \infty$ such that $P(\tilde{\psi}_n \in \Psi_{\#}|\theta_0 = \theta) > 1 - \epsilon$, for all $\theta \in \Theta$ with $\|\theta\| > L$.

Assumption 3 can also be checked using the following theorem proved by Cline

Theorem B.1 (Theorem 5.1 of Cline)**.** *Let $\{X_t\}$ be a Markov process on $\mathbb{X} \subset \mathbb{R}^K$ with transition kernel T defined by $T(x, A) = P(X_1 \in A|X_0 = x)$ and suppose that the following three conditions hold*

(i) For some $k \geq 1$, $T^k(x, \cdot)$ is absolutely continuous for all $x \in \mathbb{X}$ and, hence, for each $n \geq k$, $T^n(x, \cdot)$ is absolutely continuous with some density $g_n(x, \cdot)$.

(ii) For each $x \in \mathbb{X}$ there exists $n \geq k$ satisfying: there exists an open set $A \subset \mathbb{X}$ and $\delta > 0$ such that

$$\inf_{\hat{x}:\|\hat{x}-x\|<\delta} g_n(\hat{x}, \tilde{x}) > 0 \; for \; \tilde{x} \in A$$

(iii) There exists $\tilde{x} \in \mathbb{X}$ satisfying: for each $x \in \mathbb{X}$ and $\delta > 0$, there exists $n \geq k$ such that

$$P(\|X_n - \tilde{x}\| < \delta|X_0 = x) > 0 \text{ and } P(\|X_{n+1} - \tilde{x}\| < \delta|X_0 = x) > 0.$$

Then $\{X_t\}$ is an aperiodic r-irreducible T-chain.

To show that the B-JGJR process is stationary, we define $u \sim F_0$, D^- is a $(K \times 1)$-dimensional vector where $[D^-(u)]_i = \mathrm{I}(Q_0(1/2 - a_{i-1}) < u < Q_0(1/2 - a_i))u$ and D^+ is a $(K \times 1)$-dimensional vector where $[D^+(u)]_i = \mathrm{I}(Q_0(1/2 + a_{i-1}) < u < Q_0(1/2 + a_i))u$. We also define $D(u) = \begin{pmatrix} D^-(u) \\ D^+(u) \end{pmatrix}$ and $D^\star(u) = \begin{pmatrix} D^-(u) \\ 0_K \end{pmatrix}$. This allows us to give expressions for the terms defined by Cline (2007):

$$B(x,u) = \sqrt{B_1 x^2 + \Gamma_1 (D(u)x)^2 + \Delta_1 (D^\star(u)x)^2}$$

and

$$C(\tilde{\psi},u) = \sqrt{\mu + Bx^2 + \Gamma(D(u)x)^2 + \Delta(D^\star(u)x)^2} - \sqrt{Bx^2 + \Gamma(D(u)x)^2 + \Delta(D^\star(u)x)^2},$$

where x^2 and $\sqrt{x}$ represent the component-wise square and square root respectively. We use L_2 distance as the norm.

Assumption 1 Assumption 1 is met by the assumption of the theorem.

Assumption 2 (i) The assumption is met with $b_1 = \min\{\gamma_i + \delta_i\}$, $b_2 = 0$ and

$$\tilde{b} = \max\{\max\{\beta_i^2\}, 4K \max\{\beta_i(\gamma_i + \delta_i)\}, 2K \sum_{i=1}^{2K} (\gamma_i + \delta_i)^2\}.$$

For $\tilde{\psi} \in \Psi$ and $u < 0$,

$$\begin{aligned}
\|B(\tilde{\psi},u)\|^2 &= \sum_{i=1}^{2K} \left(\beta_i \tilde{\psi}_i + \gamma_i \sum_{j=1}^{2K} D_{i,j}(u)\tilde{\psi}_j + \delta_i \sum_{j=1}^{K} D_{i,j}(u)\tilde{\psi}_j \right)^2 \\
&\leq \sum_{i=1}^{2K} \left(\beta_i \tilde{\psi}_i + u(\gamma_i + \delta_i) \sum_{j=1}^{K} \tilde{\psi}_j \right)^2 \\
&= \sum_{i=1}^{K} \beta_i^2 \tilde{\psi}_i^2 + 2u \left(\sum_{i=1}^{K} \beta_i \tilde{\psi}_i (\gamma_i + \delta_i) \right) \sum_{j=1}^{K} \tilde{\psi}_j \\
&\quad + u^2 \sum_{i=1}^{K} (\gamma_i + \delta_i)^2 \left(\sum_{j=1}^{K} \tilde{\psi}_j \right)^2 \\
&\leq \max\{\beta_i^2\} + 4uK \max\{\beta_i(\gamma_i + \delta_i)\} + 2u^2 K \sum_{i=1}^{K} (\gamma_i + \delta_i)^2 \\
&\leq \tilde{b}(1+u)^2.
\end{aligned}$$

and for $\tilde{\psi} \in \Psi$ and $u > 0$,

$$
\begin{aligned}
\|B(\tilde{\psi}, u)\|^2 &= \sum_{i=1}^{2K} \left(\beta_i \tilde{\psi}_i + \gamma_i \sum_{j=K+1}^{2K} D_{i,j}(u) \tilde{\psi}_j \right)^2 \leq \sum_{i=1}^{2K} \left(\beta_i \tilde{\psi}_i + u\gamma_i \sum_{j=K+1}^{2K} \tilde{\psi}_j \right)^2 \\
&= \sum_{i=K+1}^{2K} \beta_i^2 \tilde{\psi}_i^2 + 2u \left(\sum_{i=K+1}^{2K} \beta_i \tilde{\psi}_i \gamma_i \right) \sum_{j=K+1}^{2K} \tilde{\psi}_j \\
&\quad + u^2 \left(\sum_{i=K+1}^{2K} \gamma_i^2 \right) \left(\sum_{j=K+1}^{2K} \tilde{\psi}_j \right)^2 \\
&\leq \max\{\beta_i^2\} + 4uK \max\{\beta_i \gamma_i\} + 2Ku^2 \sum_{i=1}^{2K} \gamma_i^2 \\
&\leq \tilde{b}(1+u)^2.
\end{aligned}
$$

Furthermore, $\|B(\tilde{\psi}, u)\|^2 > \min\{\gamma_i + \delta_i\} u^2 = (\underline{b}_1 u - \underline{b}_2)^2$.

(ii) It is clear that

$$
\begin{aligned}
\|B(\theta/\|\theta\|, \epsilon_1)\|\theta\| + C(\theta, \epsilon_1)\|^2 &= \|(B + \Gamma D(\epsilon_1) + \Delta D^\star(\epsilon_1))\theta + \mu\|^2 \\
&> \sum_{j=1}^{2K} \beta_j^2\, \theta_j^2 > \min\{\beta_j^2\} \|\theta\|^2.
\end{aligned}
$$

The assumption is trivially true with $K = \min\{\beta_j^2\}$ and $a = 1$.

Assumption 3 We check the condition in Theorem B.1

(i)

$$
\theta_t^2 = \mu + \sum_{j=1}^{n-1} B^j \mu + B^n \theta_{t-n}^2 + \sum_{k=1}^{n} B^{k-1} \gamma\, y_{t-k}^2 + \sum_{k=1}^{n} B^{k-1} \delta\, \mathrm{I}(y_{t-k} < 0) y_{t-k}^2
$$

The transition kernel $p(\theta_t^2 | \theta_{t-K}^2)$ is a linear function of $y_{t-n}, \ldots, y_{t-1}$ whose joint density is absolutely continuous with respect to $\mathbb{R}^{2K}$ and so θ_t will be absolutely continuous. Since the distribution of θ_t^2 is continuous then so is the distribution of θ_t.

(ii) Since the transition kernel in (i) is absolutely continuous and any point $\theta \in \Theta$ can be approximated using a suitably large n then the condition is met.

(iii) It is clear from the form in part (i) that for any point θ there is an n which makes any points in $(\mathbb{R}^+)^{2K}$ reachable.

Assumption 4 We set $\Psi_\# = \Psi$.

(i) Clearly θ is equicontinuous on $(\mathbb{R}^+)^{2K}$ since

$$B(\tilde{\psi}, u) = \sqrt{B\tilde{\psi}^2 + \Gamma(D(u)\tilde{\psi})^2 + \Delta(D^\star(u)\tilde{\psi})^2}$$

and $D(u)_{ij} \leq D(M)$ and $D^\star(u)_{ij} < D(M)$ for all i and j.

(ii) and (iii)

$$\tilde{\psi}_1 = \frac{\theta_1}{\|\theta_1\|} = \frac{\sqrt{\mu + B\theta_0^2 + \gamma y_t^2 + \delta \mathrm{I}(y_t < 0)y_t^2}}{\|\sqrt{\mu + B\theta_0^2 + \gamma y_t^2 + \delta \mathrm{I}(y_t < 0)y_t^2}\|} \in \Psi$$

$$\tilde{\psi}_1^\star = \frac{\sqrt{(B\psi_0^{\star 2} + \gamma D(\epsilon_1)\psi_0^{\star 2} + \delta D^\star(\epsilon_1)\psi_0^{\star 2}}}{\|\sqrt{(B\psi_0^{\star 2} + \gamma S(\epsilon_1)\psi_0^{\star 2} + \delta D^\star(\epsilon_1)\psi_0^{\star 2}}\|} \in \Psi$$

$$\frac{\theta_1^\star}{\|\theta_1^\star\|} = \frac{B\left(\frac{\theta_0^\star}{\|\theta_0^\star\|}, \epsilon_t\right)}{\|B\left(\frac{\theta_0^\star}{\|\theta_0^\star\|}, \epsilon_t\right)\|} = \frac{\theta_0^\star}{\|\theta_0^\star\|}$$

and so, clearly, conditions (ii) and (iii) will be met.

C MCMC algorithm

The algorithm samples the variables: $\tilde{\nu} = (\theta_0, \mu, \beta, \gamma, \psi_0, \phi, \eta, \mu_0, \beta_0, \gamma_0, \sigma_\mu^2, \sigma_\beta^2, \sigma_\gamma^2)$. The likelihood is

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

using the density in Equation (2.1) in the main paper. In a Metropolis-Hastings updating step, we often use the shorthand $\mathcal{L}$ to represent the value of the likelihood at the current values of the parameters and $\mathcal{L}'$ to represent the value of the likelihood at the new values of the parameters (including the current values of parameters that are not being updated). Generically, we use $p(\nu)$ to represent the prior of a parameter ν.

The algorithm makes extensive use of the idea of adaptive MCMC and, in particular, the algorithm of Atchadé and Rosenthal (2005) for Metropolis-Hastings random walk. Suppose we are updating a generic parameter ν. This algorithm adjusts the scale of the proposal distribution so that the acceptance rate converges to a value $\bar{a}$ (for which we use the asympotically optimal value 0.234 Roberts, Gelman, and Gilks (1997)). At the m-th iteration of the sampler, a value ν' is proposed from the transition $q_{s_{\nu,m}}(\nu, \nu')$ where $s_{\nu,m}$ is a scale parameter (for example, the standard deviation in a normal random walk). The proposed value is accepted using the standard Metropolis-Hastings acceptance probability $\tilde{a} = \min\left\{1, \frac{\mathcal{L}'\, p(\nu')q_{s_{\nu,m}}(\nu',\nu)}{\mathcal{L}\, p(\nu)q_{s_{\nu,m}}(\nu,\nu')}\right\}$. The approach differs from standard Metropolis-Hastings by updating the scale parameter $s_{\nu,m}$ using the recursion

$$s_{\nu,m+1} = s_{\nu,m} + w_m(\tilde{a} - \bar{a}), \tag{C.1}$$

where $w_m = O(m^{-b})$ for $1/2 < b \leq 1$. This allows the sampler to automatically find an appropriate scale for the proposal in the update. Atchadé and Rosenthal (2005) showed that this updating of the scale does not affect the ergodicity of the sampler output.

The algorithm also uses the adaptive parallel tempering algorithm of Miasojedow, Moulines, and Vihola (2013). parallel tempering has long been used to improve convergence of MCMC algorithms for multi-modal posterior distributions. For a generic set of parameters, parallel tempering uses l chains which are run in parallel with $\nu^{(j)}$ being the parameters for the j-th chain and the target density of the j-th chain is

$$\pi_j\left(\nu^{(j)}|y\right) \propto p\left(y|\nu^{(j)}\right)^{\gamma_j^\star}\, p\left(\nu^{(j)}\right)$$

where $0 < \gamma_1^\star < \cdots < \gamma_l^\star = 1$ are l different temperatures. The target distribution for the l-th chain is the posterior distribution whereas the other chain will tend to be more dispersed than the posterior distribution. We can explore each target distribution using an adaptive MCMC scheme. In addition, parameter values for different chains are proposed to be swapped with a Metropolis-Hastings acceptance probability used to ensure that the chains have the correct target distributions. The effectiveness of this algorithm depends on the choice of the temperatures. We use an adaptive approach to choosing these temperatures suggested by Miasojedow, Moulines, and Vihola (2013). Atchadé, Roberts, and Rosenthal (2011) show that it is optimal to choose the acceptance probability of swaps between consecutive parameter values at consecutive temperatures to be 0.234. The algorithm of Miasojedow, Moulines, and Vihola (2013) adjusts the temperatures to maintain this acceptance rate. Further details can be found in Miasojedow, Moulines, and Vihola (2013)

The steps of the Gibbs sampler for the j-th temperature, $\gamma_j^\star$, are provided below.

Updating θ_0

Each element of $\log\theta_0$ is updated using an adaptive Metropolis-Hastings random walk (Atchadé and Rosenthal 2005). The Metropolis-Hastings acceptance probability for $\log\theta_{i,0}^{(j)}$ is

$$\min\left\{1, \frac{\mathcal{L}'^{\gamma_j^\star}\, p\left(\log\theta_{i,0}^{(j)\prime}\right)}{\mathcal{L}^{\gamma_j^\star}\, p\left(\log\theta_{i,0}^{(j)}\right)}\right\}.$$

Each element of $\log\theta_0^{(j)}$ will have its own scaling parameter which also differs across j.

Updating μ, β and γ

The parameter $\log \mu^{(j)}$ can be updated using an adaptive Metropolis-Hastings random walk sampler which uses the proposal

$$\log \mu^{(j)\prime} = \log \mu^{(j)} + \sigma_{\mu,j} s_{\mu,j} \left(\Sigma^{\mu}\right)^{1/2} \epsilon_{\mu,j}$$

where $\epsilon_{\mu,j} \sim \mathrm{N}(0_{2K}, I_{2K})$ and $\left(\Sigma^{\mu}\right)^{1/2}$ is taken to be the Cholesky decomposition of Σ^{μ}. The proposal is accepted with probability

$$\min\left\{1, \frac{\mathcal{L}'^{\gamma_j^\star}\, p\left(\log \mu^{(j)\prime}\right)}{\mathcal{L}^{\gamma_j^\star}\, p\left(\log \mu^{(j)}\right)}\right\}.$$

The scaling value $s_{\mu,j}$ is updated using the recursion in (C.1). The parameters $\beta^{(j)}$ and $\gamma^{(j)}$ can be updated in same way with their own scaling at each temperature, $s_{\beta,j}$ and $s_{\gamma,j}$.

Updating ψ_0

We update $\log \psi_0$ using an adaptive Metropolis-Hastings random walk (Atchadé and Rosenthal 2005). The Metropolis-Hastings acceptance probability is

$$\min\left\{1, \frac{\mathcal{L}'^{\gamma_j^\star}\, p\left(\log \psi_0^{(j)\prime}\right)}{\mathcal{L}^{\gamma_j^\star}\, p\left(\log \psi_0^{(j)}\right)}\right\}.$$

There is a scaling parameter for each temperature.

Updating ϕ and η

We update $\log\phi$ and $\log\eta$ using separate adaptive Metropolis-Hastings random walks (Atchadé and Rosenthal 2005). The Metropolis-Hastings acceptance probability for $\log\phi$ is

$$\min\left\{1, \frac{\mathcal{L}'^{\gamma_j^\star}\ \log\phi'\,\mathrm{I}(\log\phi'+\log\eta<1)}{\mathcal{L}^{\gamma_j^\star}\ \log\phi\,\mathrm{I}(\log\phi+\log\eta<1)}\right\}.$$

and the Metropolis-Hastings acceptance probability for $\log\eta$ is

$$\min\left\{1, \frac{\mathcal{L}'^{\gamma_j^\star}\ \log\eta'\,\mathrm{I}(\log\phi+\log\eta'<1)}{\mathcal{L}^{\gamma_j^\star}\ \log\eta\,\mathrm{I}(\log\phi+\log\eta<1)}\right\}.$$

For both parameters, there is a scaling parameter for each temperature.

Although this update is sufficient to define a Gibbs sampler for the posterior distribution, it can lead to poor mixing. This is due to the form of the transformation G_t which suggests that $\log\eta^{(j)}$ and all $\log\theta_{i,t}^{(j)}$'s will be strongly negatively correlated. To address this, we jointly update $\log\eta^{(j)}$ and $\log\theta_{1,0}^{(j)},\ldots,\log\theta_{2K,0}^{(j)}$ and $\log\mu_1^{(j)},\ldots,\log\mu_{2K}^{(j)}$. We propose $\log\eta^{(j)\prime}=\log\eta^{(j)}+\epsilon_\eta$ where $\epsilon_\eta\sim\mathrm{N}\left(0,\sigma_{\eta,m}^{(j)\,2}\right)$ and it represents the iteration and $\sigma_{\eta,m}^{(j)\,2}$ is updated using the recursion in (C.1). We propose $\log\theta_{i,0}^{(j)\prime}=\log\theta_{i,0}^{(j)}\,\log\eta^{(j)\prime}\Big/\log\eta^{(j)}$ and $\log\mu_i^{(j)\prime}=\log\mu_i^{(j)}\,\log\eta^{(j)\prime}\Big/\log\eta^{(j)}$. The proposed move is accepted with the following acceptance probability

$$\min\left\{1, \frac{\mathcal{L}'\left(\log\eta^{(j)\prime}\right)^{2K}\prod_{i=1}^{2K}p\left(\log\mu_i^{(j)\prime}\right)}{\mathcal{L}\left(\log\eta^{(j)}\right)^{2K}\prod_{i=1}^{2K}p\left(\log\mu_i^{(j)}\right)}\right\}.$$

Updating μ_0, β_0 and γ_0

The full conditional distributions of $\log\mu_0^{(j)}$, $\log\beta_0^{(j)}$ and $\log\gamma_0^{(j)}$ are

$$\log\mu_0^{(j)}\sim\mathrm{N}\left(P_\mu\left(\sigma_\mu^{(j)\,2}\right)^{-1}1^T\Sigma_\mu^{-1}\log\mu^{(j)},P_\mu\right),$$

where

$$P_\mu = \left(\left(\sigma_\mu^{(j)\,2} \right)^{-1} 1_{2K}^T \Sigma_\mu^{-1} 1_{2K} + \sigma_0^{-2} \right)^{-1},$$

$$\log \beta_0^{(j)} \sim \mathrm{N}\left(P_\beta \left(\sigma_\beta^{(j)\,2} \right)^{-1} 1^T \Sigma_\beta^{-1} \log \beta^{(j)}, P_\beta \right),$$

where

$$P_\beta = \left(\left(\sigma_\beta^{(j)\,2} \right)^{-1} 1_{2K}^T \Sigma_\beta^{-1} 1_{2K} + \sigma_0^{-2} \right)^{-1}$$

and

$$\log \gamma_0^{(j)} \sim \mathrm{N}\left(P_\gamma \left(\sigma_\gamma^{(j)\,2} \right)^{-1} 1^T \Sigma_\gamma^{-1} \log \gamma^{(j)}, P_\gamma \right),$$

where

$$P_\gamma = \left(\left(\sigma_\gamma^{(j)\,2} \right)^{-1} 1_{2K}^T \Sigma_\gamma^{-1} 1_{2K} + \sigma_0^{-2} \right)^{-1}.$$

We also use a second update step in order to improve the mixing of these series. For this we introduce a set of interweaving steps. A second update for $\log \mu_0$, also updates $\log \mu_1, \ldots, \log \mu_{2K}$ using a Metropolis-Hastings update step. This allows to update the scale of all local processes simultaneously. We propose $\log \mu_0' = \log \mu_0 + \epsilon_\mu$ where $\epsilon_\mu \sim \mathrm{N}\left(0, \sigma^2_{\mu_0,m}\right)$ and m represents the iteration and $\sigma^2_{\mu_0,m}$ is updated using the recursion in (C.1). We propose $\log \mu_j' = \log \mu_j + \log \mu_0' - \log \mu_0$. The proposed value is then accepted according to the acceptance probability

$$\min \left\{ 1, \frac{\mathcal{L}'^{\gamma_j^\star}\, p\left(\log \mu_0^{(j)\prime}\right)}{\mathcal{L}^{\gamma_j^\star}\, p\left(\log \mu_0^{(j)}\right)} \right\}.$$

Updating σ_μ^2, σ_β^2 and σ_γ^2

We update the logarithm of these parameters. The full conditional distributions of $\log\sigma_\mu^{(j)\,2}$, $\log\sigma_\beta^{(j)\,2}$ and $\log\sigma_\gamma^{(j)\,2}$ are

$$\exp\left\{-\frac{1}{2}\left(\exp\left\{\log\sigma_\mu^{(j)\,2}\right\}\right)^{-1}A_\mu\right\}\left(\exp\left\{2(1-K)\log\sigma_\mu^{(j)\,2}\right\}\right)\left(1+\exp\left\{\log\sigma_\mu^{(j)\,2}\right\}\right)^{-K},$$

where $A_\mu = \left(\log\mu^{(j)} - \log\mu_0^{(j)}1_{2K}\right)^T\Sigma_\mu^{-1}\left(\log\mu^{(j)} - \log\mu_0^{(j)}1_{2K}\right)$,

$$\exp\left\{-\frac{1}{2}\left(\exp\left\{\log\sigma_\beta^{(j)\,2}\right\}\right)^{-1}A_\beta\right\}\left(\exp\left\{2(1-K)\log\sigma_\beta^{(j)\,2}\right\}\right)\left(1+\exp\left\{\log\sigma_\beta^{(j)\,2}\right\}\right)^{-K},$$

where $A_\beta = \left(\log\beta^{(j)} - \log\beta_0^{(j)}1_{2K}\right)^T\Sigma_\beta^{-1}\left(\log\beta^{(j)} - \log\beta_0^{(j)}1_{2K}\right)$, and

$$\exp\left\{-\frac{1}{2}\left(\exp\left\{\log\sigma_\gamma^{(j)\,2}\right\}\right)^{-1}A_\gamma\right\}\left(\exp\left\{2(1-K)\log\sigma_\gamma^{(j)\,2}\right\}\right)\left(1+\exp\left\{\log\sigma_\gamma^{(j)\,2}\right\}\right)^{-K},$$

where $A_\gamma = \left(\log\gamma^{(j)} - \log\gamma_0^{(j)}1_{2K}\right)^T\Sigma_\gamma^{-1}\left(\log\gamma^{(j)} - \log\gamma_0^{(j)}1_{2K}\right)$, respectively. These parameter are updated using an adaptive Metropolis-Hastings random walk with normal increments whose variance is adapted using the method of Atchadé and Rosenthal (2005).

We also update $\log\sigma_\mu^2$ using a second adaptive Metropolis-Hastings random walk where we propose $\log\sigma_\mu^{(j)\,2}$ (on the log scale) and produce a proposed value

$$\log\mu^{\star\,\prime} = \log\mu_0^{(j)}\mathbf{1} + \frac{\log\sigma_\mu^{(j)\,2\prime}}{\log\sigma_\mu^{(j)\,2}}\left(\log\mu^\star - \log\mu_0^{(j)}\mathbf{1}\right).$$

The proposed value is then accepted according to the acceptance probability

$$\min\left\{1, \frac{\mathcal{L}'^{\gamma_j^\star}\log\sigma_\mu^{(j)\,2\prime}p\left(\log\sigma_\mu^{(j)\,2\prime}\right)}{\mathcal{L}^{\gamma_j^\star}\log\sigma_\mu^{(j)\,2}p\left(\log\sigma_\mu^{(j)\,2}\right)}\right\}.$$

Update steps for $\log\sigma_\beta^{(j)\,2}$ and $\log\sigma_\gamma^{(j)\,2}$ can be defined in a similar way.

D Results for B-JSAV(1,1) model

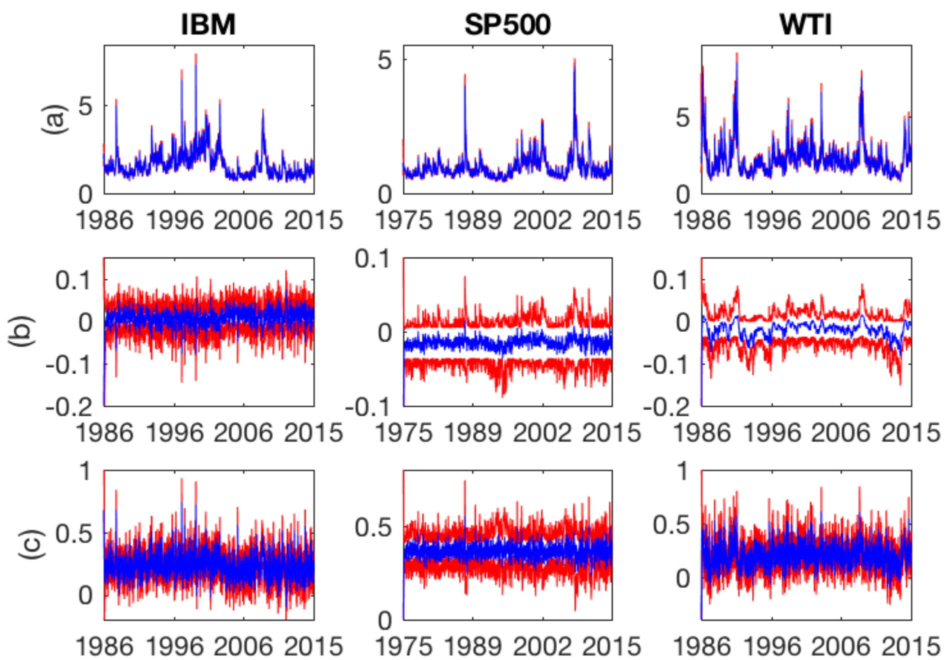

Figure 1: B-JSAV(1,1) robust quantile-based measures of the (a) scale, (b) skewness, and (c) kurtosis for IBM, S&P500, and WTI for $K = 10$ and $a = (0, 0.05, \ldots, 0.45, 0.5)$. The 95% credible intervals is in red.

The results for the B-JSAV(1,1) specification are presented in Figure 1. The model is able to capture the time-varying volatility with some short periods in which volatility increases rapidly. The S&P500 index appears to be more stable when compared to the other two assets, while WTI is the most volatile series with the highest spikes. Our model also allows for time-varying higher moments. The skewness is relatively constant for IBM and S&P500. However, there is some evidence of time-varying skewness for WTI with periods of both positive and negative skewness. There are two periods of positive skewness (1991 and 2009) while skewness becomes increasingly negative in the period from 2010 to 2014. There is also evidence of time-varying kurtosis. The variation is particularly pronounced for WTI

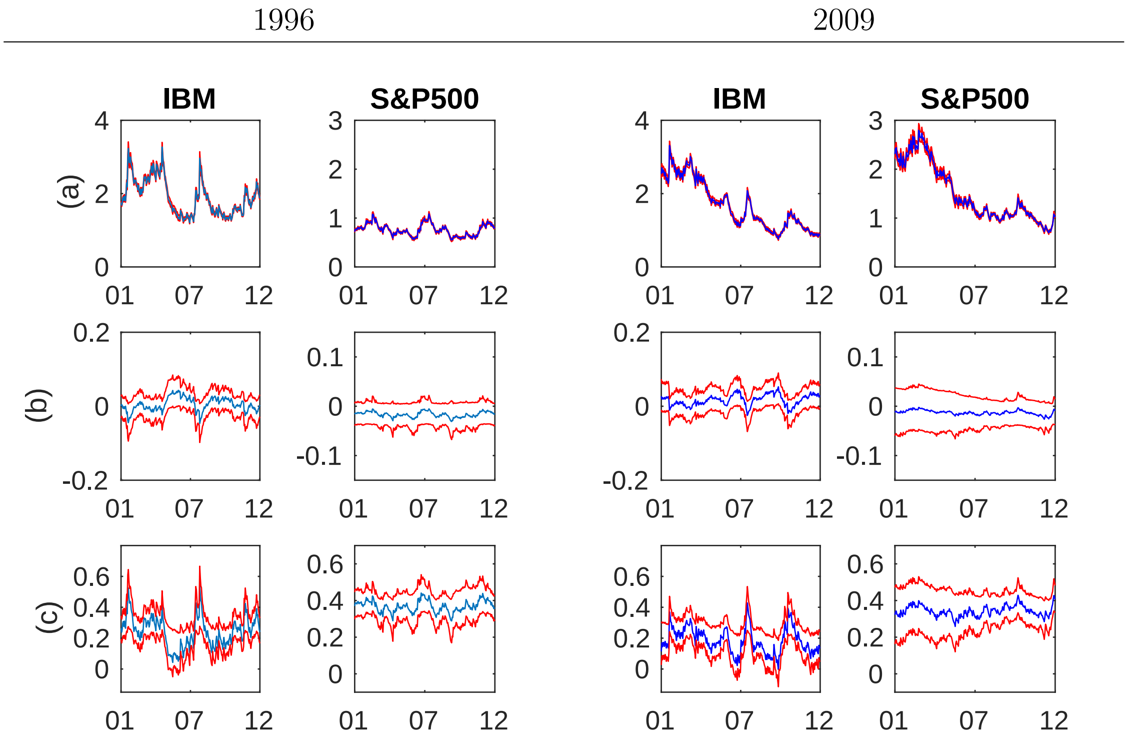

Figure 2: B-JSAV(1,1) robust quantile-based measures of the (a) scale, (b) skewness, and (c) kurtosis for IBM and S&P500 for $K = 10$ and $a = (0, 0.05, \ldots, 0.4, 0.45)$. We colour with red the 95% credible intervals.

data where the excess kurtosis rapidly increases to 0.5 during the 80's before rapidly falling to a low of -0.1 in the early 90's followed by a gradual increase to a level of around 0.2. The level is subsequently constant apart from some periods in 2014 when excess kurtosis falls to 0.1. The time variation in the S&P500 index is less clear (which is not surprising as this is an average of individual asset returns). The level is fairly constant at around 0.4. There is also evidence of time variation in kurtosis for the IBM data. Similar findings hold for the other specifications and can be found in Appendix D.

The results that we considered above cover a long period of 30 or 40 years and the plots of the conditional moments show trends in the volatility, skewness and kurtosis. However, these graphs do not illustrate the ability of the model to capture short-term changes in the higher moments, and so we consider inference for shorter periods, for example a year. We concentrate on periods that we expect to be extremely volatile such as that of 2009 following the financial crisis and that of 1996 during the Great Moderation. The results are given in Figure 2 for B-JSAV(1,1) for the three series. The results for the other 3 specifications can be found in Appendix D.

There is strong evidence of time-varying volatility in Figure 2 during 1996 and 2009. The time variation in the S&P500 index is clear now. The variation is particularly evident in 2009, where the volatility rapidly decreases by almost 2 units from the beginning to the end of the year. Similar findings hold for WTI data (which we don't present here). The year 1996 was a less volatile for S&P500 and WTI, but not for IBM. It seems that there was persistence (for example in January) with some short periods in which volatility increased rapidly (for example in February).

The skewness for both years is relatively constant for S&P500 and WTI, while there is some evidence of time-varying skewness for IBM with periods of both positive and negative skewness. More specifically, there are two periods of positive skewness and three periods of negative skewness throughout the year, with that during 1996 to be more volatile.

There is also evidence of time-varying kurtosis for all three datasets. If we turn to 1996 and IBM, we find that after the first 6 months there was a rapid drop by almost 0.3 units. The kurtosis stayed low until August when there was an increase which didn't last for a long time. The kurtosis then dropped again and stayed around 0.2 for most of the remaining months of 1996.

E Additional results

Here we include the plots of the data and the quantiles (Figure 3), and the plots of the robust quantile-based measures of the scale, skewness and kurtosis for the whole period as well as for 2009 and 1996 for B-JGJR(1,1) (Figures 4 and 5), B-JSSV(1,1) (Figures 6 and 7), and

B-JAVL(1,1) (Figures 8 and 9).

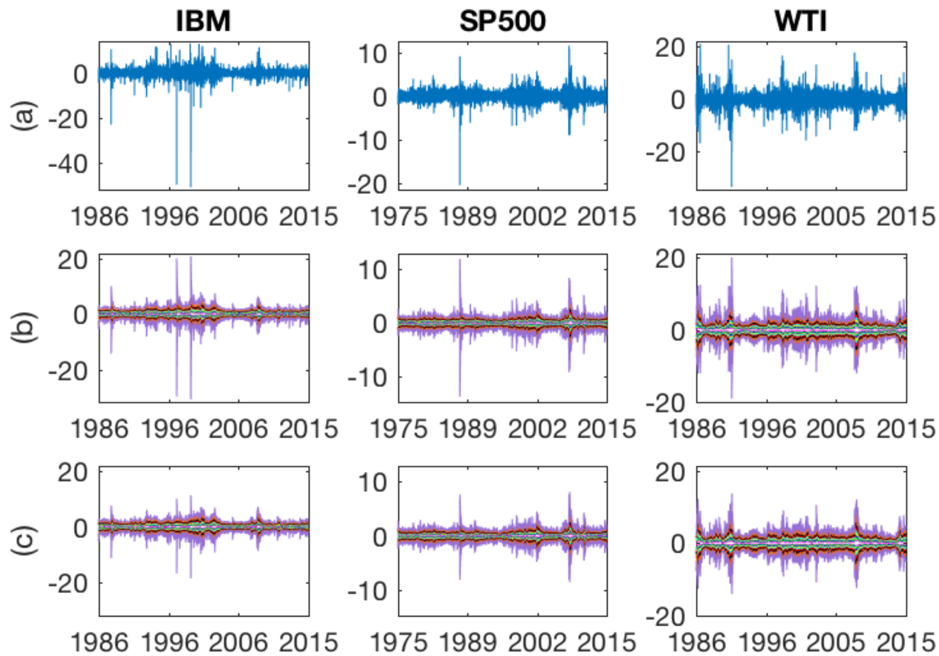

Figure 3: (a) Daily equity returns and posterior median of the conditional quantiles for (b) B-JGJR(1,1), and (c) B-JAVL(1,1) for IBM, S&P500, and WTI. We use different colours to depict the various quantiles for $K = 10$ probability levels and $a = (0, 0.05, \dots, 0.45, 0.5)$.

References

Atchadé, Y. F., G. O. Roberts, and J. S. Rosenthal. 2011. "Towards optimal scaling of Metropolis-coupled Markov chain Monte Carlo". *Statistics and Computing* 21:555–568.

Atchadé, Y., and J. S. Rosenthal. 2005. "On adaptive Markov chain Monte Carlo algorithms". *Bernoulli* 11:815–828.

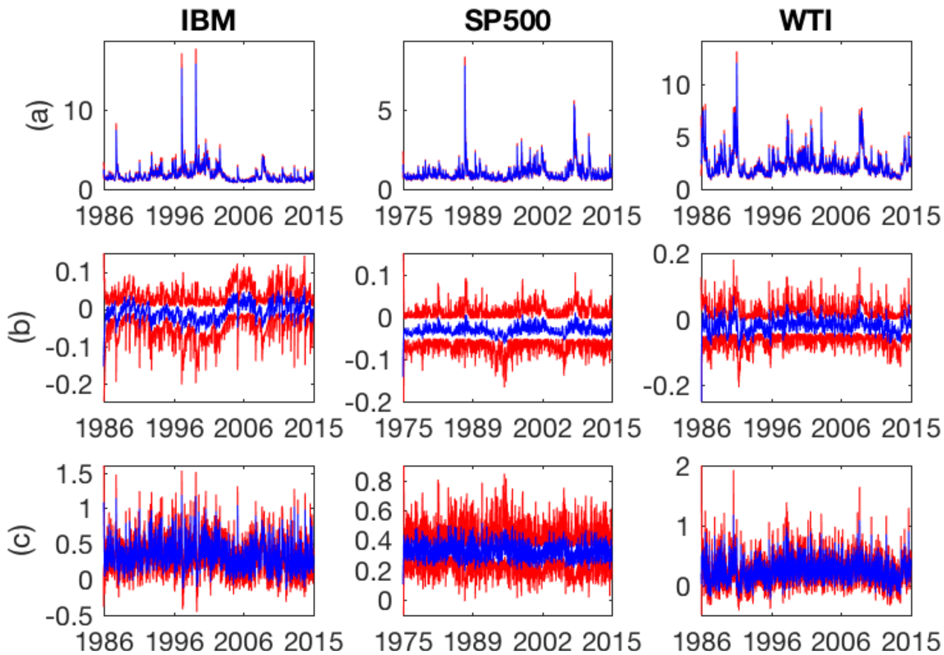

Figure 4: B-JGJR(1,1) robust quantile-based measures of the (a) scale, (b) skewness, and (c) kurtosis for IBM, S&P500, and WTI for $K = 10$ probability levels and $a = (0, 0.05, \ldots, 0.45, 0.5)$. We colour with red the 95% credible intervals.

Bougerol, P., and N. Picard. 1992. "Strict stationarity of Generalized Autoregressive Processes". *Annals of Probability* 20:1714–1730.

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.

Ding, J., and A. Zhou. 2007. "Eigenvalues of rank-one updated matrices with some applications". *Applied Mathematics Letters* 20:1223–1226.

Miasojedow, B., E. Moulines, and M. Vihola. 2013. "An adaptive parallel tempering algorithm". *Journal of Computational and Graphical Statistics* 22:649–664.

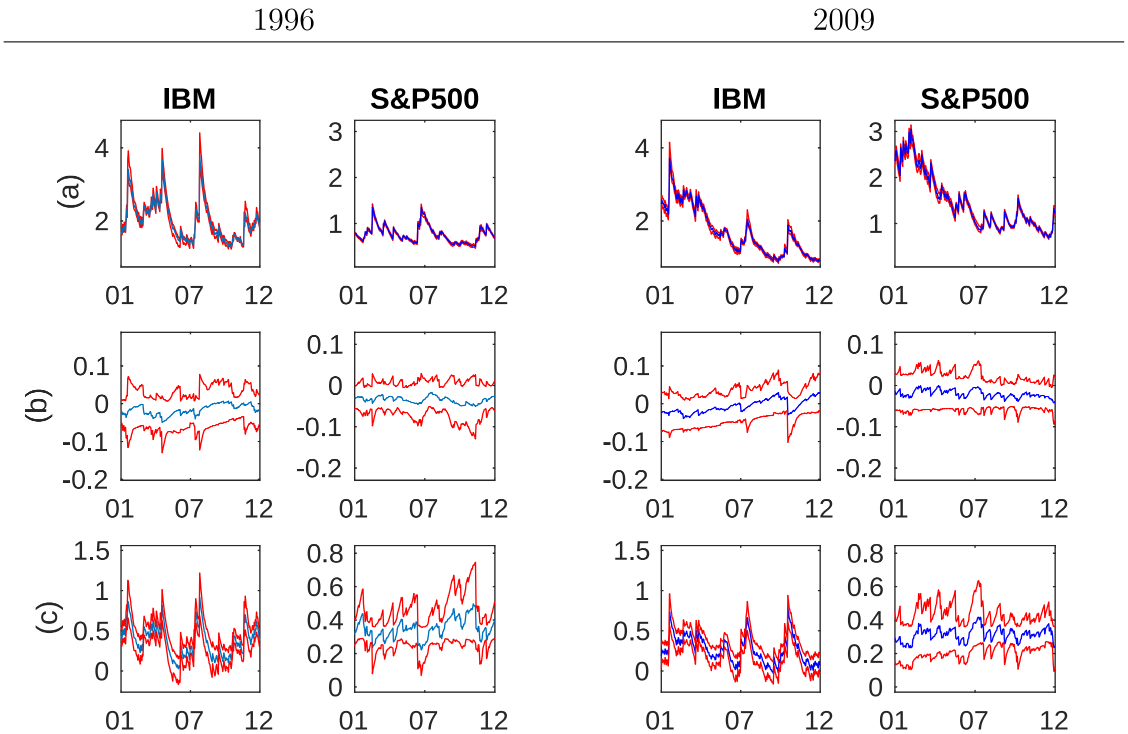

Figure 5: B-JGJR(1,1) robust quantile-based measures of the (a) scale, (b) skewness, and (c) kurtosis for IBM and S&P500 for $K = 10$ and $a = (0, 0.05, \ldots, 0.4, 0.45)$. We colour with red the 95% credible intervals.

Roberts, G. O., A. Gelman, and W. R. Gilks. 1997. "Weak convergence and optimal scaling of random walk Metropolis algorithms". *Annals of Applied Probability* 7:110–120.

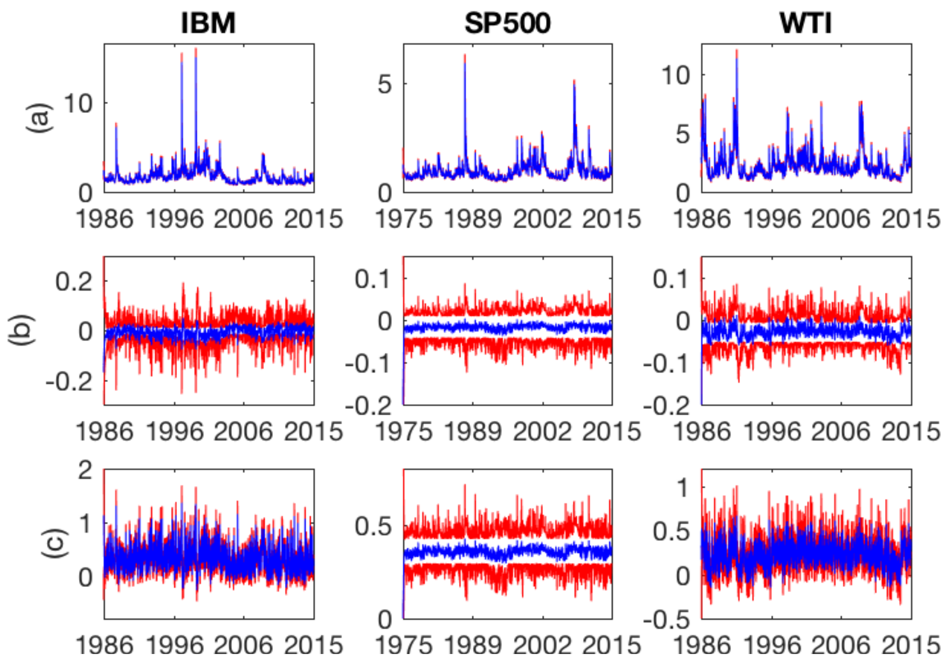

Figure 6: B-JSSV(1,1) robust quantile-based measures of the (a) scale, (b) skewness, and (c) kurtosis for IBM, S&P500, and WTI for $K = 10$ probability levels and $a = (0, 0.05, \ldots, 0.45, 0.5)$. We colour with red the 95% credible intervals.

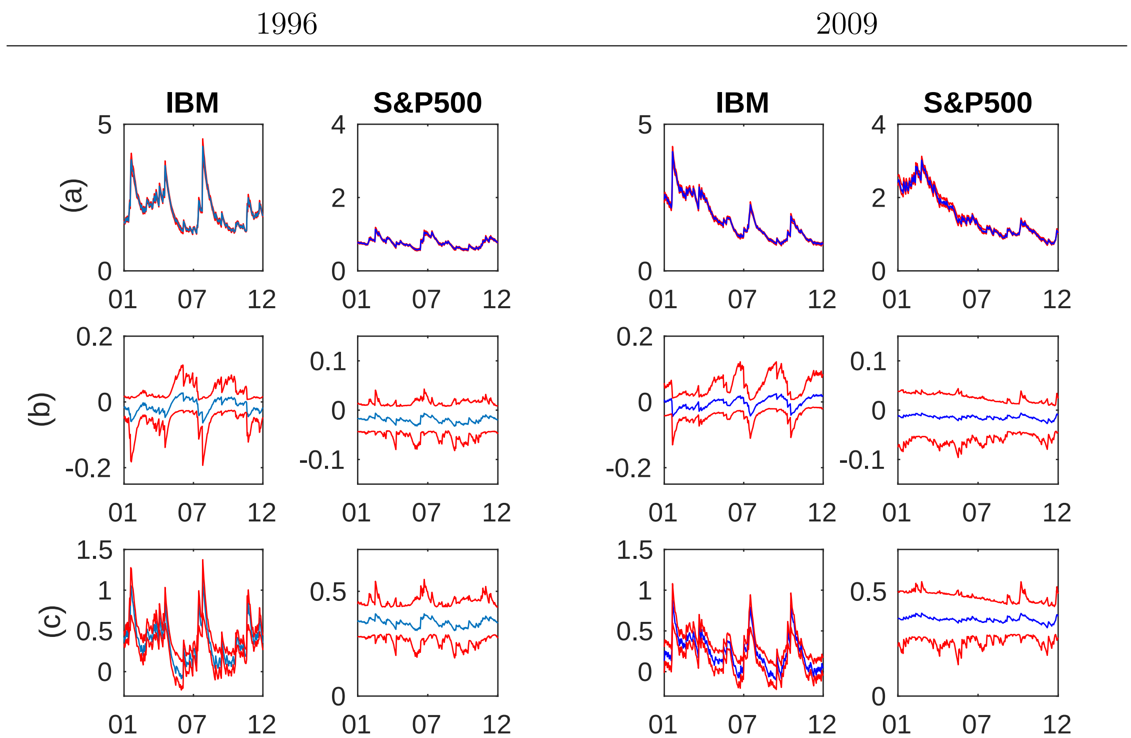

Figure 7: B-JSSV(1,1) robust quantile-based measures of the (a) scale, (b) skewness, and (c) kurtosis for IBM and S&P500 for $K = 10$ and $a = (0, 0.05, \ldots, 0.4, 0.45)$. We colour with red the 95% credible intervals.

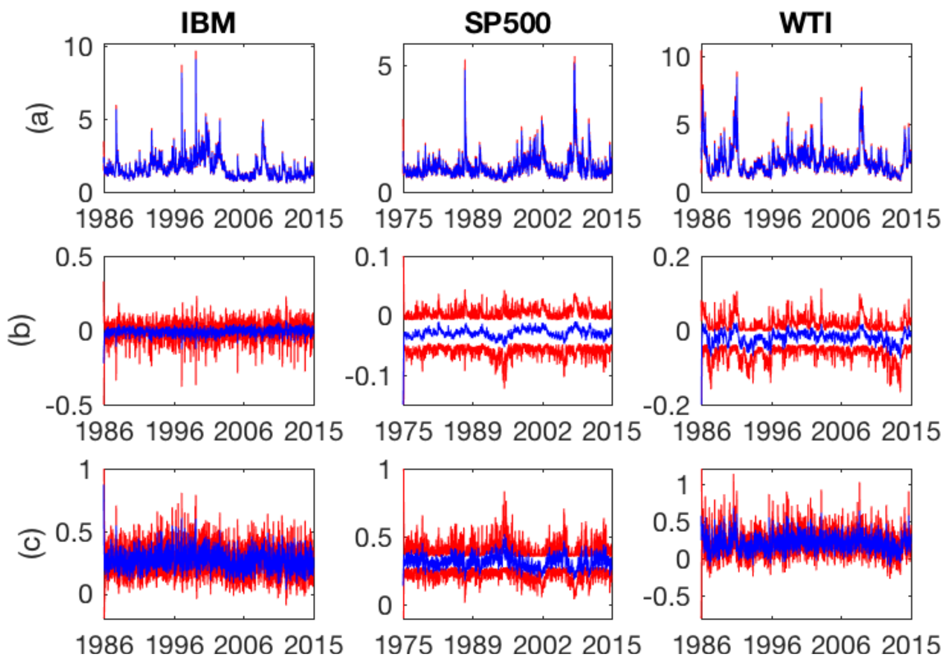

Figure 8: B-JAVL(1,1) obust quantile-based measures of the (a) scale, (b) skewness, and (c) kurtosis for IBM, S&P500, and WTI for $K = 10$ probability levels and $a = (0, 0.05, \ldots, 0.45, 0.5)$. We colour with red the 95% credible intervals.

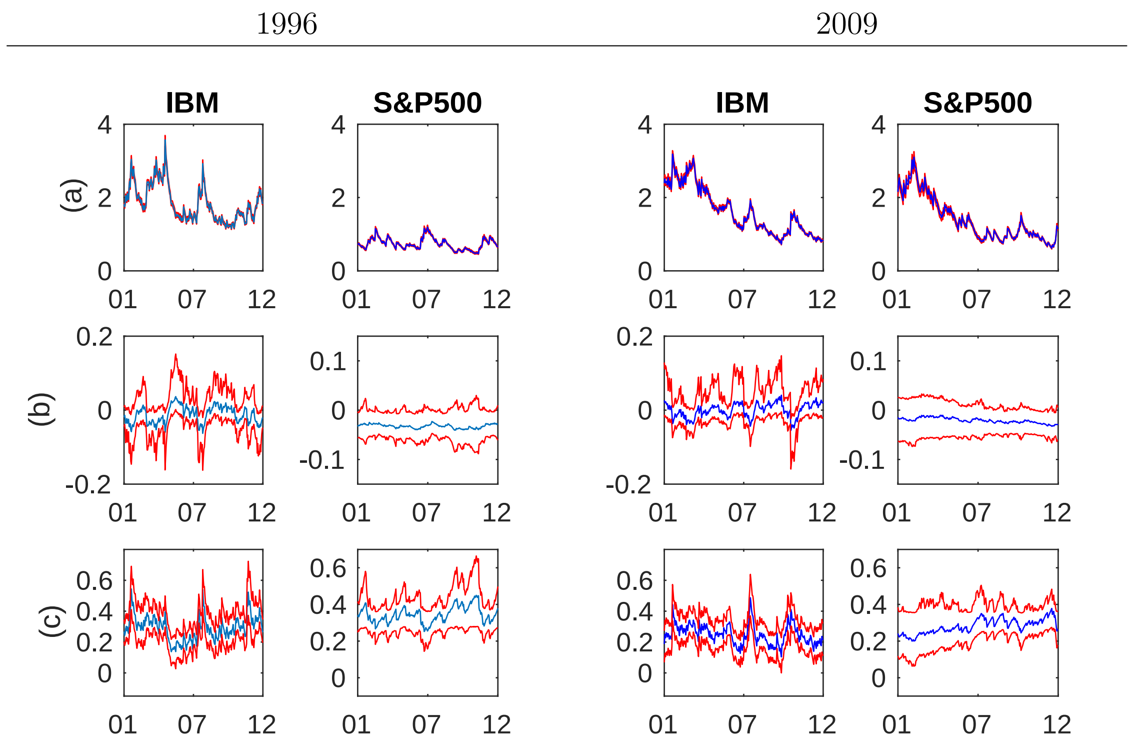

Figure 9: B-JSAVL(1,1) robust quantile-based measures of the (a) scale, (b) skewness, and (c) kurtosis for IBM and S&P500 for $K = 10$ and $a = (0, 0.05, \dots, 0.4, 0.45)$. We colour with red the 95% credible intervals.