2020-10

Working paper. Economics

ISSN 2340-5031

PREDICTION ACCURACY OF BIVARIATE SCORE-DRIVEN RISK PREMIUM AND VOLATILITY FILTERS: AN ILLUSTRATION FOR THE DOW JONES

Szabolcs Blazsek, Alvaro Escribano, and Adrian Licht.

uc3m | Universidad Carlos III de Madrid

Prediction accuracy of bivariate score-driven risk premium and volatility filters: an illustration for the Dow Jones

Szabolcs Blazsek [a], Alvaro Escribano [b,*], and Adrian Licht [a]

[a] School of Business, Universidad Francisco Marroquín, Guatemala City (Guatemala), Guatemala

[b] Department of Economics, Universidad Carlos III de Madrid, Getafe (Madrid), Spain

Abstract: In this paper, we introduce Beta-t-QVAR (quasi-vector autoregression) for the joint modelling of score-driven location and scale. Asymptotic theory of the maximum likelihood (ML) estimator is presented, and sufficient conditions of consistency and asymptotic normality of ML are proven. For the joint score-driven modelling of risk premium and volatility, Dow Jones Industrial Average (DJIA) data are used in an empirical illustration. Prediction accuracy of Beta-t-QVAR is superior to the prediction accuracies of Beta-t-EGARCH (exponential generalized AR conditional heteroscedasticity), A-PARCH (asymmetric power ARCH), and GARCH (generalized ARCH). The empirical results motivate the use of Beta-t-QVAR for the valuation of DJIA options.

Keywords: Volatility; Risk premium; Dynamic conditional score; Generalized autoregressive score

JEL Classification: C22, C58

Corresponding author. *E-mail address:* alvaroe@eco.uc3m.es (Alvaro Escribano).

1. Introduction

In this paper, Beta-t-QVAR (quasi-vector autoregression) for the joint modelling of risk premium and volatility is introduced, to improve the volatility prediction accuracies of Beta-t-EGARCH (exponential generalized AR conditional heteroscedasticity) (Harvey and Chakravarty 2008; Harvey 2013), A-PARCH (asymmetric power ARCH) (Ding et al. 1993), and GARCH (generalized ARCH) (Engle 1982; Bollerslev 1986). The Beta-t-QVAR model is a score-driven model (Creal et al. 2008, 2011, 2013; Harvey and Chakravarty 2008; Harvey 2013) for the Student's t-distribution, in which dynamic interaction effects between risk premium and volatility are measured by a bivariate score-driven filter. Score-driven models are observation-driven models (Cox 1981), which are updated by the partial derivative of the log conditional density of the dependent variable with respect to dynamic parameters (hereinafter, score function). The following contributions to the literature are provided in this paper:

Firstly, Beta-t-QVAR, in which score-driven location and score-driven scale are jointly modelled, is new in the literature, to the best of our knowledge. In Beta-t-QVAR for risk premium and volatility, score-driven filters simultaneously update risk premium and volatility. The model measures dynamic overreaction effects, dynamic risk premium effects, dynamic leverage effects, and dynamic volatility effects, which are represented by impulse response functions (IRFs). Beta-t-QVAR extends the Beta-t-EGARCH (Harvey and Chakravarty 2008) and the one-component Beta-t-EGARCH-M (Harvey and Lange 2018) models.

Secondly, the conditions of the asymptotic properties of the maximum likelihood (ML) estimates of Beta-t-QVAR are new in the literature, to the best of our knowledge. It is shown that all score functions are martingale difference sequences (MDSs), the bivariate score-driven filter converges to a unique strictly stationary and ergodic sequence, and all score functions and their derivatives have finite second moments and covariances. The ML conditions imply that the gradient and the Hessian of the ML have time-invariant expected values, and they converge to unique strictly stationary and ergodic sequences. Monte Carlo simulation-based estimation results are reported, to support the use of ML. The results of Creal et al. (2013), Harvey (2013), and Blasques et al. (2017, 2018) are extended in the present paper, because in those works ML conditions are presented for either score-driven location models or for score-driven scale models.

Thirdly, in an empirical illustration, weekly and daily data from the Dow Jones Industrial Average (DJIA) for the period of January 1985 to February 2020 are used. Out-of-sample forecasts of volatility are computed, by using the rolling-window approach, for the period of January 2010 to February 2020. We find that Beta-t-QVAR improves the performances of Beta-t-EGARCH, A-PARCH, and GARCH, which are found to be effective predictors of volatility in the literature (Hansen and Lunde 2005; Blazsek and Villatoro 2015). The empirical illustration suggests that Beta-t-QVAR may be used for DJIA options valuation at the Chicago Board Options Exchange (CBOE).

The remainder of this paper is organized as follows: Section 2 reviews the literature. Section 3 presents Beta-t-QVAR. Section 4 describes the dataset. Section 5 presents the statistical inference methods. Section 6 presents the impulse response functions. Section 7 presents the results on DJIA volatility forecasts. Section 8 concludes. Technical details are presented in Supplementary Material.

2. Review of the literature

2.1. Classical dynamic volatility models

The subject matter of this paper is in relation to several classical dynamic volatility models from the body of literature. In the work of Engle (1982), the ARCH(q) model of dynamic volatility is introduced, which is extended in the works of Bollerslev (1986, 1987) to the Gaussian-GARCH(p,q) and t-GARCH(p,q) models, respectively. In the work of Engle et al. (1987), the ARCH-in-mean (ARCH-M) model is introduced, which extends ARCH by including the conditional standard deviation of returns in the risk premium component.

In the work of Nelson (1991), the EGARCH(p,q) model is introduced, in which the dynamics of the log conditional variance of returns are formulated, and leverage effects (Black 1976) are included in the log conditional variance equation. In the work of Ding et al. (1993), the A-PARCH(p,q) model for the power $\delta \geq 0$ of the conditional standard derivation is introduced, which generalizes the ARCH and GARCH models, approximates the long memory property of stock market returns, and includes leverage effects in the filter driving conditional volatility. In the work of Glosten et al. (1993), the GARCH model is extended, by including leverage effects in the conditional variance equation.

In the present work, Student's t-distribution is assumed for the probability distribution of returns (Bollerslev 1987), EGARCH volatility formulation is used (Nelson 1991), a filter driving conditional volatility is included in the risk premium (EGARCH-M) (Engle et al. 1987), and leverage effects are included in the volatility filter (Nelson 1991; Ding et al. 1993; Glosten et al. 1993).

2.2. Score-driven volatility models

Score-driven volatility models have recently been introduced to the literature on dynamic volatility models. The first score-driven volatility model is the Beta-t-EGARCH model (Harvey and Chakravarty 2008), which assumes the Student's t-distribution for financial returns. In relation to Beta-t-EGARCH, we also refer to the works of Creal et al. (2013) and Harvey (2013). Furthermore, we also refer to an important recent score-driven volatility model that is extended in the present work, the one-component Beta-t-EGARCH-M model of Harvey and Lange (2018), in which the score-driven volatility filter is included in the risk premium component of returns.

In the work of Blasques et al. (2015), it is shown for univariate score-driven filters, such as Beta-t-EGARCH, that a score-driven update of the time series model, asymptotically and in expectation, reduces the Kullback–Leibler divergence in favour of the true data generating process at every step. The authors also show that only score-driven updates have this property. The work of Blasques et al. (2020) presents simulation-based results for finite samples, which support the use of Beta-t-EGARCH.

Alternatives to Beta-t-EGARCH are: GED-EGARCH (general error distribution EGARCH) (Harvey 2013); Beta-Skew-t-EGARCH (skewed t-distribution EGARCH) (Harvey and Sucarrat 2014); EGB2-EGARCH (exponential generalized beta distribution of the second kind EGARCH) (Caivano and Harvey 2014); Beta-Skew-Gen-t-EGARCH (skewed generalized t-distribution EGARCH) (Harvey and Lange 2017); NIG-EGARCH (normal-inverse Gaussian distribution EGARCH) (Blazsek et al. 2018); MXN-EGARCH (Meixner distribution EGARCH) model (Blazsek and Haddad 2020).

In this section, we also refer to the following works from the literature, in which practically relevant applications of score-driven expected return plus volatility models are presented: Blazsek and Mendoza (2016), Blazsek et al. (2016), Blazsek and Monteros (2017a, 2017b), Blazsek and Hernandez (2018), Blazsek et al. (2018), Ayala and Blazsek (2018a, 2018b, 2019a, 2019b), and Blazsek and Licht (2018, 2020). Sufficient conditions of ML are not presented in those works, which motivates the development of the ML conditions for Beta-t-QVAR in the present paper.

2.3. Multivariate score-driven filters

In the work of Harvey (2013), the dynamic conditional score (DCS) model for the multivariate t-distribution is introduced, which is abbreviated as t-QVAR(1), for the modelling of $I(0)$ or co-integrated $I(1)$ variables (Granger 1981; Engle and Granger 1987). In the work of Creal et al. (2014), t-QVAR(1) is extended to t-QVARMA(p, q). In the recent works of Blazsek and Licht (2020) and Blazsek et al. (2020), applications of t-QVAR(1) and t-QVARMA(p, q), respectively, are presented. In the work of Blazsek et al. (2019), t-QVARMA(p,q,r) is introduced, extending the score-driven location models of

Harvey (2013) and Creal et al. (2014). The t-QVARMA(p,q,r) model is for a joint modelling of $I(0)$ and co-integrated $I(1)$ variables. In the present work, the multivariate Beta-t-QVAR(1) filter is used for risk premium and volatility, by using the univariate Student's t-distribution.

3. Econometric model

The score-driven model of log-return, $y_t = 100 \times \ln(s_t/s_{t-1})$ for $t = 1, \ldots, T$, is:

$$y_t = \mu_t + v_t = \mu_t + \exp(\lambda_t)\epsilon_t = \mu_t + \exp(\omega + \lambda_t^\dagger)\epsilon_t \tag{1}$$

where pre-sample data define s_0, μ_t is a dynamic parameter representing the risk premium, v_t is the unexpected return, and $\exp(\lambda_t) \equiv \exp(\omega + \lambda_t^\dagger)$ is the dynamic scale parameter with time-invariant parameter ω and filter $\lambda_t^\dagger$. The error term $\epsilon_t \sim t(\nu)$, with degrees of freedom $2 < \nu < \infty$, is independent and identically distributed (i.i.d.) with respect to the Student's t-distribution. Hence, the second moment of ϵ_t is finite, and the conditional volatility for period t is $\sigma_t = \exp(\lambda_t)[\nu/(\nu-2)]^{1/2}$.

In the work of Harvey and Lange (2018), for the one-component score-driven EGARCH-M model, risk premium is specified as $\mu_t = c + \beta_2 \exp(\lambda_t)$, where c is a time-invariant parameter, $\beta_2 \exp(\lambda_t)$ represents the in-mean (M) component, and λ_t is formulated according to a score-driven EGARCH model. In the present paper, two alternatives are considered, named Beta-t-QVAR(1) and Beta-t-QVAR(1)-M, respectively, in which the risk premium is specified as follows:

$$\mu_t = c + \beta_1 \mu_t^\dagger \tag{2}$$

$$\mu_t = c + \beta_1 \mu_t^\dagger + \beta_2 \exp(\lambda_t) = c + \beta_1 \mu_t^\dagger + \beta_2 \exp(\omega + \lambda_t^\dagger) \tag{3}$$

where $\mu_t^\dagger$ is a new filter in the score-driven EGARCH. The latter specification is an extension of the one-component model of Harvey and Lange (2018), by adding $\beta_1 \mu_t^\dagger$ into risk premium. In Eqs. (1)-(3), $E(\lambda_t) = \omega$ and $E(\mu_t^\dagger) = E(\lambda_t^\dagger) = 0$. The $\beta_1 \mu_t^\dagger$ term in the risk premium measures an extra serial correlation in the mean of the log-return on a financial asset, in addition to $\beta_2 \exp(\omega + \lambda_t^\dagger)$.

The following updating mechanism is used for $\mu_t^\dagger$ and $\lambda_t^\dagger$:

$$\begin{bmatrix} \mu_t^\dagger \\ \lambda_t^\dagger \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \mu_{t-1}^\dagger \\ \lambda_{t-1}^\dagger \end{bmatrix} + \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix} \begin{bmatrix} u_{\mu,t-1} \\ u_{\lambda,t-1} \end{bmatrix} \tag{4}$$

where updating terms $u_{\mu,t}$ and $u_{\lambda,t}$ are named score functions. Updating term $u_{\mu,t}$ is the scaled conditional score of the LL (log-likelihood) with respect to μ_t. Updating term $u_{\lambda,t}$ is the conditional score of the LL with respect to λ_t. Both $u_{\mu,t}$ and $u_{\lambda,t}$ are MDSs with zero mean and finite variance. Hence, $u_{\mu,t}$ and $u_{\lambda,t}$ are white noise variables (Harvey 2013, p. 6, Definition 1). Formal definitions and stochastic properties of $u_{\mu,t}$ and $u_{\lambda,t}$ are presented in Section 5.2.

In matrix form, Eq. (4) can be written as:

$$\theta_t = \Phi\theta_{t-1} + \Psi u_{t-1} \tag{5}$$

where $\theta_t = (\theta_{1,t}, \theta_{2,t})' \equiv (\mu_t^\dagger, \lambda_t^\dagger)'$ and $u_t = (u_{1,t}, u_{2,t})' \equiv (u_{\mu,t}, u_{\lambda,t})'$. In Eq. (5), a first-order specification, i.e. Beta-t-QVAR(1), is assumed, which can be extended to Beta-t-QVARMA(p,q) in future works. Moreover, it is assumed that the maximum modulus of eigenvalues of Φ is less than one. Filter θ_t is initialized by its deterministic unconditional mean: $\theta_1 = E(\theta_t) = 0_{2\times 1}$.

The work of Hansen and Lunde (2005) motivates the use of leverage effects for forecasting the volatility of stock returns. Therefore, we include leverage effects in θ_t in the following filter:

$$\begin{bmatrix} \mu_t^\dagger \\ \lambda_t^\dagger \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \mu_{t-1}^\dagger \\ \lambda_{t-1}^\dagger \end{bmatrix} + \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix} \begin{bmatrix} u_{\mu,t-1} \\ u_{\lambda,t-1} \end{bmatrix} + \psi^* \begin{bmatrix} 0 \\ \mathrm{sgn}(-\epsilon_{t-1})(u_{\lambda,t-1}+1) \end{bmatrix} \tag{6}$$

where parameter $\psi^* \in \mathbb{R}$ measures leverage effects, $\mathrm{sgn}(\cdot)$ is the signum function, and the leverage effects formulation is from Harvey (2013, p. 105). In matrix form, Eq. (6) is:

$$\theta_t = \Phi\theta_{t-1} + g(u_{t-1}) \tag{7}$$

where $\theta_t = (\theta_{1,t}, \theta_{2,t})' \equiv (\mu_t^\dagger, \lambda_t^\dagger)'$. Moreover, $g(u_t) = g[(u_{1,t}, u_{2,t})'] \equiv g[(u_{\mu,t}, u_{\lambda,t})']$ represents the second and third terms on the right side of Eq. (6), and $\theta_1 = E(\theta_t) = 0_{2\times 1}$ is used for initialization. In Eq. (7), we assume that the maximum modulus of eigenvalues of Φ is less than one. The Beta-t-

QVAR(1) and Beta-t-QVAR(1)-M specifications with leverage effects are named Beta-t-QVAR(1)-lev and Beta-t-QVAR(1)-M-lev, respectively.

We note that for Beta-t-QVAR(1)-M and Beta-t-QVAR(1)-M-lev, the most recent information on asset return influences the risk premium in two ways: (i) through $\mu_t^\dagger$, which is updated by a linear transformation of u_t; (ii) through $\exp(\omega + \lambda_t^\dagger)$, which is updated by a nonlinear transformation of u_t. The differences between those updates are the functional forms of the updates and the parameters that scale u_t. Details of the model formulation for all Beta-t-QVAR(1) specifications, by showing the dynamics of risk premium and volatility with explicit formulas, are presented in the Supplementary Material.

4. Data

Weekly and daily log-return data are used from the DJIA stock market index for the period of January 1985 to February 2020 (source: Yahoo Finance). In this paper, gross return is used instead of excess return over a risk-free rate, since it is assumed that the forecast users of the empirical illustration are DJIA options investors. Weekly data frequency is motivated by the volatility forecast study for Beta-t-EGARCH-M in the work of Harvey and Lange (2018). Daily data frequency is motivated by the volatility forecast study for GARCH-type volatility models in the work of Hansen and Lunde (2005).

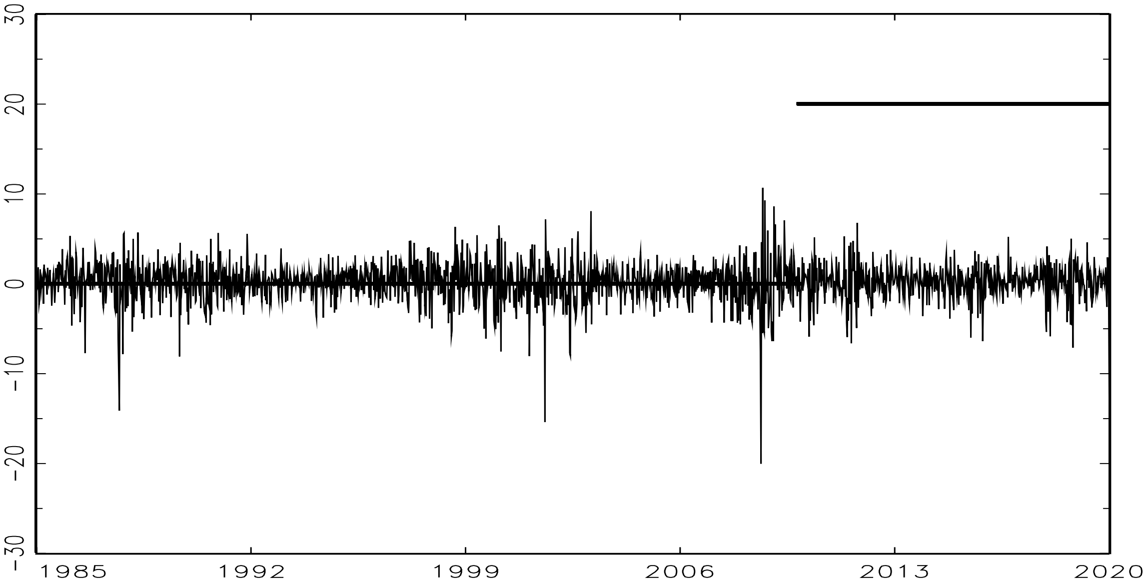
In the volatility forecasting case study of the present paper, the full data window is divided into the initial estimation and the forecasting windows (Table 1). A rolling-window approach is used for forecasting. After the use of the initial estimation window, for the remaining estimation windows, the first observation is excluded from the estimation window and a new last observation is added to the estimation window. Descriptive statistics are presented in Table 1. The evolution of weekly and daily DJIA log-returns is presented for the period of January 1985 to February 2020 in Fig. 1.

Table 1. Descriptive statistics of DJIA log-return y_t (% points).

Full data window:	Weekly	Daily
Start date	28 January 1985	31 January 1985
End date	24 February 2020	27 February 2020
Sample size	$1,830$	$8,840$
Forecasting window:	Weekly	Daily
Start date	4 January 2010	4 January 2010
End date	24 February 2020	27 February 2020
Sample size	530	$2,555$
Initial estimation window:	Weekly	Daily
Start date	28 January 1985	31 January 1985
End date	28 December 2009	31 December 2009
Sample size	$1,300$	$6,285$
Final estimation window:	Weekly	Daily
Start date	27 March 1995	10 March 1995
End date	17 February 2020	26 February 2020
Sample size	$1,300$	$6,285$
Statistics:	Weekly	Daily
Minimum	−20.0298	−25.6315
Maximum	10.6977	10.5083
Average	0.1634	0.0339
Standard deviation	2.2516	1.0956
Skewness	−1.0302	−1.6607
Excess kurtosis	7.2389	41.4272
Corr(y_t, y_{t-1})	−0.0624	−0.0330
Corr$(y_t, \|y_{t-1}\|)$	−0.0149	0.0302
Corr$(\|y_t\|, y_{t-1})$	−0.1996	−0.1242
Corr$(\|y_t\|, \|y_{t-1}\|)$	0.2610	0.2319
Shapiro–Wilk test	0.9382***	0.8779***

Notes: The null hypothesis of the Shapiro and Wilk (1965) test is normal distribution. *** indicates significance at the 1% level.

(a) Weekly DJIA log-return
Initial estimation window: 28 January 1985 to 28 December 2009 ($T = 1,300$)
Forecasting window: 4 January 2010 to 24 February 2020 ($T_f = 530$).



(b) Daily DJIA log-return
Initial estimation window: 31 January 1985 to 31 December 2009 ($T = 6,285$)
Forecasting window: 4 January 2010 to 27 February 2020 ($T_f = 2,555$).

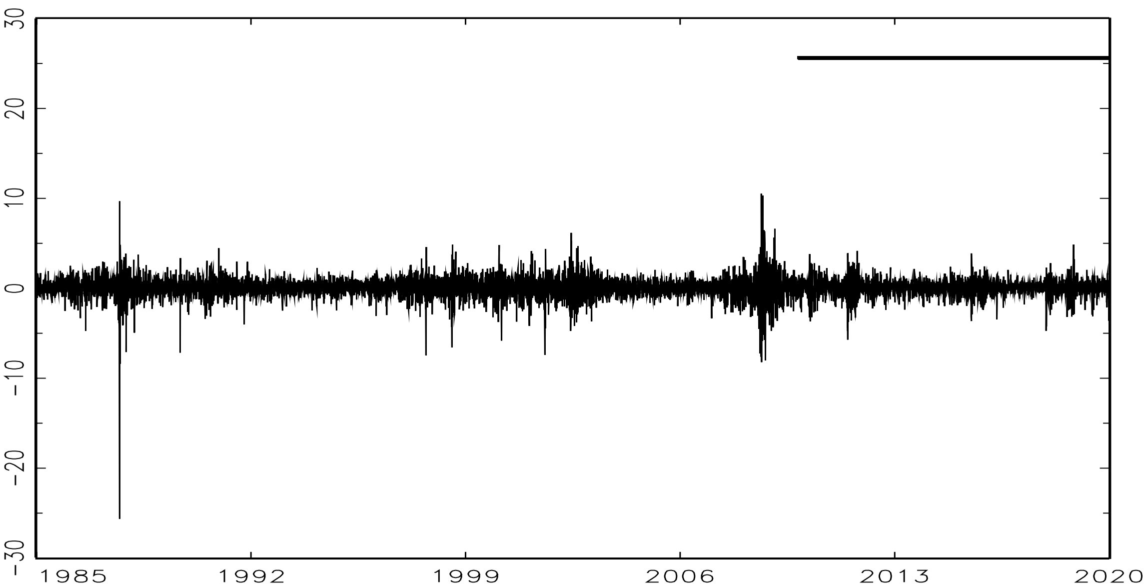


Fig. 1. DJIA log-return $y_t = 100 \times \ln(p_t/p_{t-1})$ for the period of January 1985 to February 2020.
Notes: In each panel, a tick constant line indicates the forecasting window.

5. Statistical inference

5.1. ML estimator

The parameters of Beta-t-QVAR are estimated by using the ML method, as follows:

$$\hat{\Theta} = \arg\max_{\Theta} \mathrm{LL}(y_1, \ldots, y_T|\Theta) = \arg\max_{\Theta} \sum_{t=1}^{T} \ln f(y_t|\mathcal{F}_{t-1}, \Theta) \tag{8}$$

where Θ is a $S \times 1$ vector of time-invariant parameters, the information set is $\mathcal{F}_{t-1} = (\theta_1, y_1, \ldots, y_{t-1})$ (Blasques et al. 2017), and the log conditional density of $y_t|\mathcal{F}_{t-1}$ is:

$$\ln f(y_t|\mathcal{F}_{t-1}, \Theta) = \tag{9}$$

$$\ln\Gamma\left(\frac{\nu+1}{2}\right) - \ln\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\ln(\pi\nu) - \lambda_t - \frac{\nu+1}{2}\ln\left[1 + \frac{(y_t-\mu_t)^2}{\nu\exp(2\lambda_t)}\right]$$

For the Beta-t-QVAR(1) model, $\Theta = [c, \beta_1, \mathrm{vec}(\Phi)', \mathrm{vec}(\Psi)', \omega, \nu]'$, where $S = 12$. For the Beta-t-QVAR(1)-M model, $\Theta = [c, \beta_1, \beta_2, \mathrm{vec}(\Phi)', \mathrm{vec}(\Psi)', \omega, \nu]'$, where $S = 13$. For the Beta-t-QVAR(1)-lev model, $\Theta = [c, \beta_1, \mathrm{vec}(\Phi)', \mathrm{vec}(\Psi)', \psi^*, \omega, \nu]'$, where $S = 13$. For the Beta-t-QVAR(1)-M-lev model, $\Theta = [c, \beta_1, \beta_2, \mathrm{vec}(\Phi)', \mathrm{vec}(\Psi)', \psi^*, \omega, \nu]'$, where $S = 14$. Within Θ, $\mathrm{vec}(x)$ indicates that the columns of matrix x are being stacked one upon the other. With respect to the parameter set for Θ, we assume that $2 < \nu < \infty$, and each of the remaining parameters take finite values within the set of real numbers $\mathbb{R}$. As a consequence, $-\infty < \lambda_t < \infty$, and $-\infty < \mu_t < \infty$. We use the following assumption:

(A1) Parameter set $\tilde{\Theta} \subset \mathbb{R}^S$ is compact (Wooldridge 1994, Theorem 4.1).

In the following, the gradient vector $G_t(\Theta)$ and the Hessian matrix $H_t(\Theta)$ of LL are defined. The $T \times S$ matrix of contributions to the gradient $G(y_1, \ldots, y_T, \Theta)$ is defined by its elements:

$$G_{t,i}(\Theta) = -\frac{\partial \ln f(y_t|\mathcal{F}_{t-1}; \Theta)}{\partial \Theta_i} \tag{10}$$

for period $t = 1, \ldots, T$, and parameter $i = 1, \ldots, S$ (Wooldridge 1994, p. 2674). The t-th row of $G(y_1, \ldots, y_T, \Theta)$ is denoted by using $G_t(\Theta)$, which is the score vector for the t-th observation. Under the ML assumptions of this paper, the maximization problem of Eq. (8) is equivalent to:

$$\frac{1}{T}\sum_{t=1}^{T} G_t(\hat{\Theta})' = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} G_{t,1}(\hat{\Theta}) \\ \vdots \\ G_{t,S}(\hat{\Theta}) \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} -\frac{\partial \ln f(y_t|\mathcal{F}_{t-1};\hat{\Theta})}{\partial \Theta_1} \\ \vdots \\ -\frac{\partial \ln f(y_t|\mathcal{F}_{t-1};\hat{\Theta})}{\partial \Theta_S} \end{bmatrix} = 0_{S\times 1} \tag{11}$$

According to the mean-value expansion about the true values of parameters Θ_0:

$$\frac{1}{T}\sum_{t=1}^{T} G_t(\hat{\Theta})' = \frac{1}{T}\sum_{t=1}^{T} G_t(\Theta_0)' + \frac{1}{T}\left[\sum_{t=1}^{T} H_t(\bar{\Theta})\right](\hat{\Theta}-\Theta_0) \tag{12}$$

where each row of the $S \times S$ Hessian matrix (Wooldridge 1994, p. 2674):

$$H_t(\Theta) = -\frac{\partial^2 \ln f(y_t|\mathcal{F}_{t-1};\Theta)}{\partial \Theta \partial \Theta'} \tag{13}$$

is evaluated at S different mean values $\bar{\Theta}$ (Eq. (12)). Each $\bar{\Theta}$ is located between Θ_0 and $\hat{\Theta}$: $||\bar{\Theta}-\Theta_0|| \leq ||\hat{\Theta}-\Theta_0||$, where $||\cdot||$ is the Euclidean norm. From Eqs. (11) and (12):

$$\sqrt{T}(\hat{\Theta}-\Theta_0) = \left[-\frac{1}{T}\sum_{t=1}^{T} H_t(\bar{\Theta})\right]^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} G_t(\Theta_0)'\right] \tag{14}$$

The asymptotic covariance matrix of parameters $\hat{\Theta}$ is estimated by using the inverse information matrix: $\{(1/T)\sum_{t=1}^{T}[G_t(\hat{\Theta})'G_t(\hat{\Theta})]\}^{-1}$ (Creal et al. 2013; Harvey 2013; Blasques et al. 2017).

In the proofs of the stochastic properties of the score functions and in the asymptotic theory of the ML estimator, the following maintained assumptions (A2) and (A3) are used:

(A2) $f(y_t|\mathcal{F}_{t-1};\Theta_0) = p_0(y_t|\mathcal{F}_{t-1};\Theta_0)$ for Θ_0 from the parameter set $\tilde{\Theta} \subset \mathbb{R}^S$, where p_0 is the true conditional density, and Θ_0 represents the true values of Θ (Wooldridge 1994, p. 2672).

(A3) $f(y_t|\mathcal{F}_{t-1};\Theta_0) = p_0(y_t|\mathcal{F}_{t-1};\Theta_0)$ for Θ_0 is a dynamically complete density (Wooldridge 1994, p. 2677; Woodridge 2010, p. 408).

One of the consequences of (A3) is that $G_t(\Theta_0)'$ is a MDS (Wooldridge 1994, p. 2677). Further assumptions for the asymptotic theory of ML are presented in Section 5.3.

5.2. Score functions of dynamic parameters

In this section, we present the properties of score functions $u_{\mu,t}$ and $u_{\lambda,t}$ at the true values of parameters $\Theta = \Theta_0$. We use the following maintained assumption for the log-scale parameter:

(A4) $|\lambda_t| < \lambda_{\max} < \infty, \qquad \forall t$

where $\lambda_{\max}$ is a finite parameter that does not depend on ϵ_t for all t. (A4) assumes that volatility cannot go to infinity, given that $\nu > 2$. Several properties of the score functions are based on this assumption.

Updating term $u_{\mu,t}$ is the scaled conditional score of the LL with respect to μ_t (Harvey 2013):

$$\frac{\partial \ln f(y_t|\mathcal{F}_{t-1}, \Theta)}{\partial \mu_t} = \frac{\nu + 1}{\nu \exp(2\lambda_t)} \times u_{\mu,t} \tag{15}$$

$$u_{\mu,t} = \frac{\nu \exp(2\lambda_t)(y_t - \mu_t)}{\nu \exp(2\lambda_t) + (y_t - \mu_t)^2} = \frac{\nu \exp(\lambda_t)\epsilon_t}{\nu + \epsilon_t^2} \tag{16}$$

is a continuously differentiable function of ϵ_t. In the last equation we replace λ_t by $\lambda_{\max}$. Hence,

$$|u_{\mu,t}| = \left| \frac{\nu \exp(\lambda_t)\epsilon_t}{\nu + \epsilon_t^2} \right| < \left| \frac{\nu \exp(\lambda_{\max})\epsilon_t}{\nu + \epsilon_t^2} \right| \tag{17}$$

The right side of Eq. (17) includes a continuously differentiable function of ϵ_t, and if $|\epsilon_t| \to \infty$, hence the right side of Eq. (17) goes to zero. Therefore, $u_{\mu,t}$ is a bounded function of ϵ_t.

Updating term $u_{\lambda,t}$ is the conditional score of the LL with respect to λ_t (Harvey 2013, p. 99):

$$u_{\lambda,t} = \frac{\partial \ln f(y_t|\mathcal{F}_{t-1}, \Theta)}{\partial \lambda_t} = \frac{(\nu + 1)(y_t - \mu_t)^2}{\nu \exp(2\lambda_t) + (y_t - \mu_t)^2} - 1 = \frac{(\nu + 1)\epsilon_t^2}{\nu + \epsilon_t^2} - 1 \tag{18}$$

If $\nu > 0$, then $u_{\lambda,t}$ is a continuously differentiable and bounded function of ϵ_t. Variable $|u_{\lambda,t}|$ is bounded because $u_{\lambda,t}$ is a transformation of the random variable $[\epsilon_t^2/(\nu + \epsilon_t^2)] \in (0, 1)$ (Harvey 2013).

In the following, we study the boundedness of the derivatives of score functions with respect to μ_t and λ_t. The derivative of $u_{\mu,t}$ with respect to μ_t is:

$$\frac{\partial u_{\mu,t}}{\partial \mu_t} = \frac{\nu \epsilon_t^2 - \nu^2}{(\nu + \epsilon_t^2)^2} \tag{19}$$

which is a continuously differentiable function of $\epsilon_t = (y_t - \mu_t)\exp(-\lambda_t)$. Hence,

$$\left|\frac{\partial u_{\mu,t}}{\partial \mu_t}\right| = \left|\frac{\nu\epsilon_t^2 - \nu^2}{(\nu + \epsilon_t^2)^2}\right| < \left|\frac{\nu\epsilon_t^2 - \nu^2}{\nu + \epsilon_t^2}\right| < \left|\frac{\nu\epsilon_t^2}{\nu + \epsilon_t^2}\right| < \infty \tag{20}$$

where the first inequality is true for $\nu > 1$, and the last inequality is true because $[\epsilon_t^2/(\nu + \epsilon_t^2)] \in (0, 1)$. As a consequence, $\partial u_{\mu,t}/\mu_t$ is a bounded function of ϵ_t. The derivative of $u_{\mu,t}$ with respect to λ_t is:

$$\frac{\partial u_{\mu,t}}{\partial \lambda_t} = \frac{2\nu\exp(\lambda_t)\epsilon_t^3}{(\nu + \epsilon_t^2)^2} \tag{21}$$

which is a continuously differentiable function of $\epsilon_t = (y_t - \mu_t)\exp(-\lambda_t)$. Hence,

$$\left|\frac{\partial u_{\mu,t}}{\partial \lambda_t}\right| = \left|\frac{2\nu\exp(\lambda_t)\epsilon_t^3}{(\nu + \epsilon_t^2)^2}\right| < \left|\frac{2\nu\exp(\lambda_{\max})\epsilon_t^3}{(\nu + \epsilon_t^2)^2}\right| \tag{22}$$

The right side of Eq. (22) includes a continuous function of ϵ_t. Moreover, if $|\epsilon_t| \to \infty$, hence, the right side of Eq. (22) goes to zero. Hence, the right side of Eq. (22) is finite for all ϵ_t. Therefore, $\partial u_{\mu,t}/\partial\lambda_t$ is a bounded function of ϵ_t. The derivative of $u_{\lambda,t}$ with respect to μ_t is:

$$\frac{\partial u_{\lambda,t}}{\partial \mu_t} = -\frac{2\nu(\nu+1)\exp(-\lambda_t)\epsilon_t}{(\nu + \epsilon_t^2)^2} \tag{23}$$

which is a continuously differentiable function of $\epsilon_t = (y_t - \mu_t)\exp(-\lambda_t)$. Hence,

$$\left|\frac{\partial u_{\lambda,t}}{\partial \mu_t}\right| = \left|\frac{2\nu(\nu+1)\exp(-\lambda_t)\epsilon_t}{(\nu + \epsilon_t^2)^2}\right| < \left|\frac{2\nu(\nu+1)\exp(\lambda_{\max})\epsilon_t}{(\nu + \epsilon_t^2)^2}\right| \tag{24}$$

The right side of Eq. (24) includes a continuous function of ϵ_t. Moreover, if $|\epsilon_t| \to \infty$, then the right side of Eq. (24) goes to zero. Hence, the right side of Eq. (24) is finite for all ϵ_t. Therefore, $\partial u_{\lambda,t}/\partial\mu_t$ is a bounded function of ϵ_t. The derivative of $u_{\lambda,t}$ with respect to λ_t is:

$$\frac{\partial u_{\lambda,t}}{\partial \lambda_t} = -\frac{2\nu(\nu+1)\epsilon_t^2}{(\nu + \epsilon_t^2)^2} \tag{25}$$

which is a continuously differentiable function of $\epsilon_t = (y_t - \mu_t)\exp(-\lambda_t)$. Hence,

$$\left|\frac{\partial u_{\lambda,t}}{\partial \lambda_t}\right| = \left|\frac{2\nu(\nu+1)\epsilon_t^2}{(\nu + \epsilon_t^2)^2}\right| < \left|\frac{2\nu(\nu+1)\epsilon_t^2}{\nu + \epsilon_t^2}\right| < \infty \tag{26}$$

where the first inequality is due to $\nu > 1$, and the second inequality is true because $[\epsilon_t^2/(\nu+\epsilon_t^2)] \in (0,1)$. As a consequence, $\partial u_{\lambda,t}/\lambda_t$ is a bounded function of ϵ_t.

Due to (A1), (A4), $\mathrm{Var}(\epsilon_t) < \infty$ as $\nu > 2$, and the boundedness of score functions, their derivatives, and their products (products of bounded functions are also bounded), we have the following result:

$$E\left[(u_{j,t})^{2-i}\left(\frac{\partial u_{k,t}}{\partial l_t}\right)^i\right] < \infty \tag{27}$$

$$E\left[\left(\frac{\partial u_{j,t}}{\partial m_t}\right)^{2-i}\left(\frac{\partial u_{k,t}}{\partial l_t}\right)^i\right] < \infty \tag{28}$$

for $i = 0, 1, 2$, and $j, k, l, m = \mu, \lambda$.

In the following, we study the consequences of (A3) on the score functions:

(i) Score function $u_{\mu,t}$ is a MDS due to the following arguments. Due to (A3), $G_t(\Theta_0)'$ is a MDS:

$$E_{t-1}\left[\frac{\partial \ln f(y_t|\mathcal{F}_{t-1},\Theta)}{\partial \Theta'}\right] = E_{t-1}\left[\frac{\partial \ln f(y_t|\mathcal{F}_{t-1},\Theta)}{\partial \mu_t}\right] \times \frac{\partial \mu_t}{\partial \Theta'} = 0 \tag{29}$$

where index $t-1$ indicates expectations that are conditional on $\mathcal{F}_{t-1}$. Since $(\partial \mu_t/\partial \Theta') \neq 0$,

$$E_{t-1}\left[\frac{\partial \ln f(y_t|\mathcal{F}_{t-1},\Theta)}{\partial \mu_t}\right] = E_{t-1}\left[\frac{\nu+1}{\nu \exp(2\lambda_t)} u_{\mu,t}\right] = E_{t-1}(u_{\mu,t})\frac{\nu+1}{\nu \exp(2\lambda_t)} = 0 \tag{30}$$

Thus, $E_{t-1}(u_{\mu,t}) = 0$.

(ii) Score function $u_{\lambda,t}$ is a MDS due to the following arguments. Due to (A3), $G_t(\Theta_0)'$ is a MDS:

$$E_{t-1}\left[\frac{\partial \ln f(y_t|\mathcal{F}_{t-1},\Theta)}{\partial \Theta'}\right] = E_{t-1}\left[\frac{\partial \ln f(y_t|\mathcal{F}_{t-1},\Theta)}{\partial \lambda_t}\right] \times \frac{\partial \lambda_t}{\partial \Theta'} = 0 \tag{31}$$

Since $(\partial \lambda_t/\partial \Theta') \neq 0$,

$$E_{t-1}\left[\frac{\partial \ln f(y_t|\mathcal{F}_{t-1},\Theta)}{\partial \lambda_t}\right] = E_{t-1}(u_{\lambda,t}) = 0 \tag{32}$$

(iii) $E(u_{\mu,t}) = 0$ and $E(u_{\lambda,t}) = 0$, due to the law of iterated expectations.

(iv) $u_{\mu,t}$ is a MDS and $\text{Var}(u_{\mu,t}) < \infty$, hence $u_{\mu,t}$ is white noise.

(v) $u_{\lambda,t}$ is a MDS and $\text{Var}(u_{\lambda,t}) < \infty$, hence $u_{\lambda,t}$ is white noise.

(iv) Score function $u_{\mu,t}$ is not i.i.d., because λ_t depends on λ_{t-1}.

(v) Score function $u_{\lambda,t}$ is i.i.d., because $u_{\lambda,t}$ is a continuous function of ϵ_t and ϵ_t is i.i.d. (White 2001).

(vi) Score function $u_{\lambda,t}$ is an $\mathcal{F}$-measurable function of ϵ_t (White 2001), because $u_{\lambda,t}$ is a continuous function of ϵ_t (Harvey 2013).

(vii) Score function $u_{\lambda,t}$ is strictly stationary and ergodic, because $u_{\lambda,t}$ is an $\mathcal{F}$-measurable function of ϵ_t, and because ϵ_t is strictly stationary and ergodic (White 2001, Theorem 3.35).

(viii) If the maximum modulus of eigenvalues of Φ is less than one, then, due to the finite variances of score functions: λ_t is a sum of nonlinear functions of lags of ϵ_t. As a consequence, the continuous $u_{\mu,t}$ function, which includes λ_t, is an $\mathcal{F}$-measurable function of ϵ_t (White 2001).

(ix) Score function $u_{\mu,t}$ is strictly stationary and ergodic, because $u_{\mu,t}$ is an $\mathcal{F}$-measurable function of $(\epsilon_1, \ldots, \epsilon_t)$, and because ϵ_t is strictly stationary and ergodic (White 2001, Theorem 3.35).

In Fig. 2, the in-sample estimates, for the period of January 1985 to February 2020, of $u_{\mu,t}$ and $u_{\lambda,t}$ as a function of ϵ_t are presented for Beta-t-QVAR(1) by using DJIA data. The figure indicates the continuity and boundedness of the score functions for the ML estimates.

Proposition 1(a): For Beta-t-QVAR(1) and Beta-t-QVAR(1)-M, if the maximum modulus of eigenvalues of Φ is less than one, and Ψ is non-zero, then θ_t is covariance stationary.

Proof: For filter $\theta_t = \Phi\theta_{t-1} + \Psi u_{t-1}$, u_t is white noise, with zero mean and a well-defined covariance matrix for $\nu > 2$. If the maximum modulus of eigenvalues of Φ is < 1 and Ψ is non-zero, then θ_t is covariance stationary. *QED*

Proposition 1(b): For Beta-t-QVAR(1)-lev and Beta-t-QVAR(1)-M-lev, if the maximum modulus of eigenvalues of Φ is less than one, and Ψ or ψ^* is non-zero, then θ_t is covariance stationary.

Proof: For filter $\theta_t = \Phi\theta_{t-1}+g(u_{t-1})$, $g(u_t)$ is white noise, with zero mean and a well-defined covariance matrix for $\nu > 2$. If the maximum modulus of eigenvalues of Φ is < 1 and Ψ or ψ^* is non-zero, then θ_t is covariance stationary. *QED*

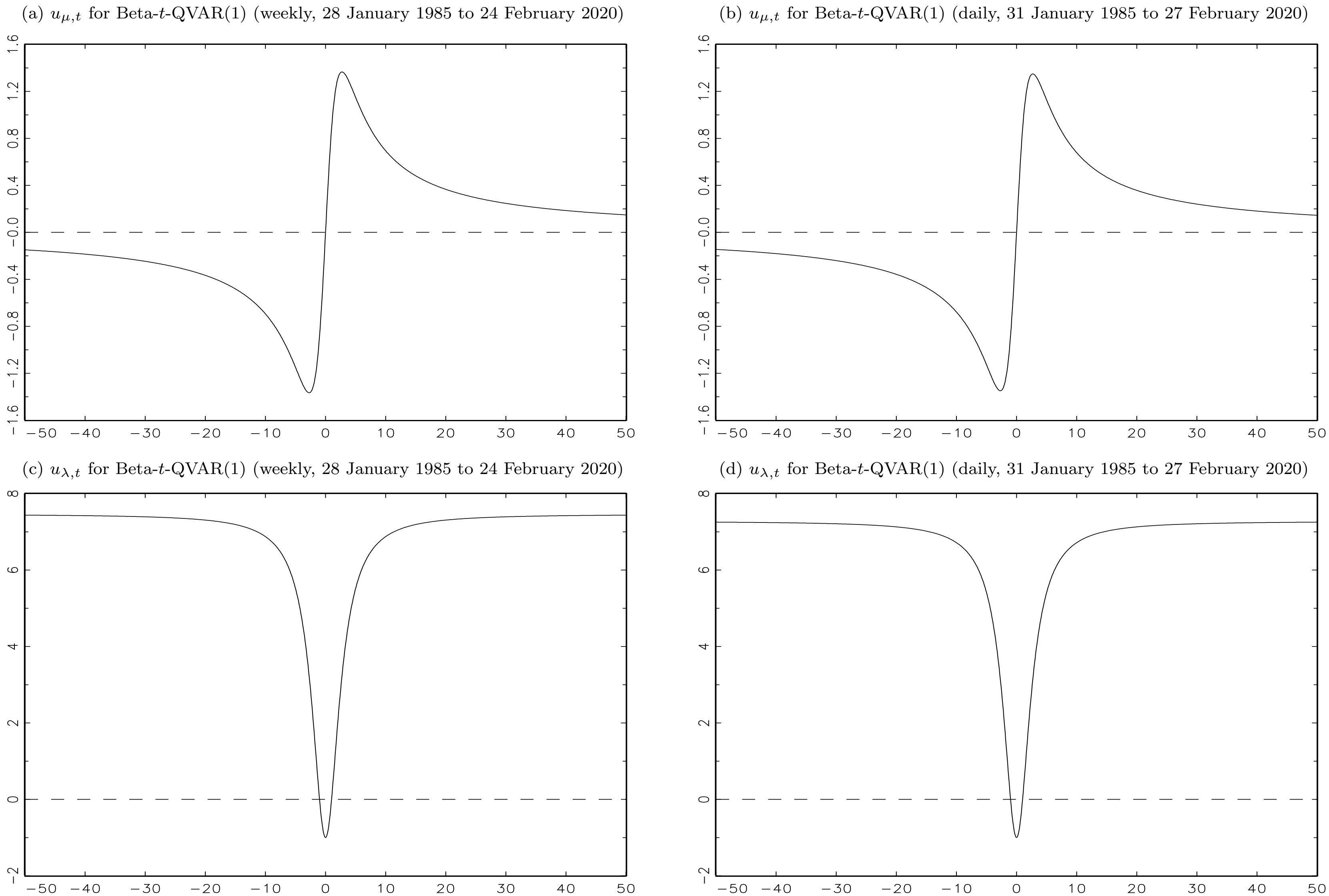


Fig. 2. Score functions $u_{\mu,t}$ ("asymptotic trimming") and $u_{\lambda,t}$ ("asymptotic Winsorizing"), as functions of ϵ_t for the estimation window.
Notes: $\lambda_t = 0$, $\hat{\nu} = 7.4613$ (weekly data), and $\hat{\nu} = 7.2765$ (daily data). Estimates for Beta-t-QVAR(1) are presented because the Akaike information criterion (AIC), Bayesian information criterion (BIC), and Hannan–Quinn criterion (HQC) of this model are superior to the statistical performance metrics of other Beta-t-QVAR(1) specifications.

5.3. Asymptotic theory of the ML estimator

We use the following assumptions for the asymptotic theory of ML (Wooldridge 1994):

(A5) y_t is strictly stationary and ergodic on $\mathbb{R}$ (Wooldridge 1994, Theorem 4.1).

(A6) $\ln f(\cdot|\mathcal{F}_{t-1};\Theta) : \mathbb{R} \times \tilde{\Theta} \rightarrow \mathbb{R}$ is a real-valued function, where $\tilde{\Theta} \subset \mathbb{R}^S$ is the parameter set.

(A7) For each $\Theta \in \tilde{\Theta}$, $\ln f(\cdot|\mathcal{F}_{t-1};\Theta)$ is a Borel measurable function on $\mathbb{R}$ (Wooldridge 1994, Theorem 4.1 and Definition A2).

(A8) For each $y_t \in \mathbb{R}$, $\ln f(y_t|\mathcal{F}_{t-1};\cdot)$ is a continuous function on $\tilde{\Theta}$ (Wooldridge 1994, Theorem 4.1 and Definition A2).

(A9) $\exists$ function $b(\cdot)$ such that $|\ln f(y_t|\mathcal{F}_{t-1};\Theta)| \leq b(y_t)$ for all Θ, and $E[b(y_t)] < \infty$ (Wooldridge 1994, Theorem 4.1).

(A10) $\int_{\mathbb{R}} f(y_t|\mathcal{F}_{t-1};\Theta)dy_t = 1$ for all Θ (Wooldridge 1994, p. 2673).

(A11) Θ_0 is a unique solution to (Wooldridge 1994, Theorem 4.3):

$$\max_{\Theta \in \tilde{\Theta}} \lim_{T \rightarrow \infty} \sum_{t=1}^{T} \ln f(y_t|\mathcal{F}_{t-1}, \Theta) \tag{33}$$

(A12) Each element of $H_t(\Theta)$ is strictly stationary and ergodic (Wooldridge 1994, Theorem 4.1).

(A13) For each element of $H_t(\Theta)$, $H_{i,j,t}(\Theta) : \mathbb{R} \times \tilde{\Theta} \rightarrow \mathbb{R}$ is a real-valued function (Wooldridge 1994, Theorem 4.1).

(A14) For each $\Theta \in \tilde{\Theta}$, each element of $H_t(\Theta)$ is a Borel measurable function on $\mathbb{R}$ (Wooldridge 1994, Theorems 4.1 and 4.4, and Definition A2).

(A15) For each $y_t \in \mathbb{R}$, each element of $H_t(\Theta)$ is a continuous function on $\tilde{\Theta}$ (Wooldridge 1994, Theorems 4.1 and 4.4, and Definition A2).

(A16) $\exists$ function $b(\cdot)$ such that, for all elements of $H_t(\Theta)$, $|H_{i,j,t}(\Theta)| \leq b[H_{i,j,t}(\Theta)]$ for all Θ, and $E\{b[H_{i,j,t}(\Theta)]\} < \infty$ (Wooldridge 1994, Theorems 4.1 and 4.4).

(A17) $E[G_t(\Theta_0)G_t(\Theta_0)'] < \infty$ (Wooldridge 1994, Definition 4.3).

(A18) $(1/\sqrt{T})\sum_{t=1}^{T} E[G_t(\Theta_0)'] \to 0_{S\times 1}$ for $T \to \infty$ (Wooldridge 1994, Definition 4.3).

(A19) $(1/\sqrt{T})\sum_{t=1}^{T} G_t(\Theta_0)' \to_d N(0, B_0)$ for $T \to \infty$, where (Wooldridge 1994, Definition 4.3):

$$B_0 = \lim_{T\to\infty} \mathrm{Var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} G_t(\Theta_0)'\right] \tag{34}$$

(A20) Θ_0 is an interior point within $\tilde{\Theta} \subset \mathbb{R}^S$ (Wooldridge 1994, Theorem 4.4).

(A21) For each $y_t \in \mathbb{R}$, $\ln f(y_t|\mathcal{F}_{t-1};\cdot)$ is twice continuously differentiable on all of the interior points of $\tilde{\Theta}$ (Wooldridge 1994, Theorem 4.4).

(A22) $\partial[\int_{\mathbb{R}} f(y_t|\mathcal{F}_{t-1};\Theta)dy_t]/\partial\Theta = \int_{\mathbb{R}}[\partial f(y_t|\mathcal{F}_{t-1};\Theta)/\partial\Theta]dy_t$ (Wooldridge 1994, p. 2674).

(A23) $\partial[\int_{\mathbb{R}} G_t(\Theta)' f(y_t|\mathcal{F}_{t-1};\Theta)dy_t]/\partial\Theta = \int_{\mathbb{R}}[\partial G_t(\Theta)' f(y_t|\mathcal{F}_{t-1};\Theta)/\partial\Theta]dy_t$ (Wooldridge 1994, p. 2675).

(A24) Matrix

$$A_0 = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[H_t(\Theta_0)] = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathrm{Var}[G_t(\Theta_0)'] \tag{35}$$

is positive definite (Wooldridge 1994, Theorem 4.4).

Conditions (A1), (A6), (A7), (A8), (A10), (A13), (A14), (A15), (A20), (A21), (A22), and (A23) are standard regularity conditions in the literature, which hold for the Beta-t-QVAR model. Conditions (A2), (A3), (A11), and (A24) are standard maintained assumptions in the literature, which are assumed to hold in this paper. Condition (A4) can be interpreted as a regularity condition of non-infinite volatility. Condition (A9) holds for the parameter set of the Beta-t-QVAR model, because $|\ln f(y_t|\mathcal{F}_{t-1};\Theta)|$ is bounded for $2 < \nu < \infty$, $-\infty < \lambda_t < \lambda_{\max}$, and $-\infty < \mu_t < \infty$, by using the triangle inequality for Eq. (9). Condition (A16) holds due to the bounded derivatives of the score functions for $\tilde{\Theta}$ under (A4). Conditions (A5), (A12), (A17), (A18), and (A19) for the Beta-t-QVAR model are proven in the remainder of this section. Consistency of the ML estimator of Beta-t-QVAR is studied by using Monte Carlo simulation experiments of Section 5.4.

In the following, several propositions about the asymptotic properties of the ML estimates (i.e. consistency and asymptotic normality of $\hat{\theta}$) are proven. We refer to the work of Wooldridge (1994).

Proposition 2: If assumptions (A1), (A5), (A6), (A7), (A8), and (A9) hold. Hence, $\ln f(y_t|\mathcal{F}_{t-1};\Theta)$ satisfies the uniform weak law of large numbers (UWLLN) (Wooldridge 1994, Definition 4.2).

Proof: It follows from Wooldridge (1994, Theorem 4.1). For (A5), we use the following result for Beta-t-QVAR: $\hat{\theta}_t$ converges exponentially almost surely (e.a.s.) to the unique strictly stationary and ergodic sequence $\theta_t(\Theta_0)$ for $T \to \infty$, which is proven in Propositions 7(a-b) (Supplementary Material). Moreover, for (A5), we also use the work of White (2001, Theorem 3.35): In Eqs. (1)-(3), θ_t is transformed to y_t, by using an $\mathcal{F}$-measurable function. Therefore, y_t is strictly stationary and ergodic for $T \to \infty$. (A1), (A6), (A7), (A8), and (A9) hold for the Beta-t-QVAR models of this paper. *QED*

Proposition 3: If the following assumptions hold: (A1), (A2), (A7), (A8), (A10), (A11), and log-density $\ln f(y_t|\mathcal{F}_{t-1};\Theta)$ satisfies the UWLLN. Hence, $\hat{\Theta}$ is weakly consistent, i.e. $\hat{\Theta} \to_p \Theta_0$.

Proof: It follows from Wooldridge (1994, Theorem 5.1). (A2) and (A11) are maintained assumptions. (A1), (A7), (A8), and (A10) hold for the Beta-t-QVAR models of this paper. The assumption '$\ln f(y_t|\mathcal{F}_{t-1};\Theta)$ satisfies the UWLLN' holds due to Proposition 2. *QED*

Proposition 4: If the following assumptions hold: (A1), (A12), (A13), (A14), (A15), and (A16). Hence, $H_t(\Theta)$ satisfies the UWLLN.

Proof: It follows from Wooldridge (1994, Theorem 4.1). For (A12), we use the following result for Beta-t-QVAR: $H_t(\hat{\Theta})$ converges e.a.s. to the unique strictly stationary and ergodic sequence $H_t(\Theta_0)$ for $T \to \infty$ (conditions are in the Supplementary Material). (A1), (A13), (A14), and (A15) hold for the Beta-t-QVAR models of this paper. (A16) is a maintained assumption. *QED*

Proposition 5: If the following assumptions hold: (A17), (A18), and (A19). Hence, $G_t(\Theta_0)$ satisfies the central limit theorem (CLT) with asymptotic variance B_0.

Proof: It follows from Wooldridge (1994, Definition 4.3). For (A17), we use the following result: $E[H_t(\Theta_0)] = \text{Var}[G_t(\Theta_0)')] = E[G_t(\Theta_0)'G_t(\Theta_0)] < \infty$, where the equalities hold due to (A22) (Wooldridge 1994, p. 2674) and (A23) (Wooldridge 1994, p. 2675), respectively. Inequality $E[G_t(\Theta_0)'G_t(\Theta_0)] < \infty$ is shown in Propositions 9(a-b) (Supplementary Material), which implies $E[G_t(\Theta_0)G_t(\Theta_0)'] < \infty$, because the terms of the sum defined by $G_t(\Theta_0)G_t(\Theta_0)'$ are in the diagonal of $G_t(\Theta_0)'G_t(\Theta_0)$. For (A18), we use the following result: $E[G_t(\Theta_0)'] = 0_{S\times 1}$, which holds under (A22) (Wooldridge 1994, p. 2674). For (A19) we, use White (2001, Theorem 5.16): (i) $G_t(\Theta_0)'$ is a MDS, which holds under (A3) (Wooldridge 1994, p. 2677). Therefore, $G_t(\Theta_0)$ is an adapted mixingale (White 2001, Definition 5.15, p. 125). (ii) $G_t(\hat{\Theta})'$ converges e.a.s. to the unique strictly stationary and ergodic sequence $G_t(\Theta_0)'$ for $T \to \infty$, which is proven in Propositions 10(a-b) (Supplementary Material). (i) and (ii) provide (A18). *QED*

Proposition 6: If the following assumptions hold: (A1), (A2), (A3), (A7), (A8), (A10), (A11), (A20), (A21), (A22), (A23), (A24), $\ln f(y_t|\mathcal{F}_{t-1};\Theta)$ satisfies the UWLLN, $H_t(\Theta)$ satisfies the UWLLN, and $G_t(\Theta_0)$ satisfies the CLT with asymptotic variance:

$$B_0 = \lim_{T\to\infty} \mathrm{Var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} G_t(\Theta_0)'\right] \tag{36}$$

then,

$$\sqrt{T}(\hat{\Theta} - \Theta_0) \to_d N_S\left(0_{S\times 1}, A_0^{-1}B_0A_0^{-1}\right) = N_S\left(0_{S\times 1}, A_0^{-1}\right) \quad \text{as} \quad T \to \infty \tag{37}$$

The equality in Eq. (37) is due to (A3), which provides: (i) $G_t(\Theta_0)'$ is a MDS (Wooldridge 1994, p. 2677), (ii) $G_t(\Theta_0)'$ is serially uncorrelated (Wooldridge 1994, pp. 2676-2677), and (iii) $A_0 = B_0$ (Wooldridge 1994, p. 2676). The equality in Eq. (37) is due to (iii).

Proof: It follows from Wooldridge (1994, Theorem 5.2). (A2), (A3), (A11), and (A24) are maintained assumptions. (A1), (A7), (A8), (A10), (A20), (A21), (A22), and (A23) hold for the Beta-t-QVAR models. The assumption '$\ln f(y_t|\mathcal{F}_{t-1};\Theta)$ satisfies the UWLLN' holds due to Proposition 2. The assumption '$H_t(\Theta)$ satisfies the UWLLN' holds due to Proposition 4. The assumption '$G_t(\Theta_0)$ satisfies the CLT asymptotic variance B_0' holds due to Proposition 5. *QED*

5.4. Monte Carlo results

In Table 2, we present the results of a Monte Carlo study, which is based on the simulations of 200 trajectories of y_t with $T_{\mathrm{MC}} = 2,500$. The sample size T_{MC} is larger than the sample size of the weekly DJIA data, and it is smaller than the sample size of the daily DJIA data. Each trajectory is drawn independently of each other. In the Monte Carlo study, we use one set of true parameter values that is similar to the parameter estimates for Beta-t-QVAR(1)-M-lev of weekly DJIA for the period of January 1985 to February 2020. Beta-QVAR(1)-M-lev is used in the Monte Carlo study, because that model has the most accurate prediction accuracy for volatility from the Beta-t-QVAR specifications (Section 7). Due to the specific set of true parameter values, the specific econometric model, and the relatively limited number of trajectories, it is important to note that the Monte Carlo results provide only an illustration of the consistency of ML estimation of score-driven location plus scale models.

To compare true and estimated parameter values, the Sign test (i.e. distribution-free or non-parametric test) is used, which compares the median of the parameter estimates with the true value for each parameter. An advantage in using the Sign test is that it is based on few assumptions about the true data generating process. The test results indicate that the Sign test statistic is not significantly different from zero for any of the parameters (Table 2).

Table 2. Sign test results for simulated data

	True value	Median	Sign statistic	p-value
c	0.3000	0.2865	0.5000	0.4795
β_1	0.1000	0.1019	0.7200	0.3961
β_2	-0.0400	-0.0353	0.5000	0.4795
$\phi_{1,1}$	0.8024	0.8029	0.0200	0.8875
$\phi_{1,2}$	-0.0191	-0.0177	0.3200	0.5716
$\phi_{2,1}$	-0.0226	-0.0219	0.7200	0.3961
$\phi_{2,2}$	0.9776	0.9726	1.6200	0.2031
$\psi_{1,1}$	-0.6100	-0.6090	0.3200	0.5716
$\psi_{1,2}$	0.2200	0.2198	0.0000	1.0000
$\psi_{2,1}$	-0.0800	-0.0798	0.0800	0.7773
$\psi_{2,2}$	0.0600	0.0590	0.9800	0.3222
ψ^*	-0.0200	-0.0203	0.9800	0.3222
ω	0.5600	0.5612	0.1800	0.6714
ν	7.5000	7.5885	0.9800	0.3222

Notes: The Monte Carlo study is based on the simulations of 200 trajectories of y_t with $T_{\text{MC}} = 2,500$ observations for each trajectory.

6. Impulse response functions (IRFs)

For Beta-t-QVAR(1) and Beta-t-QVAR(1)-M, the IRFs of filter θ_t are:

$$\text{IRF}_{j+1,t} = \frac{\partial \theta_{t+j+1}}{\partial (v_t, v_t)} = \frac{\partial \theta_{t+j+1}}{\partial u_t'} \begin{bmatrix} \partial u_{\mu,t}/\partial v_t & \partial u_{\mu,t}/\partial v_t \\ \partial u_{\lambda,t}/\partial v_t & \partial u_{\lambda,t}/\partial v_t \end{bmatrix} = \Phi^j \Psi \begin{bmatrix} \partial u_{\mu,t}/\partial v_t & \partial u_{\mu,t}/\partial v_t \\ \partial u_{\lambda,t}/\partial v_t & \partial u_{\lambda,t}/\partial v_t \end{bmatrix} \quad (38)$$

for $j = 0, 1, \ldots, \infty$. For Beta-t-QVAR(1)-lev and Beta-t-QVAR(1)-M-lev, the IRFs of filter θ_t are:

$$\text{IRF}_{j+1,t} = \frac{\partial \theta_{t+j+1}}{\partial(v_t, v_t)} = \frac{\partial \theta_{t+j+1}}{\partial u_t'} \begin{bmatrix} \partial u_{\mu,t}/\partial v_t & \partial u_{\mu,t}/\partial v_t \\ \partial u_{\lambda,t}/\partial v_t & \partial u_{\lambda,t}/\partial v_t \end{bmatrix} = \Phi^j \Psi_k \begin{bmatrix} \partial u_{\mu,t}/\partial v_t & \partial u_{\mu,t}/\partial v_t \\ \partial u_{\lambda,t}/\partial v_t & \partial u_{\lambda,t}/\partial v_t \end{bmatrix} \quad (39)$$

for $j = 0, 1, \ldots, \infty$. For the latter IRF, two versions are defined, $k = P, N$, where $k = P$ assumes positive unexpected return $v_{t-1} > 0$, and $k = N$ assumes negative unexpected return $v_{t-1} < 0$, and

$$\Psi_P = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} - \psi^* \end{bmatrix}, \quad \Psi_N = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} + \psi^* \end{bmatrix} \quad (40)$$

respectively. In Eqs. (38) and (39), the derivatives are given by:

$$\frac{\partial u_{\mu,t}}{\partial v_t} = \frac{\nu \exp(2\lambda_t)[\nu \exp(2\lambda_t) - \epsilon_t^2]}{[\nu \exp(2\lambda_t) + v_t^2]^2} \quad (41)$$

$$\frac{\partial u_{\lambda,t}}{\partial v_t} = \frac{2\nu(\nu+1)\exp(2\lambda_t) v_t}{[\nu \exp(2\lambda_t) + v_t^2]^2} \quad (42)$$

respectively. The IRFs of Eqs. (38) and (39) are time-dependent, since they correspond to nonlinear models. For the definitions of IRFs of nonlinear models there are several alternatives in the literature (e.g. Lütkepohl 2005; Herwartz and Lütkepohl 2000).

In Figs. 3 and 4, for weekly and daily DJIA data, respectively, the estimates of $\partial \theta_{t+j+1}/\partial u_t'$ are presented. In the panels of Figs. 3 and 4, the following significant dynamic interaction effects are shown: dynamic overreaction effects, dynamic risk premium effects, dynamic leverage effects, and dynamic volatility effects. The measurement of the dynamic effects of the score functions is motivated by Fig. 2, in which it is presented that the score functions are robust to extreme observations. Therefore, the dynamic interaction effects $\partial \theta_{t+j+1}/\partial u_t'$ can provide robust information about the significance and the signs of the dynamic interaction effects between risk premium and volatility.

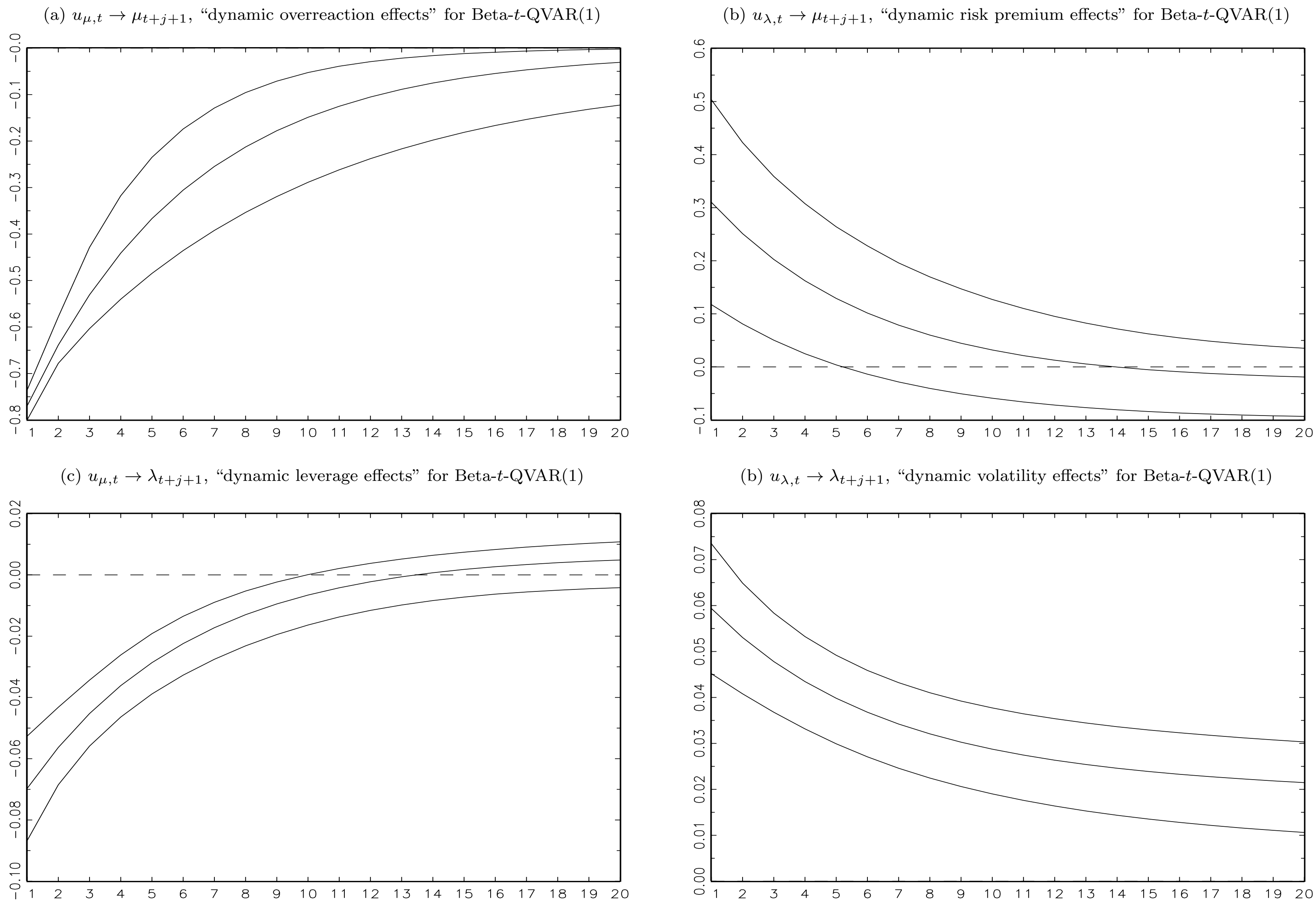


Fig. 3. IRFs $\partial\theta_{t+j+1}/\partial u_t' = \Phi^j \Psi$ for $j+1 = 1, \ldots, 20$ for weekly DJIA for the estimation window (28 January 1985 to 24 February 2020). *Notes:* 90% Monte Carlo confidence interval for $\hat{\Phi}^j \hat{\Psi}$ is presented. Beta-t-QVAR(1) is presented since the AIC, BIC, and HQC of this model are superior to the same metrics of other Beta-t-QVAR(1) specifications.

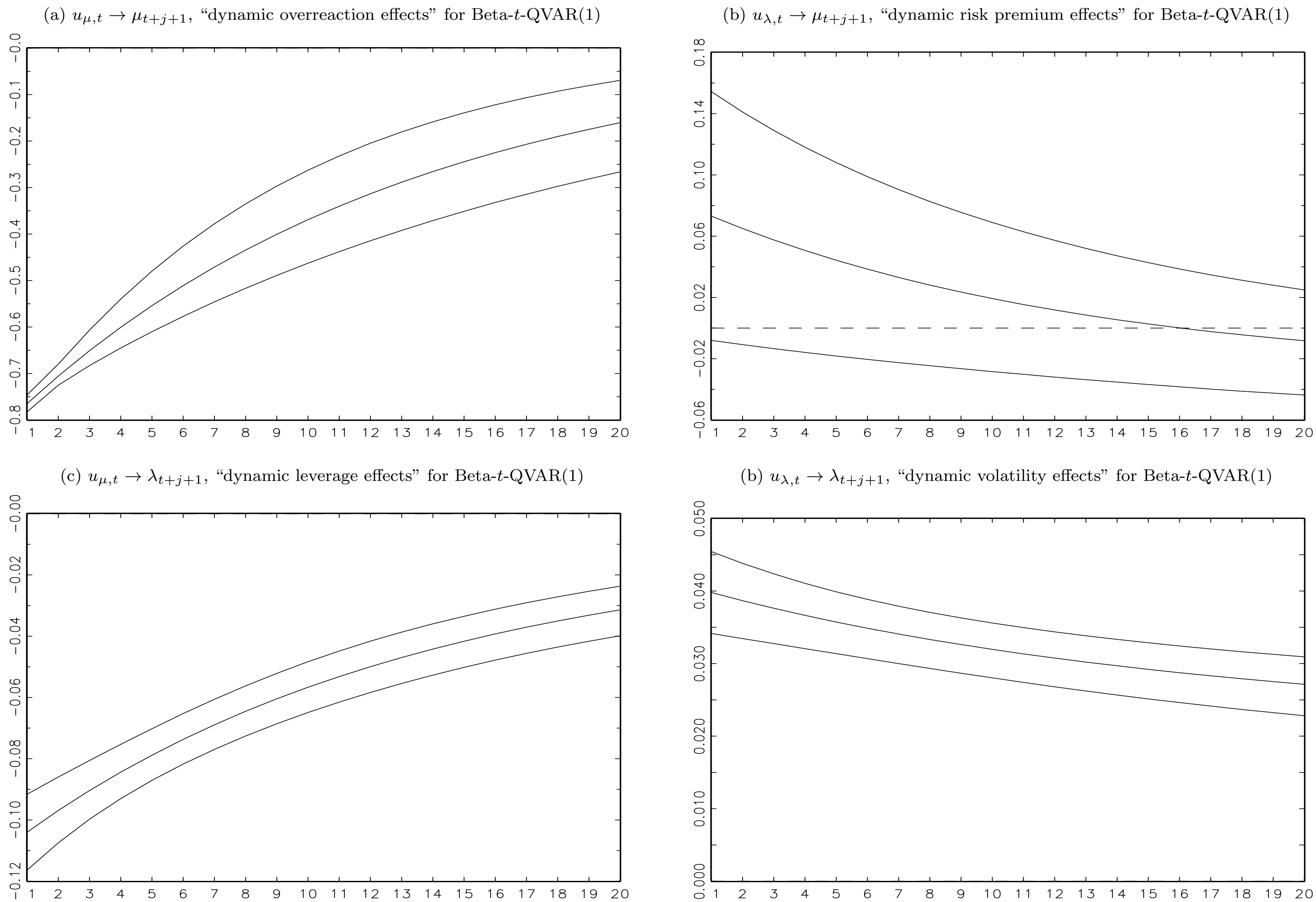


Fig. 4. IRFs $\partial\theta_{t+j+1}/\partial u_t' = \Phi^j\Psi$ for $j+1 = 1, \ldots, 20$ for daily DJIA for the estimation window (31 January 1985 to 27 February 2020). *Notes:* 90% Monte Carlo confidence interval for $\hat{\Phi}^j\hat{\Psi}$ is presented. Beta-t-QVAR(1) is presented since the AIC, BIC, and HQC of this model are superior to the same metrics of other Beta-t-QVAR(1) specifications.

7. Prediction accuracy

One-step ahead forecasts of volatility, $\sigma_t = \exp(\lambda_t)[\nu/(\nu-2)]^{1/2}$, are computed for the forecasting window. For each Beta-t-QVAR(1) specification, λ_t is given by Eqs. (1)-(7). As a proxy of true volatility σ_t^*, the square root of realized variance is used (source: Oxford-Man Institute of Quantitative Finance (OMI), https://realized.oxford-man.ox.ac.uk/data/download). From the realized variance data file of OMI, variable 'rv5' is used (Harvey and Lange 2018). Prediction accuracies are compared by using the Diebold–Mariano test for the following loss functions (Hansen and Lunde 2005; Patton 2011):

$$\begin{array}{ll} \mathrm{MSE}_{1,i,t} = (\sigma_t^* - \sigma_{i,t})^2 & \mathrm{MSE}_{2,i,t} = [(\sigma_t^*)^2 - \sigma_{i,t}^2]^2 \\ \mathrm{QLIKE}_{i,t} = \left\{ \frac{(\sigma_t^*)^2}{\sigma_{i,t}^2} - \ln\left[\frac{(\sigma_t^*)^2}{\sigma_{i,t}^2}\right] - 1 \right\} & \mathrm{R^2LOG}_{i,t} = \left\{ \ln\left[\frac{(\sigma_t^*)^2}{\sigma_{i,t}^2}\right] \right\}^2 \\ \mathrm{MAE}_{1,i,t} = |\sigma_t^* - \sigma_{i,t}| & \mathrm{MAE}_{2,i,t} = |(\sigma_t^*)^2 - \sigma_{i,t}^2| \end{array} \tag{43}$$

for model i and for each period of the forecasting window $t = 1, \ldots, T_f$. We use the general $\mathrm{LOSS}_{i,t}$ notation for all loss functions. The Diebold–Mariano test studies the significance of the mean difference between the loss functions of two forecasting models:

$$\overline{d}_{i,j} = \frac{1}{T_f}\sum_{t=1}^{T_f} d_{i,j,t} = \frac{1}{T_f}\sum_{t=1}^{T_f}(\mathrm{LOSS}_{i,t} - \mathrm{LOSS}_{j,t}) \tag{44}$$

If $\overline{d}_{i,j}$ is significantly positive, then the forecast accuracy of model j is superior to that of model i. If $\overline{d}_{i,j}$ is significantly negative, then the forecast accuracy of model i is superior to that of model j. Otherwise, the forecast accuracy of the two models does not differ significantly.

With respect to the statistical properties of Beta-t-QVAR for the rolling windows, the empirical estimate of the maximum modulus of eigenvalues of Φ is denoted C_{Stat}. In Fig. 5, rolling-window estimates, for the period of January 2010 to February 2020, of C_{Stat} for DJIA are presented for Beta-t-QVAR(1). Furthermore, the empirical estimate of the Lyapunov exponent is C_{Inv}. In Fig. 5, rolling-window estimates, for the period of January 2010 to February 2020, of C_{Inv} for DJIA are presented for Beta-t-QVAR(1). The condition of the negative Lyapunov exponent for Beta-t-QVAR(1) (Brandt 1986; Elton 1990; Alsmeyer 2003; Straumann and Mikosch 2006):

$$\inf_{n\geq 1}\left\{ n^{-1}E\left[\ln\left|\left|\prod_{t=1}^{n}\frac{\partial\theta_t}{\partial(\theta_{t-1})'}\right|\right|_1\right]\right\} = \inf_{n\geq 1}\left\{ n^{-1}E\left(\ln\left|\left|\prod_{t=1}^{n} X_{t-1}\right|\right|_1\right)\right\} < 0 \tag{45}$$

where

$$X_{t-1} \equiv \frac{\partial \theta_t}{\partial(\theta_{t-1})'} = \Phi + \Psi \frac{\partial u_{t-1}}{\partial(\theta_{t-1})'} \tag{46}$$

which is one of the conditions of the e.a.s. convergence of $\hat{\theta}_t$ to $\theta_t(\Theta_0)$ for $T \to \infty$. We present further technical details in the Supplementary Material. In Fig. 5, $C_{\text{Stat}} < 1$ and $C_{\text{Inv}} < 0$ are supported.

The prediction accuracy of Beta-t-QVAR is compared with the prediction accuracies of the following models: (i) Motivated by the work of Hansen and Lunde (2005), A-PARCH, normal-GARCH, and t-GARCH are used. For normal-GARCH and t-GARCH, we study prediction accuracies with and without leverage effects in the model specifications. Similarly to the work of Hansen and Lunde (2005), the zero mean, constant mean, and GARCH-in-mean (GARCH-M) alternatives are used for expected return in the benchmark models. With respect to the lag-orders, Beta-t-EGARCH(1,1), A-PARCH(1,1), normal-GARCH(1,1), and t-GARCH(1,1) specifications are considered. (ii) Beta-t-EGARCH, for which the zero mean, constant mean, and EGARCH-in-mean (EGARCH-M) cases are considered, and we study prediction accuracies with and without leverage effects in each specification.

In Tables 3 and 4, for weekly and daily data, respectively, mean loss functions for the forecasting window are presented. For each mean loss function, the significance of the Diebold–Mariano test statistic is presented (without presenting the estimate of the Diebold–Mariano test statistic). We use Beta-t-QVAR(1)-M-lev as the benchmark volatility forecasting model, for which the mean loss function estimates are highlighted by bold numbers in Tables 3 and 4. We find that Beta-t-EGARCH, A-PARCH, and GARCH are not superior to Beta-t-QVAR for any of the loss functions, and, for several loss functions, Beta-t-QVAR provides significantly more accurate forecasts than the competing models.

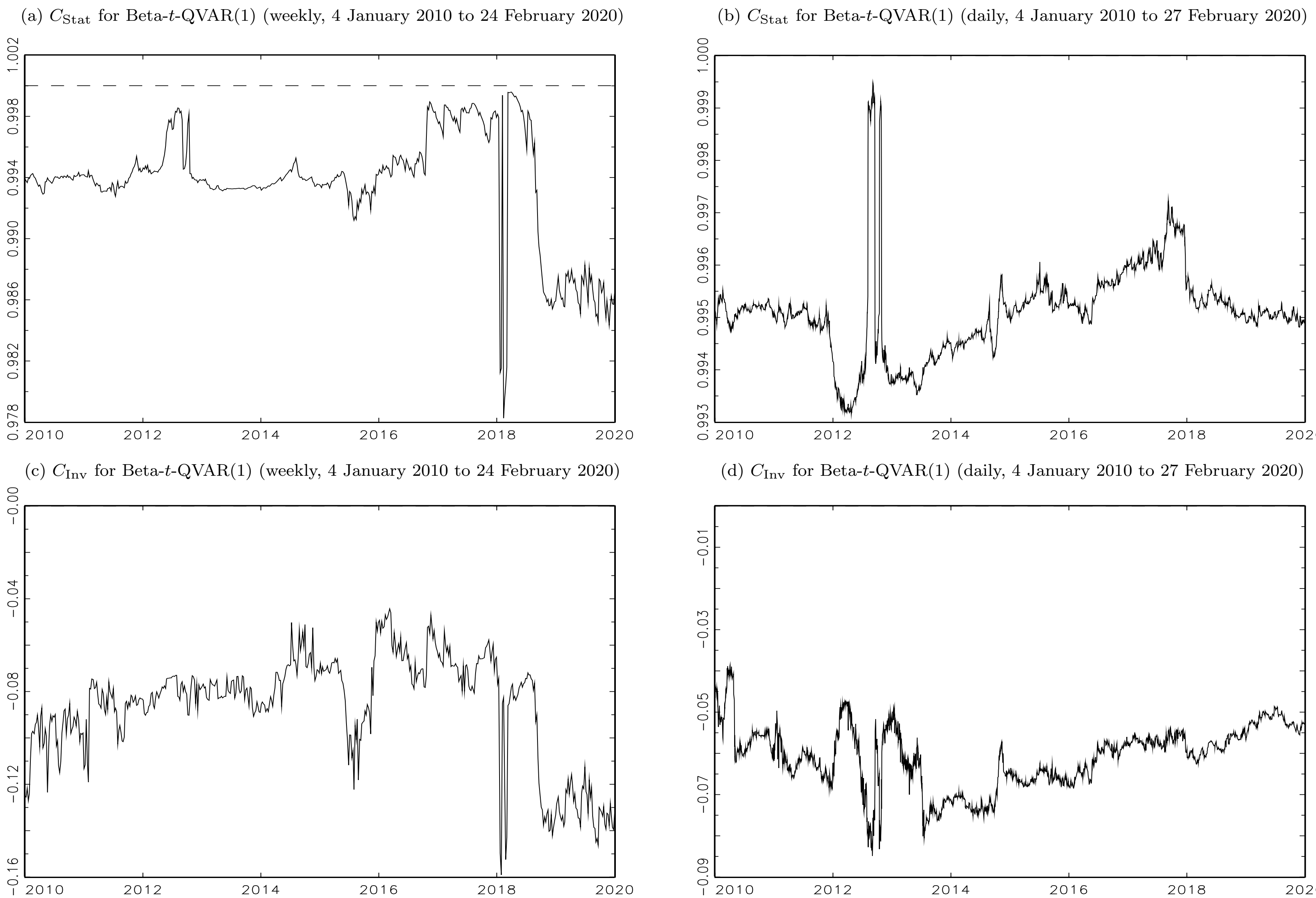


Fig. 5. Rolling-window-based estimates of C_{Stat} and C_{Inv} for the forecasting window.
Notes: Beta-t-QVAR(1) is presented since the AIC, BIC, and HQC of this model are superior to the same metrics of other Beta-t-QVAR(1) specifications.

Table 3. Out-of-sample prediction accuracy for weekly data (forecasting window: 4 January 2010 to 24 February 2020)

Model	MSE_1	MSE_2	QLIKE	R^2LOG	MAE_1	MAE_2
Beta-*t*-QVAR(1)	0.4362	20.8477	0.2426	0.5820	0.4899	1.9073
Beta-*t*-QVAR(1)-lev	0.4324	20.9979	0.2432	0.5694	0.4811	1.8749
Beta-*t*-QVAR(1)-M	0.4313	20.7235	0.2393	0.5729	0.4850	1.8919
Beta-*t*-QVAR(1)-M-lev	**0.4253**	**20.7816**	**0.2399**	**0.5598**	**0.4755**	**1.8511**
Beta-*t*-EGARCH(1,1) zero mean	0.6090***	23.8141**	0.3365***	0.8523***	0.5969***	2.2899***
Beta-*t*-EGARCH(1,1)-lev zero mean	0.4719+	19.8236	0.2624	0.6875***	0.5359***	2.0559**
Beta-*t*-EGARCH(1,1) constant mean	0.5647***	23.0201**	0.3085***	0.7755***	0.5677***	2.1915***
Beta-*t*-EGARCH(1,1)-lev constant mean	0.4714+	20.2329	0.2586	0.6669***	0.5293***	2.0413**
Beta-*t*-EGARCH(1,1)-M	0.5581***	22.8170**	0.3073***	0.7746***	0.5650***	2.1775***
Beta-*t*-EGARCH(1,1)-M-lev	0.4651	20.0838	0.2569	0.6629***	0.5258***	2.0249**
A-PARCH(1,1) zero mean	0.4373	20.2683	0.2521	0.6302**	0.4974+	1.8923
A-PARCH(1,1) constant mean	0.4370	20.3250	0.2514	0.6257**	0.4960	1.8894
A-PARCH(1,1)-M	0.4313	20.2729	0.2495	0.6179**	0.4903	1.8647
Gaussian-GARCH(1,1) zero mean	0.5816***	21.9378	0.3272***	0.8880***	0.5957***	2.2549***
Gaussian-GARCH(1,1)-lev zero mean	0.4514	18.5132	0.2596	0.7033***	0.5277**	2.0139+
Gaussian-GARCH(1,1) constant mean	0.5328***	20.8898	0.2996***	0.8112***	0.5664***	2.1544***
Gaussian-GARCH(1,1)-lev constant mean	0.4441	18.4878	0.2546	0.6854***	0.5208**	1.9894
Gaussian-GARCH(1,1)-M	0.5296***	20.7867	0.2995***	0.8107***	0.5652***	2.1463***
Gaussian-GARCH(1,1)-M-lev	0.4422	18.4652	0.2550	0.6869***	0.5202**	1.9828
t-GARCH(1,1) zero mean	0.6064***	23.2001**	0.3395***	0.8921***	0.6040***	2.2900***
t-GARCH(1,1)-lev zero mean	0.4623	19.0508	0.2669	0.7182***	0.5345***	2.0294*
t-GARCH(1,1) constant mean	0.5577***	22.1752	0.3109***	0.8155***	0.5761***	2.1947***
t-GARCH(1,1)-lev constant mean	0.4607	19.2999	0.2634	0.7014***	0.5309***	2.0225*
t-GARCH(1,1)-M	0.5526***	21.9898	0.3100***	0.8155***	0.5744***	2.1859***
t-GARCH(1,1)-M-lev	0.4586	19.2604	0.2639	0.7039***	0.5297***	2.0143*

Notes: Bold numbers indicate the loss functions of Beta-*t*-QVAR(1)-M-lev, which we use as the reference specification of Beta-*t*-QVAR. The Diebold–Mariano test compares the loss functions of Beta-*t*-QVAR(1)-M-lev with the loss functions of Beta-*t*-EGARCH(1,1), A-PARCH(1,1), Gaussian-GARCH(1,1), and *t*-GARCH(1,1). For each loss function, the significance of the Diebold–Mariano test statistic is presented. Note that the lowest loss function values do not always correspond to Beta-*t*-QVAR(1)-M-lev. If the difference between loss functions is significant, then the prediction accuracy of Beta-*t*-QVAR(1)-M-lev is superior to the prediction accuracy of Beta-*t*-EGARCH(1,1), A-PARCH(1,1), Gaussian-GARCH(1,1), or *t*-GARCH(1,1). The prediction accuracies of Beta-*t*-EGARCH(1,1), A-PARCH(1,1), Gaussian-GARCH(1,1), and *t*-GARCH(1,1) are never significantly superior to the prediction accuracy of Beta-*t*-QVAR(1)-M-lev. $^+$, *, **, and *** indicate significance at the 15%, 10%, 5%, and 1% levels, respectively.

Table 4. Out-of-sample prediction accuracy for daily data (forecasting window: 4 January 2010 to 27 February 2020)

Model	MSE_1	MSE_2	QLIKE	R^2LOG	MAE_1	MAE_2
Beta-*t*-QVAR(1)	0.1211	2.1494	0.3053	0.7215	0.2434	0.4546
Beta-*t*-QVAR(1)-lev	0.1192	2.0711	0.3012	0.7234	0.2452	0.4590
Beta-*t*-QVAR(1)-M	0.1213	2.1531	0.3065	0.7260	0.2436	0.4537
Beta-*t*-QVAR(1)-M-lev	**0.1168**	**2.0491**	**0.2995**	**0.7191**	**0.2443**	**0.4562**
Beta-*t*-EGARCH zero mean	0.1551***	2.3138***	0.3785***	0.9475***	0.2852***	0.5305***
Beta-*t*-EGARCH-lev zero mean	0.1382***	2.1394***	0.3340***	0.8573***	0.2731***	0.5134***
Beta-*t*-EGARCH constant mean	0.1519***	2.3002***	0.3669***	0.9136***	0.2807***	0.5252***
Beta-*t*-EGARCH-lev constant mean	0.1381***	2.1472***	0.3329***	0.8479***	0.2720***	0.5122***
Beta-*t*-EGARCH-M	0.1511***	2.2952***	0.3660***	0.9114***	0.2801***	0.5238***
Beta-*t*-EGARCH-M-lev	0.1378***	2.1450***	0.3323***	0.8463***	0.2716***	0.5116***
A-PARCH(1,1) zero mean	0.1323***	2.1083+	0.3339***	0.8680***	0.2715***	0.4991***
A-PARCH(1,1) constant mean	0.1327***	2.1095+	0.3348***	0.8710***	0.2720***	0.5000***
A-PARCH(1,1)-M	0.1324***	2.1085+	0.3341***	0.8690***	0.2716***	0.4993***
Gaussian-GARCH(1,1) zero mean	0.1498***	2.2180**	0.3746***	0.9936***	0.2881***	0.5296***
Gaussian-GARCH(1,1)-lev zero mean	0.1406***	2.1348	0.3440***	0.9188***	0.2785***	0.5196***
Gaussian-GARCH(1,1) constant mean	0.1479***	2.2152**	0.3672***	0.9686***	0.2851***	0.5268***
Gaussian-GARCH(1,1)-lev constant mean	0.1402***	2.1302	0.3433***	0.9147***	0.2786***	0.5195***
Gaussian-GARCH(1,1)-M	0.1471***	2.2080**	0.3665***	0.9666***	0.2845***	0.5252***
Gaussian-GARCH(1,1)-lev-M	0.1402***	2.1328	0.3433***	0.9147***	0.2785***	0.5196***
t-GARCH(1,1) zero mean	0.1538***	2.2616***	0.3796***	0.9889***	0.2894***	0.5348***
t-GARCH(1,1)-lev zero mean	0.1443***	2.1669**	0.3443***	0.9100***	0.2802***	0.5281***
t-GARCH(1,1) constant mean	0.1510***	2.2477***	0.3689***	0.9591***	0.2856***	0.5306***
t-GARCH(1,1)-lev constant mean	0.1433***	2.1603**	0.3424***	0.9019***	0.2790***	0.5257***
t-GARCH(1,1)-M	0.1503***	2.2428***	0.3681***	0.9570***	0.2851***	0.5294***
t-GARCH(1,1)-lev-M	0.1434***	2.1623**	0.3426***	0.9026***	0.2791***	0.5260***

Notes: Bold numbers indicate the loss functions of Beta-*t*-QVAR(1)-M-lev, which we use as the reference specification of Beta-*t*-QVAR. The Diebold–Mariano test compares the loss functions of Beta-*t*-QVAR(1)-M-lev with the loss functions of Beta-*t*-EGARCH(1,1), A-PARCH(1,1), Gaussian-GARCH(1,1), and *t*-GARCH(1,1). For each loss function, the significance of the Diebold–Mariano test statistic is presented. Note that the lowest loss function values do not always correspond to Beta-*t*-QVAR(1)-M-lev. If the difference between loss functions is significant, then the prediction accuracy of Beta-*t*-QVAR(1)-M-lev is superior to the prediction accuracy of Beta-*t*-EGARCH(1,1), A-PARCH(1,1), Gaussian-GARCH(1,1), or *t*-GARCH(1,1). The prediction accuracies of Beta-*t*-EGARCH(1,1), A-PARCH(1,1), Gaussian-GARCH(1,1), and *t*-GARCH(1,1) are never significantly superior to the prediction accuracy of Beta-*t*-QVAR(1)-M-lev. $^+$, *, **, and *** indicate significance at the 15%, 10%, 5%, and 1% levels, respectively.

8. Conclusions

In the present paper, a new bivariate score-driven model has been introduced, that captures dynamic interaction effects between the risk premium and volatility of financial assets, in order to improve volatility forecast precision. Beta-t-QVAR is a score-driven location plus scale model, for which the asymptotic properties of the ML estimator have been proven. The one-step ahead volatility forecasting performance of the new score-driven model has been illustrated for weekly and daily DJIA data for the period of January 1985 to February 2020. The prediction accuracy of Beta-t-QVAR has been compared to the volatility prediction accuracies of Beta-t-EGARCH, A-PARCH, and GARCH. The results of the empirical illustration have indicated that the volatility forecasts of Beta-t-QVAR are superior to the volatility forecasts of Beta-t-EGARCH, A-PARCH, and GARCH for several loss functions.

The empirical illustration of this paper is specific in several aspects, which may be generalized in future works: (i) A particular stock market index of 30 large United States firms is used. Alternatively, stock indices involving the prices of more firms, e.g. the S&P 500 (Standard & Poor's 500 Index), or other financial assets may be considered. (ii) A particular set of rolling estimation and forecasting windows are used for the data. (iii) Only one-step ahead volatility forecasts are used. (iv) Motivated by studies in the body of literature on forecasting asset price volatility, a particular set of competing dynamic volatility models are used. (v) For all econometric models of this paper, first-order dynamics for all filters are used. (vi) A particular definition of IRFs is used. Alternative definitions of IRFs from the literature on nonlinear econometric models may be considered. (vii) Motivated by studies from the literature on forecasting asset price volatility, a particular set of loss functions is used. (viii) A particular statistical test of out-of-sample forecast accuracy comparison is used.

Due to these specifications, the empirical illustration cannot be interpreted as a general finding on Beta-t-QVAR that is valid for all financial assets. Our objectives are (i) to highlight the potential usefulness of Beta-t-QVAR, as an alternative to Beta-t-EGARCH, A-PARCH, and GARCH, and (ii) to present new conditions of statistical inference for score-driven location plus scale models.

Acknowledgements

The authors are thankful for the comments of Matthew Copley. Blazsek and Licht acknowledge funding from Universidad Francisco Marroquín. Escribano acknowledges funding from Ministerio de Economía, Industria y Competitividad (ECO2016-00105-001 and MDM 2014-0431), Comunidad de Madrid (MadEco-CM S2015/HUM-3444), and Agencia Estatal de Investigación (2019/00419/001).

References

Alsmeyer, G., 2003. On the Harris recurrence of iterated random Lipschitz functions and related convergence rate results. *Journal of Theoretical Probability* 16, 217–247. doi: 10.1023/A:1022290807360.

Ayala, A., Blazsek, S., 2018a. Equity market neutral hedge funds and the stock market: an application of score-driven copula models. *Applied Economics* 50, 4005–4023. doi: 10.1080/00036846.2018.1440062.

Ayala, A., Blazsek, S., 2018b. Score-driven copula models for portfolios of two risky assets. *The European Journal of Finance* 24, 1861–1884. doi: 10.1080/1351847X.2018.1464488.

Ayala, A., Blazsek, S., 2019a. Score-driven currency exchange rate seasonality as applied to the Guatemalan Quetzal/US Dollar. *SERIEs* 10, 65–92. doi: 10.1007/s13209-018-0186-0.

Ayala, A., Blazsek, S., 2019b. Score-driven models of stochastic seasonality in location and scale: An application case study of the Indian rupee to USD exchange rate. *Applied Economics* 51, 4083–4103. doi: 10.1080/00036846.2019.1588952.

Black, F., 1976. Studies of stock market volatility changes. 1976 Proceedings of the American Statistical Association Business and Economic Statistics Section.

Blasques, F., Gorgi, P., Koopman, S.J., Wintenberger, O., 2018. Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics* 12, 1019–1052. doi: 10.1214/18-EJS1416.

Blasques, F., Koopman, S.J., Lucas, A., 2015. Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika* 102, 325–343. doi: 10.1093/biomet/asu076.

Blasques, F., Koopman, S.J., Lucas, A., 2017. Maximum likelihood estimation for score-driven models. TI 2014-029/III Tinbergen Institute Discussion Paper. https://papers.tinbergen.nl/14029.pdf.

Blasques, F., Lucas, A., van Vlodrop, A., 2020. Finite sample optimality of score-driven volatility models: some Monte Carlo evidence. *Econometrics and Statistics*. doi: 10.1016/j.ecosta.2020.03.010.

Blazsek, S., Carrizo, D., Eskildsen, R., Gonzalez, H., 2018. Forecasting rate of return after extreme values when using AR-t-GARCH and QAR-Beta-t-EGARCH. *Finance Research Letters* 24, 193–198. doi: 10.1016/j.frl.2017.09.006.

Blazsek, S., Chavez, H., Mendez, C., 2016. Model stability and forecast performance of Beta-t-EGARCH. *Applied Economics Letters* 23, 1219–1223. doi: 10.1080/13504851.2016.1145343.

Blazsek, S., Escribano, A, Licht, A., 2019. Co-integration and common trends analysis with score-driven models: an application to the Federal Funds effective rate and the US inflation rate. Working Paper 19-08, University Carlos III of Madrid, Department of Economics. https://e-archivo.uc3m.es/bitstream/handle/10016/28451/we1908.pdf.

Blazsek, S., Escribano, A, Licht, A., 2020. Identification of seasonal effects in impulse responses using score-driven multivariate location models. *Journal of Econometric Methods*. doi: 10.1515/jem-2020-0003.

Blazsek, S., Haddad, M., 2020. Estimation and statistical performance of Markov-switching score-driven volatility models: the case of G20 stock markets. Discussion Paper 1/2020, Francisco Marroquin University, School of

Business, 2020. https://en.ufm.edu/wp-content/uploads/2020/02/Blazsek-Haddad-2020-GESG-Disussion-Paper-no.1-2020-1.pdf

Blazsek, S., Hernandez, H., 2018. Analysis of electricity prices for Central American countries using dynamic conditional score models. *Empirical Economics* 55, 1807–1848. doi: 10.1007/s00181-017-1341-3.

Blazsek, S., Ho, H.-C., Liu, S.-P., 2018. Score-driven Markov-switching EGARCH models: an application to systematic risk analysis. *Applied Economics* 50, 6047–6060. doi: 10.1080/00036846.2018.1488073.

Blazsek, S., Licht, A., 2018. Robustness of score-driven location and scale models to extreme observations: an application to the Chinese stock market. *Financial Statistical Journal* 1, 507–516. doi: 10.24294/fsj.v1i2.699.

Blazsek, S., Licht, A., 2020. Dynamic conditional score models: a review of their applications. *Applied Economics* 52, 1181–1199. doi: 10.1080/00036846.2019.1659498.

Blazsek, S., Mendoza, V., 2016. QARMA-Beta-t-EGARCH versus ARMA-GARCH: an application to S&P 500. *Applied Economics* 48, 1119–1129. doi: 10.1080/00036846.2015.1093086.

Blazsek, S., Monteros, L.A., 2017a. Event-study analysis by using dynamic conditional score models. *Applied Economics* 49, 4530–4541. doi: 10.1080/00036846.2017.1284996.

Blazsek, S., Monteros, L.A., 2017b. Dynamic conditional score models of degrees of freedom: filtering with score-driven heavy tails. *Applied Economics* 49, 5426–5440. doi: 10.1080/00036846.2017.1307935.

Blazsek, S., Villatoro, M., 2015. Is Beta-t-EGARCH(1,1) superior to GARCH(1,1)? *Applied Economics* 47, 1764–1774. doi: 10.1080/00036846.2014.1000536.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327. doi: 10.1016/0304-4076(86)90063-1.

Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics* 69, 542–547. doi: 10.2307/1925546.

Brandt, A., 1986. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability* 18, 211–220. doi: 10.2307/1427243.

Caivano, M., Harvey, A., 2014. Time-series models with an EGB2 conditional distribution. *Journal of Time Series Analysis* 35, 558–571. doi: 10.1111/jtsa.12081.

Creal, D., Koopman, S.J., Lucas, A., 2008. A general framework for observation driven time-varying parameter models. Tinbergen Institute Discussion Paper 08-108/4. https://papers.tinbergen.nl/08108.pdf.

Creal, D., Koopman, S.J., Lucas, A., 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics* 29, 552–563. doi: 10.1198/jbes.2011.10070.

Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28, 777–795. doi: 10.1002/jae.1279.

Creal, D., Schwaab, B., Koopman, S.J., Lucas, A., 2014. Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *The Review of Economics and Statistics* 96, 898–915. doi: 10.1162/REST_a_00393.

Cox, D. R., 1981. Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics* 8, 93–115.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 253–263. doi: 10.1080/07350015.1995.10524599.

Ding, Z., Granger, C.W.J., Engle, R.F., 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106. doi: 10.1016/0927-5398(93)90006-D.

Elton, J.H., 1990. A multiplicative ergodic theorem for Lipschitz maps. *Stochastic Processes and their Applications* 34, 39–47. doi: 10.1016/0304-4149(90)90055-W.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007. doi: 10.2307/1912773.

Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: representation, estimation, and testing. *Econometrica* 55, 251–276. doi: 10.2307/1913236.

Engle, R.F., Lilien, D.M., Robins, R.P., 1987. Estimating time-varying risk premia in the term structure: the ARCH-M model. *Econometrica* 55, 391–407. doi: 10.2307/1913242.

Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* 48, 1779–1801. doi: 10.1111/j.1540-6261.1993.tb05128.x.

Granger, C.W.J., 1981. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16, 121–130. doi: 10.1016/0304-4076(81)90079-8.

Hansen, P.R., Lunde, A., 2005. A forecast comparison of volatility models: does anything beat a GARCH(1,1)?" *Journal of Applied Econometrics* 20, 873–889. doi: 10.1002/jae.800.

Harvey, A.C., 2013. *Dynamic Models for Volatility and Heavy Tails.* Cambridge University Press, Cambridge.

Harvey A.C., Chakravarty, T., 2008. Beta-t-(E)GARCH. Cambridge working papers in Economics 0840, Faculty of Economics, University of Cambridge, Cambridge. http://www.econ.cam.ac.uk/research/repec/cam/pdf/cwpe0840.pdf.

Harvey, A., Lange, R.J., 2017. Volatility modeling with a generalized t distribution. *Journal of Time Series Analysis* 38, 175–190. doi: 10.1111/jtsa.12224.

Harvey, A.C., Lange, R.J., 2018. Modeling the interactions between volatility and returns using EGARCH-M. *Journal of Time Series Analysis* 39, 909–919. doi: 10.1111/jtsa.12419.

Harvey, A., Sucarrat, G., 2014. EGARCH models with fat tails, skewness and leverage. *Computational Statistics & Data Analysis* 76, 320–338. doi: 10.1016/j.csda.2013.09.022.

Herwartz, H., Lütkepohl, H., 2000 Multivariate volatility analysis of VW stock prices. *International Journal of Intelligent Systems in Accounting, Finance & Management* 9, 35–54. doi: 10.1002/(SICI)1099-1174(200003)9:1<35::AID-ISAF176>3.0.CO;2-V.

Lütkepohl, H., 2005. *New Introduction to Multivariate Time Series Analysis.* Springer-Verlag, Berlin Heidelberg.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–370. doi: 10.2307/2938260.

Patton, A.J., 2011. Data-based ranking of realised volatility estimators. *Journal of Econometrics* 161, 284–303. doi: 10.1016/j.jeconom.2010.12.010.

Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611. doi: 10.1093/biomet/52.3-4.591.

Straumann, D., Mikosch, T. 2006, Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics* 34, 2449–2495. doi: 10.1214/009053606000000803.

White, H., 2001. *Asymptotic Theory for Econometricians,* revised edition. Academic Press, San Diego.

Wooldridge, J.M., 1994. Estimation and inference for dependent processes. In: Engle, R.F., McFadden, D.L. (Eds.), *Handbook of Econometrics*, Volume 4, pp. 2639–2738. North-Holland, Amsterdam.

Wooldridge, J.M., 2010. *Econometric Analysis of Cross Section and Panel Data*, second edition. MIT Press, Cambridge.

Supplementary Material:

Prediction accuracy of bivariate score-driven risk premium and volatility filters: an illustration for the Dow Jones

Szabolcs Blazsek [a], Alvaro Escribano [b,*], and Adrian Licht [a]

[a] School of Business, Universidad Francisco Marroquín, Guatemala City (Guatemala), Guatemala

[b] Department of Economics, Universidad Carlos III de Madrid, Getafe (Madrid), Spain

Abstract: In this paper, we introduce Beta-t-QVAR (quasi-vector autoregression) for the joint modelling of score-driven location and scale. Asymptotic theory of the maximum likelihood (ML) estimator is presented, and sufficient conditions of consistency and asymptotic normality of ML are proven. For the joint score-driven modelling of risk premium and volatility, Dow Jones Industrial Average (DJIA) data are used in an empirical illustration. Prediction accuracy of Beta-t-QVAR is superior to the prediction accuracies of Beta-t-EGARCH (exponential generalized AR conditional heteroscedasticity), A-PARCH (asymmetric power ARCH), and GARCH (generalized ARCH). The empirical results motivate the use of Beta-t-QVAR for the valuation of DJIA options.

Keywords: Volatility; Risk premium; Dynamic conditional score; Generalized autoregressive score

JEL Classification: C22, C58

Corresponding author. *E-mail address:* alvaroe@eco.uc3m.es (Alvaro Escribano).

1. Introduction

The consequences of (A1), (A4), and the functional forms of $u_{\mu,t}$ and $u_{\lambda,t}$ are:

$$E\left[(u_{j,t})^{2-i}\left(\frac{\partial u_{k,t}}{\partial l_t}\right)^i\right] < \infty \tag{1}$$

$$E\left[\left(\frac{\partial u_{j,t}}{\partial m_t}\right)^{2-i}\left(\frac{\partial u_{k,t}}{\partial l_t}\right)^i\right] < \infty \tag{2}$$

for $i = 0, 1, 2$, and $j, k, l, m = \mu, \lambda$. In the proofs of the Supplementary Material, we use the finite second moments and covariances of the derivatives of u_t with respect to θ_t. Based on Eqs. (1)-(2), we show that those moments are finite by using the chain rule:

$$\frac{\partial u_t}{\partial \theta_t'} = \frac{\partial u_t}{\partial(\mu_t, \lambda_t)}\frac{\partial(\mu_t, \lambda_t)'}{\partial \theta_t'} \tag{3}$$

The elements of $\partial u_t/\partial(\mu_t, \lambda_t)$ are bounded (Section 5.2). The elements of $\partial(\mu_t, \lambda_t)'/\partial\theta_t'$, for the Beta-$t$-QVAR(1)-M-lev specification, are bounded because:

$$\left|\frac{\partial \mu_t}{\partial \mu_t^\dagger}\right| = |\beta_1| < \infty \tag{4}$$

$$\left|\frac{\partial \mu_t}{\partial \lambda_t^\dagger}\right| = |\beta_2 \exp(\omega + \lambda_t^\dagger)| = |\beta_2 \exp(\lambda_t)| < |\beta_2 \exp(\lambda_{\max})| < \infty \tag{5}$$

$$\left|\frac{\partial \lambda_t}{\partial \mu_t^\dagger}\right| = 0 < \infty \tag{6}$$

$$\left|\frac{\partial \lambda_t}{\partial \lambda_t^\dagger}\right| = 1 < \infty \tag{7}$$

For the other Beta-t-QVAR(1) specifications, the latter result can be applied by using parameter restrictions in Eqs. (4)-(7). Since the product of bounded functions is also a bounded function,

$$E\left[(u_{j,t})^{2-i}\left(\frac{\partial u_{k,t}}{\partial \theta_{l,t}}\right)^{i}\right] < \infty \tag{8}$$

$$E\left[\left(\frac{\partial u_{j,t}}{\partial \theta_{m,t}}\right)^{2-i}\left(\frac{\partial u_{k,t}}{\partial \theta_{l,t}}\right)^{i}\right] < \infty \tag{9}$$

for $i = 0, 1, 2$, and $j, k, l, m = 1, 2$. We define:

$$X_{t-1} \equiv \frac{\partial \theta_t}{\partial (\theta_{t-1})'} = \Phi + \Psi \frac{\partial u_{t-1}}{\partial (\theta_{t-1})'} \tag{10}$$

From Eq. (8), we have that $E(X_{t-1}) < \infty$ and $E(X_{t-1} \otimes X_{t-1}) < \infty$. We also define:

$$X^*_{t-1} \equiv \frac{\partial \theta_t}{\partial (\theta_{t-1})'} = \Phi + \Psi^*_t \frac{\partial u_{t-1}}{\partial (\theta_{t-1})'} \equiv \Phi + \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} + \psi^* \operatorname{sgn}(-\epsilon_{t-1}) \end{bmatrix} \frac{\partial u_{t-1}}{\partial (\theta_{t-1})'} \tag{11}$$

where sgn$(\cdot)$ is the signum function. From Eq. (8), $E(X^*_{t-1}) < \infty$ and $E(X^*_{t-1} \otimes X^*_{t-1}) < \infty$. In relation to Eqs. (8) to (11), we refer to Creal et al. (2013), Harvey (2013), and Blasques et al. (2017).

The remainder of the Supplementary Material is organized as follows: In Section 2, it is proven in Propositions 7(a-b) that filter $\hat{\theta}_t$ converges exponentially almost surely (e.a.s.) to the strictly stationary and ergodic sequence $\theta_t(\Theta_0)$ for $T \to \infty$. Propositions 7(a-b) are applied to Proposition 2. In Section 3, in Propositions 8(a-b), it is proven that $E[G_t(\Theta_0)'] < \infty$. Those proofs involve technical details that are used in the proofs of $E[H_t(\Theta_0)] = \text{Var}[G_t(\Theta_0)')] = E[G_t(\Theta_0)'G_t(\Theta_0)] < \infty$ in Propositions 9(a-b). In Section 4, Propositions 9(a-b) are proven, which are applied to Proposition 5. In Section 5, it is proven in Propositions 10(a-b) that $G_t(\hat{\Theta})'$ converges e.a.s. to the strictly stationary and ergodic sequence $G_t(\Theta_0)'$ for $T \to \infty$. Propositions 10(a-b) are applied to Proposition 5. In Section 6, we present how the conditions of the following result can be obtained: $H_t(\hat{\Theta})$ converges e.a.s. to the strictly stationary and ergodic sequence $H_t(\Theta_0)$ for $T \to \infty$. This result is applied to Proposition 4. In Sections 7 and 8, the dynamics of y_t and σ_t, respectively, are presented by using explicit formulas.

2. Stationarity and ergodicity of $\hat{\theta}_t$

In Propositions 7(a-b), conditions of e.a.s. convergence of $\hat{\theta}_t$ to the unique strictly stationary and ergodic sequence $\theta_t(\Theta_0)$ are presented. Results from Brandt (1986), Elton (1990), and Alsmeyer (2003) are used. We also refer to Straumann and Mikosch (2006), and Blasques et al. (2017, 2018).

Proposition 7(a): For Beta-t-QVAR(1) and Beta-t-QVAR(1)-M, $\hat{\theta}_t$ converges e.a.s. to the unique strictly stationary and ergodic sequence $\theta_t(\Theta_0)$, i.e. $||\hat{\theta}_t - \theta_t(\Theta_0)||_1 \xrightarrow{e.a.s.} 0$ for $t \to \infty$, when the following conditions hold: (i) $E(\ln^+ ||\Psi u||_1) < \infty$, where $||\Psi u||_1 \equiv \sup\{||\Psi u_1 x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\}$, and $\ln^+(x) = 0$ if $0 \leq x \leq 1$ and $\ln^+(x) = \ln(x)$ if $x > 1$. (ii) $E(\ln^+ ||X||_1) < \infty$, where $||X||_1 \equiv \sup\{||X_1 x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\}$. (iii) The Lyapunov exponent is negative:

$$\inf_{n\geq 1} \left\{ n^{-1} E \left[\ln \left|\left| \prod_{t=1}^{n} \frac{\partial \theta_t}{\partial (\theta_{t-1})'} \right|\right|_1 \right] \right\} = \inf_{n\geq 1} \left\{ n^{-1} E \left(\ln \left|\left| \prod_{t=1}^{n} X_{t-1} \right|\right|_1 \right) \right\} < 0 \tag{12}$$

(iv) Ψu_t is strictly stationary and ergodic. (v) X_t is strictly stationary and ergodic. In this paper, matrix norm $||A||_1 = \max_{1\leq j\leq 2} \sum_{i=1}^{2} |a_{i,j}|$ is used, where $A = \{a_{i,j}\}$ for $i, j = 1, 2$.

Proof: It follows from Brandt (1986), Elton (1990), and Alsmeyer (2003). (i) and (ii) hold for the Beta-t-QVAR models. (iii) is a maintained assumption, which can be verified empirically. (iv) is due to the properties of the score functions of Beta-t-QVAR. (v) is due to the following arguments: According to Eqs. (10) and (3), by using the properties of the score functions (Section 5.2), X_t is an $\mathcal{F}$-measurable function of $(\epsilon_1, \ldots, \epsilon_t)$. Variable X_t is strictly stationary and ergodic, because ϵ_t is strictly stationary and ergodic (White 2001, Theorem 3.35). *QED*

Proposition 7(b): For Beta-t-QVAR(1)-lev and Beta-t-QVAR(1)-M-lev, $\hat{\theta}_t$ converges e.a.s. to the unique strictly stationary and ergodic sequence $\theta_t(\Theta_0)$, i.e. $||\hat{\theta}_t - \theta_t(\Theta_0)||_1 \xrightarrow{e.a.s.} 0$ for $t \to \infty$, if: (i) $E(\ln^+ ||g(u)||_1) < \infty$, where $||g(u)||_1 \equiv \sup\{||\Psi u_1 + \psi^*[0, \mathrm{sgn}(-\epsilon_1)(u_{\lambda,1} + 1)]' x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\}$. (ii) $E(\ln^+ ||X^*||_1) < \infty$, where $||X^*||_1 \equiv \sup\{||X_1^* x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\}$. (iii) The Lyapunov exponent is negative:

$$\inf_{n\geq 1} \left\{ n^{-1} E \left[\ln \left|\left| \prod_{t=1}^{n} \frac{\partial \theta_t}{\partial (\theta_{t-1})'} \right|\right|_1 \right] \right\} = \inf_{n\geq 1} \left\{ n^{-1} E \left(\ln \left|\left| \prod_{t=1}^{n} X_{t-1}^* \right|\right|_1 \right) \right\} < 0 \tag{13}$$

(iv) $g(u_t)$ is strictly stationary and ergodic. (v) X_t^* is strictly stationary and ergodic.

Proof: It follows from Brandt (1986), Elton (1990), and Alsmeyer (2003). (i) and (ii) hold for the Beta-t-QVAR models. (iii) is a maintained assumption, which can be verified empirically. (iv) is due to the properties of the score functions of Beta-t-QVAR. (v) is due to the following arguments: According to Eqs. (11) and (3), by using the properties of the score functions (Section 5.2), X_t^* is an $\mathcal{F}$-measurable function of $(\epsilon_1, \ldots, \epsilon_t)$. Variable X_t^* is strictly stationary and ergodic, because ϵ_t is strictly stationary and ergodic (White 2001, Theorem 3.35). *QED*

3. Time-invariant expected value of the gradient, $\mathbf{E}[\mathbf{G_t}(\boldsymbol{\Theta_0})'] < \infty$

In Propositions 8(a-b), the conditions of the time-invariant expected value of the gradient for Beta-t-QVAR(1) are presented. Expected value $E[G_t(\Theta_0)'] < \infty$ if $E(\partial\theta_t(\Theta_0)/\partial\Theta') < \infty$ (Harvey 2013). For the proofs, arguments from the work of Harvey (2013) are extended.

Proposition 8(a): For Beta-t-QVAR(1) and Beta-t-QVAR(1)-M, the expected value of the gradient is time-invariant if the maximum modulus of eigenvalues of $E(X_{t-1})$ is less than one.

Proof: (i) In the first part of the proof, we focus on the derivatives of θ_t with respect to $\psi_{i,j}$, which are in $G_t(\Theta_0)'$ and $H_t(\Theta_0)$. The derivatives of θ_t with respect to $\psi_{i,j}$ are:

$$\frac{\partial\theta_t}{\partial\psi_{i,j}} = \Phi\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}} + \Psi\frac{\partial u_{t-1}}{\partial\psi_{i,j}} + W_{i,j}u_{t-1} \tag{14}$$

for $i, j = 1, 2$; $W_{i,j}$ is a 2×2 matrix, in which element (i, j) is one, and the rest of the elements are zero. Therefore, $W_{i,j}u_{t-1}$ is the j-th element of u_{t-1}. By using the chain rule, from Eq. (14):

$$\frac{\partial\theta_t}{\partial\psi_{i,j}} = \left(\Phi + \Psi\frac{\partial u_{t-1}}{\partial\theta'_{t-1}}\right)\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}} + W_{i,j}u_{t-1} = X_{t-1}\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}} + W_{i,j}u_{t-1} \tag{15}$$

The expectation of the latter equation, that is conditional on $\mathcal{F}_{t-2}$, is:

$$E\left(\frac{\partial\theta_t}{\partial\psi_{i,j}}|\mathcal{F}_{t-2}\right) = E(X_{t-1}|\mathcal{F}_{t-2})\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}} + W_{i,j}E(u_{t-1}|\mathcal{F}_{t-2}) \tag{16}$$

where $\partial\theta_{t-1}/\partial\psi_{i,j}$ is outside the conditional expectation, because it is determined by $\mathcal{F}_{t-2}$. We consider the unconditional expectation of Eq. (16). Firstly, we focus on the term

$$E\left[E(X_{t-1}|\mathcal{F}_{t-2})\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}}\right] \equiv E(A\times B) = E\left[\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\right] \tag{17}$$

where

$$a_{1,1} = E\left(\phi_{1,1} + \psi_{1,1}\frac{\partial u_{\mu,t-1}}{\partial\mu^{\dagger}_{t-1}} + \psi_{1,2}\frac{\partial u_{\lambda,t-1}}{\partial\mu^{\dagger}_{t-1}}|\mathcal{F}_{t-2}\right) \tag{18}$$

$$a_{1,2} = E\left(\phi_{1,2} + \psi_{1,1}\frac{\partial u_{\mu,t-1}}{\partial\lambda^{\dagger}_{t-1}} + \psi_{1,2}\frac{\partial u_{\lambda,t-1}}{\partial\lambda^{\dagger}_{t-1}}|\mathcal{F}_{t-2}\right) \tag{19}$$

$$a_{2,1} = E\left(\phi_{2,1} + \psi_{2,1}\frac{\partial u_{\mu,t-1}}{\partial\mu^{\dagger}_{t-1}} + \psi_{2,2}\frac{\partial u_{\lambda,t-1}}{\partial\mu^{\dagger}_{t-1}}|\mathcal{F}_{t-2}\right) \tag{20}$$

$$a_{2,2} = E\left(\phi_{2,2} + \psi_{2,1}\frac{\partial u_{\mu,t-1}}{\partial\lambda^{\dagger}_{t-1}} + \psi_{2,2}\frac{\partial u_{\lambda,t-1}}{\partial\lambda^{\dagger}_{t-1}}|\mathcal{F}_{t-2}\right) \tag{21}$$

$$b_1 = \frac{\partial\mu^{\dagger}_{t-1}}{\partial\psi_{i,j}} \tag{22}$$

$$b_2 = \frac{\partial\lambda^{\dagger}_{t-1}}{\partial\psi_{i,j}} \tag{23}$$

Hence, $E(A\times B)$ from Eq. (17) can be written as:

$$E(A\times B) = E(A)E(B) + \mathrm{Cov}(A,B) = E(A)E(B) + \begin{bmatrix} \mathrm{Cov}(a_{1,1},b_1) + \mathrm{Cov}(a_{1,2},b_2) \\ \mathrm{Cov}(a_{2,1},b_1) + \mathrm{Cov}(a_{2,2},b_2) \end{bmatrix} \tag{24}$$

In the following, we show that each term in Eq. (24) is finite. Firstly,

$$E(A) = E[E(X_{t-1}|\mathcal{F}_{t-2})] = E(X_{t-1}) < \infty \tag{25}$$

due to Eqs. (8) and (10). Secondly,

$$E(B) = E\left(\frac{\partial \theta_{t-1}}{\partial \psi_{i,j}}\right) = E\left[\frac{\partial\left(\sum_{j=0}^{\infty}\Phi^j \Psi u_{t-j-2}\right)}{\partial \psi_{i,j}}\right] < \infty \tag{26}$$

where the second equality is under the assumption of covariance stationary θ_t, and finiteness is due to Eq. (8). Thirdly, we study the covariances in Eq. (24), and we prove the finiteness of the variances of the variables in Eqs. (18) to (23). We consider from Eq. (18):

$$\mathrm{Var}(a_{1,1}) = \mathrm{Var}\left[E\left(\phi_{1,1} + \psi_{1,1}\frac{\partial u_{\mu,t-1}}{\partial \mu^{\dagger}_{t-1}} + \psi_{1,2}\frac{\partial u_{\lambda,t-1}}{\partial \mu^{\dagger}_{t-1}}|\mathcal{F}_{t-2}\right)\right] \tag{27}$$

$$= E\left[E^2\left(\phi_{1,1} + \psi_{1,1}\frac{\partial u_{\mu,t-1}}{\partial \mu^{\dagger}_{t-1}} + \psi_{1,2}\frac{\partial u_{\lambda,t-1}}{\partial \mu^{\dagger}_{t-1}}|\mathcal{F}_{t-2}\right)\right] - E^2\left[E\left(\phi_{1,1} + \psi_{1,1}\frac{\partial u_{\mu,t-1}}{\partial \mu^{\dagger}_{t-1}} + \psi_{1,2}\frac{\partial u_{\lambda,t-1}}{\partial \mu^{\dagger}_{t-1}}|\mathcal{F}_{t-2}\right)\right]$$

$$= E\left[E^2\left(\phi_{1,1} + \psi_{1,1}\frac{\partial u_{\mu,t-1}}{\partial \mu^{\dagger}_{t-1}} + \psi_{1,2}\frac{\partial u_{\lambda,t-1}}{\partial \mu^{\dagger}_{t-1}}|\mathcal{F}_{t-2}\right)\right] - E^2\left(\phi_{1,1} + \psi_{1,1}\frac{\partial u_{\mu,t-1}}{\partial \mu^{\dagger}_{t-1}} + \psi_{1,2}\frac{\partial u_{\lambda,t-1}}{\partial \mu^{\dagger}_{t-1}}\right)$$

$$\leq E\left\{E\left[\left(\phi_{1,1} + \psi_{1,1}\frac{\partial u_{\mu,t-1}}{\partial \mu^{\dagger}_{t-1}} + \psi_{1,2}\frac{\partial u_{\lambda,t-1}}{\partial \mu^{\dagger}_{t-1}}\right)^2|\mathcal{F}_{t-2}\right]\right\} - E^2\left(\phi_{1,1} + \psi_{1,1}\frac{\partial u_{\mu,t-1}}{\partial \mu^{\dagger}_{t-1}} + \psi_{1,2}\frac{\partial u_{\lambda,t-1}}{\partial \mu^{\dagger}_{t-1}}\right)$$

$$= E\left[\left(\phi_{1,1} + \psi_{1,1}\frac{\partial u_{\mu,t-1}}{\partial \mu^{\dagger}_{t-1}} + \psi_{1,2}\frac{\partial u_{\lambda,t-1}}{\partial \mu^{\dagger}_{t-1}}\right)^2\right] - E^2\left(\phi_{1,1} + \psi_{1,1}\frac{\partial u_{\mu,t-1}}{\partial \mu^{\dagger}_{t-1}} + \psi_{1,2}\frac{\partial u_{\lambda,t-1}}{\partial \mu^{\dagger}_{t-1}}\right) < \infty$$

for which both terms of the latter equation are finite due to Eqs. (8) and (9), and the inequality follows from the Jensen inequality. For the finiteness of $\mathrm{Var}(a_{1,2})$, $\mathrm{Var}(a_{2,1})$, and $\mathrm{Var}(a_{2,2})$, the same arguments are used, hence the proofs are not reported. Next, we consider:

$$\mathrm{Var}(b_1) = \mathrm{Var}\left(\frac{\partial \mu^{\dagger}_{t-1}}{\partial \psi_{i,j}}\right) < \infty \tag{28}$$

where the finiteness is due to the following arguments. Under the assumption of covariance stationarity, $\mu^{\dagger}_{t-1}$ is a linear combination of lags of the score functions. Therefore, the derivative with respect to $\psi_{i,j}$ involves a linear combination of the score functions, which have finite variance due to the properties of the score functions for Beta-t-QVAR. For the finiteness of $\mathrm{Var}(b_2)$, the same arguments are used.

Therefore, the unconditional expectation of Eq. (16) can be written as:

$$E\left[E\left(\frac{\partial\theta_t}{\partial\psi_{i,j}}|\mathcal{F}_{t-2}\right)\right] = E\left[E(X_{t-1}|\mathcal{F}_{t-2})\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}}\right] + W_{i,j}E[E(u_{t-1}|\mathcal{F}_{t-2})] \tag{29}$$

which is equivalent to

$$E\left(\frac{\partial\theta_t}{\partial\psi_{i,j}}\right) = E(X_{t-1})E\left(\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}}\right) + \mathrm{Cov}(A,B) + W_{i,j}E(u_{t-1}) \tag{30}$$

Due to Eq. (8), and due to the previous arguments from Eqs. (17) to (28), $E(\partial\theta_t/\partial\psi_{i,j}) < \infty$ if the maximum modulus of eigenvalues of $E(X_{t-1})$ is less than one.

(ii) In the second part of the proof, we focus on the derivatives of θ_t with respect to $\phi_{i,j}$, which are in $G_t(\Theta_0)'$ and $H_t(\Theta_0)$. The derivatives of θ_t with respect to $\phi_{i,j}$ are:

$$\frac{\partial\theta_t}{\partial\phi_{i,j}} = \left(\Phi + \Psi\frac{\partial u_{t-1}}{\partial\theta'_{t-1}}\right)\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}} + W_{i,j}\theta_{t-1} = X_{t-1}\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}} + W_{i,j}\theta_{t-1} \tag{31}$$

for $i,j = 1,2$. Therefore, $W_{i,j}\theta_{t-1}$ is the j-th element of θ_{t-1}. The expectation of the latter equation, that is conditional on $\mathcal{F}_{t-2}$, is:

$$E\left(\frac{\partial\theta_t}{\partial\phi_{i,j}}|\mathcal{F}_{t-2}\right) = E(X_{t-1}|\mathcal{F}_{t-2})\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}} + W_{i,j}E(\theta_{t-1}|\mathcal{F}_{t-2}) \tag{32}$$

where $\partial\theta_{t-1}/\partial\phi_{i,j}$ is outside the conditional expectation, because it is determined by $\mathcal{F}_{t-2}$. We use the unconditional expectation of Eq. (32), for which the finiteness of the terms is proven by using the same arguments as in the first part of this proof. Therefore, $E(\partial\theta_t/\partial\phi_{i,j})$ is finite if the maximum modulus of eigenvalues of $E(X_{t-1})$ is less than one. *QED*

Proposition 8(b): For Beta-t-QVAR(1)-lev and Beta-t-QVAR(1)-M-lev, the expected value of the gradient is time-invariant if the maximum modulus of eigenvalues of $E(X^*_{t-1})$ is less than one.

Proof: We only summarize the key details of the proof, since the proof of Proposition 8(b) is similar to the proof of Proposition 8(a). The derivatives of θ_t with respect to $\psi_{i,j}$, $\phi_{i,j}$, and ψ^* are:

$$\frac{\partial \theta_t}{\partial \psi_{i,j}} = X^*_{t-1}\frac{\partial \theta_{t-1}}{\partial \psi_{i,j}} + W_{i,j}u_{t-1} \tag{33}$$

$$\frac{\partial \theta_t}{\partial \phi_{i,j}} = X^*_{t-1}\frac{\partial \theta_{t-1}}{\partial \phi_{i,j}} + W_{i,j}\theta_{t-1} \tag{34}$$

$$\frac{\partial \theta_t}{\partial \psi^*} = X^*_{t-1}\frac{\partial \theta_{t-1}}{\partial \psi^*} + W_{2,2}u_{t-1} + \begin{bmatrix} 0 \\ \mathrm{sgn}(-\epsilon_{t-1}) \end{bmatrix} \tag{35}$$

respectively, for $i, j = 1, 2$. Eqs. (33) to (35) indicate that the condition for the existence of the unconditional mean of the gradient is that the maximum modulus of eigenvalues of $E(X^*_{t-1})$ is less than one. To validate this statement, firstly, the expectations of Eqs. (33) to (35), that are conditional on $\mathcal{F}_{t-2}$, are written. The law of iterated expectations, and Eqs. (8), (9), and (11) are used, to show that the unconditional means of Eqs. (33) to (35) are finite. *QED*

4. Time-invariant expected value of the Hessian matrix, $\mathbf{E[H_t(\Theta_0)']} < \infty$

In Propositions 9(a-b), the conditions of the time-invariant expected value of the Hessian, $E[H_t(\Theta_0)] < \infty$, for the Beta-$t$-QVAR(1) models are presented. Expected value $E[H_t(\Theta_0)] < \infty$ if expected value $E\{[\partial\theta_t(\Theta_0)/(\partial\Theta\partial\Theta')]'\} < \infty$. For the proofs, arguments from the work of Harvey (2013) are extended. We also refer to the works of Blasques et al. (2017, 2018).

Proposition 9(a): For Beta-t-QVAR(1) and Beta-t-QVAR(1)-M, the expected value of the Hessian matrix is time-invariant if the maximum modulus of eigenvalues of $E(X_{t-1} \otimes X_{t-1})$ is less than one, where $\otimes$ is the Kronecker product.

Proof: (i) In the first part, we focus on the following derivative, which contributes to $H_t(\Theta_0)$:

$$\frac{\partial \theta_t}{\partial \psi_{i,j}}\left(\frac{\partial \theta_t}{\partial \psi_{k,l}}\right)' = X_{t-1}\frac{\partial \theta_{t-1}}{\partial \psi_{i,j}}\left(\frac{\partial \theta_{t-1}}{\partial \psi_{k,l}}\right)'(X_{t-1})' + X_{t-1}\frac{\partial \theta_{t-1}}{\partial \psi_{i,j}}u'_{t-1}W'_{k,l} \tag{36}$$

$$+W_{i,j}u_{t-1}\left(\frac{\partial\theta_{t-1}}{\partial\psi_{k,l}}\right)'X'_{t-1}+W_{i,j}u_{t-1}u'_{t-1}W'_{k,l}$$

which, by using the equation $\mathrm{vec}(ABC)=(C'\otimes A)\mathrm{vec}(B)$, Eq. (36) can be written as:

$$\mathrm{vec}\left[\frac{\partial\theta_t}{\partial\psi_{i,j}}\left(\frac{\partial\theta_t}{\partial\psi_{k,l}}\right)'\right]=(X_{t-1}\otimes X_{t-1})\mathrm{vec}\left[\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}}\left(\frac{\partial\theta_{t-1}}{\partial\psi_{k,l}}\right)'\right] \tag{37}$$

$$+[(W_{k,l}u_{t-1})\otimes X_{t-1}]\mathrm{vec}\left(\frac{\partial\theta_{t-1}}{\partial\psi_{k,l}}\right)+[X_{t-1}\otimes(W_{i,j}u_{t-1})]\mathrm{vec}\left(\frac{\partial\theta_{t-1}}{\partial\psi_{k,l}}\right)'+\mathrm{vec}\left(W_{i,j}u_{t-1}u'_{t-1}W'_{k,l}\right)$$

The expectation of the latter equation, that is conditional on $\mathcal{F}_{t-2}$, is:

$$E\left\{\mathrm{vec}\left[\frac{\partial\theta_t}{\partial\psi_{i,j}}\left(\frac{\partial\theta_t}{\partial\psi_{k,l}}\right)'\right]|\mathcal{F}_{t-2}\right\}=E(X_{t-1}\otimes X_{t-1}|\mathcal{F}_{t-2})\mathrm{vec}\left[\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}}\left(\frac{\partial\theta_{t-1}}{\partial\psi_{k,l}}\right)'\right] \tag{38}$$

$$+E[(W_{k,l}u_{t-1})\otimes X_{t-1}|\mathcal{F}_{t-2}]\mathrm{vec}\left(\frac{\partial\theta_{t-1}}{\partial\psi_{k,l}}\right)+E[X_{t-1}\otimes(W_{i,j}u_{t-1})|\mathcal{F}_{t-2}]\mathrm{vec}\left(\frac{\partial\theta_{t-1}}{\partial\psi_{k,l}}\right)'$$

$$+\mathrm{vec}\left[W_{i,j}E\left(u_{t-1}u'_{t-1}|\mathcal{F}_{t-2}\right)W'_{k,l}\right]$$

where $\mathrm{vec}[(\partial\theta_{t-1}/\partial\psi_{i,j})(\partial\theta_{t-1}/\partial\psi_{k,l})']$, $\mathrm{vec}[(\partial\theta_{t-1}/\partial\psi_{i,j})]$, and $\mathrm{vec}[(\partial\theta_{t-1}/\partial\psi_{k,l})']$ are outside the conditional expectations, because they are determined by $\mathcal{F}_{t-2}$. We use the unconditional expectation of Eq. (38). For the unconditional expectation of the first three terms on the right side of Eq. (38), covariances appear in the same way as explained for Eq. (24). We summarize those covariance terms by using the notation Cov*. The unconditional expectation of Eq. (38) is:

$$E\left\{\mathrm{vec}\left[\frac{\partial\theta_t}{\partial\psi_{i,j}}\left(\frac{\partial\theta_t}{\partial\psi_{k,l}}\right)'\right]\right\}=E(X_{t-1}\otimes X_{t-1})E\left\{\mathrm{vec}\left[\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}}\left(\frac{\partial\theta_{t-1}}{\partial\psi_{k,l}}\right)'\right]\right\} \tag{39}$$

$$+E\left[(W_{k,l}u_{t-1})\otimes X_{t-1}\right]E\left\{\mathrm{vec}\left(\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}}\right)\right\}+E\left[X_{t-1}\otimes(W_{i,j}u_{t-1})\right]E\left\{\mathrm{vec}\left[\left(\frac{\partial\theta_{t-1}}{\partial\psi_{k,l}}\right)'\right]\right\}$$

$$+\mathrm{vec}\left[W_{i,j}E\left(u_{t-1}u'_{t-1}\right)W'_{k,l}\right]+\mathrm{Cov}^*$$

The terms on the right side of Eq. (39) are finite due to Eqs. (8) to (11), by using the same arguments as for Propositions 8(a-b). Therefore, $E[(\partial\theta_t/\partial\psi_{i,j})(\partial\theta_t/\partial\psi_{k,l})']$ is finite if the maximum modulus of eigenvalues of $E(X_{t-1}\otimes X_{t-1})$ is less than one.

(ii) In the second part, we focus on the following derivative, which contributes to $H_t(\Theta_0)$:

$$\text{vec}\left[\frac{\partial\theta_t}{\partial\phi_{i,j}}\left(\frac{\partial\theta_t}{\partial\phi_{k,l}}\right)'\right] = (X_{t-1}\otimes X_{t-1})\text{vec}\left[\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}}\left(\frac{\partial\theta_{t-1}}{\partial\phi_{k,l}}\right)'\right] \tag{40}$$

$$+[(W_{k,l}\theta_{t-1})\otimes X_{t-1}]\text{vec}\left(\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}}\right)+[X_{t-1}\otimes(W_{i,j}\theta_{t-1})]\text{vec}\left[\left(\frac{\partial\theta_{t-1}}{\partial\phi_{k,l}}\right)'\right]+\text{vec}\left(W_{i,j}\theta_{t-1}\theta_{t-1}'W_{k,l}'\right)$$

The expectation of the latter equation, that is conditional on $\mathcal{F}_{t-2}$, is:

$$E\left\{\text{vec}\left[\frac{\partial\theta_t}{\partial\phi_{i,j}}\left(\frac{\partial\theta_t}{\partial\phi_{k,l}}\right)'\right]|\mathcal{F}_{t-2}\right\} = E(X_{t-1}\otimes X_{t-1}|\mathcal{F}_{t-2})\text{vec}\left[\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}}\left(\frac{\partial\theta_{t-1}}{\partial\phi_{k,l}}\right)'\right] \tag{41}$$

$$+E[(W_{k,l}\theta_{t-1})\otimes X_{t-1}|\mathcal{F}_{t-2}]\text{vec}\left(\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}}\right)+E[X_{t-1}\otimes(W_{i,j}\theta_{t-1})|\mathcal{F}_{t-2}]\text{vec}\left[\left(\frac{\partial\theta_{t-1}}{\partial\phi_{k,l}}\right)'\right]$$

$$+\text{vec}\left[W_{i,j}E\left(\theta_{t-1}\theta_{t-1}'|\mathcal{F}_{t-2}\right)W_{k,l}'\right]$$

By using the additional $\text{Cov}^\dagger$ term, which represents the covariances from the first three terms on the right side of Eq. (41), the unconditional expectation of Eq. (41) is:

$$E\left\{\text{vec}\left[\frac{\partial\theta_t}{\partial\phi_{i,j}}\left(\frac{\partial\theta_t}{\partial\phi_{k,l}}\right)'\right]\right\} = E(X_{t-1}\otimes X_{t-1})E\left\{\text{vec}\left[\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}}\left(\frac{\partial\theta_{t-1}}{\partial\phi_{k,l}}\right)'\right]\right\} \tag{42}$$

$$+E[(W_{k,l}\theta_{t-1})\otimes X_{t-1}]E\left[\text{vec}\left(\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}}\right)\right]+E[X_{t-1}\otimes(W_{i,j}\theta_{t-1})]E\left\{\text{vec}\left[\left(\frac{\partial\theta_{t-1}}{\partial\phi_{k,l}}\right)'\right]\right\}$$

$$+\text{vec}\left[W_{i,j}E\left(\theta_{t-1}\theta_{t-1}'\right)W_{k,l}'\right]+\text{Cov}^\dagger$$

The terms on the right side of Eq. (42) are finite due to Eqs. (8) and (9), by using the same arguments as for Propositions 8(a-b). Therefore, $E[(\partial\theta_t/\partial\phi_{i,j})(\partial\theta_t/\partial\phi_{k,l})']$ is finite if the maximum modulus of eigenvalues of $E(X_{t-1}\otimes X_{t-1})$ is less than one. *QED*

Proposition 9(b): For Beta-t-QVAR(1)-lev and Beta-t-QVAR(1)-M-lev, the expected value of the Hessian matrix is time-invariant if the maximum modulus of eigenvalues of $E(X_{t-1}^*\otimes X_{t-1}^*)$ is less than one.

Proof: We only summarize the key details of the proof, since the proof of Proposition 9(b) is very similar to the proof of Proposition 9(a). By using Eqs. (33) to (35), dynamic equations of the second-derivatives with respect to $\psi_{i,j}$, $\phi_{i,j}$, and ψ^*, respectively, can be written as follows:

$$\mathrm{vec}\left[\frac{\partial\theta_t}{\partial\psi_{i,j}}\left(\frac{\partial\theta_t}{\partial\psi_{k,l}}\right)'\right] = (X^*_{t-1}\otimes X^*_{t-1})\mathrm{vec}\left[\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}}\left(\frac{\partial\theta_{t-1}}{\partial\psi_{k,l}}\right)'\right] \tag{43}$$

$$+[(W_{k,l}u_{t-1})\otimes X^*_{t-1}]\mathrm{vec}\left(\frac{\partial\theta_{t-1}}{\partial\psi_{k,l}}\right)+[X^*_{t-1}\otimes(W_{i,j}u_{t-1})]\mathrm{vec}\left(\frac{\partial\theta_{t-1}}{\partial\psi_{k,l}}\right)'+\mathrm{vec}\left(W_{i,j}u_{t-1}u'_{t-1}W'_{k,l}\right)$$

and

$$\mathrm{vec}\left[\frac{\partial\theta_t}{\partial\phi_{i,j}}\left(\frac{\partial\theta_t}{\partial\phi_{k,l}}\right)'\right] = (X^*_{t-1}\otimes X^*_{t-1})\mathrm{vec}\left[\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}}\left(\frac{\partial\theta_{t-1}}{\partial\phi_{k,l}}\right)'\right] \tag{44}$$

$$+[(W_{k,l}\theta_{t-1})\otimes X^*_{t-1}]\mathrm{vec}\left(\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}}\right)+[X^*_{t-1}\otimes(W_{i,j}\theta_{t-1})]\mathrm{vec}\left[\left(\frac{\partial\theta_{t-1}}{\partial\phi_{k,l}}\right)'\right]+\mathrm{vec}\left(W_{i,j}\theta_{t-1}\theta'_{t-1}W'_{k,l}\right)$$

and

$$\mathrm{vec}\left[\frac{\partial\theta_t}{\partial\psi^*}\left(\frac{\partial\theta_t}{\partial\psi^*}\right)'\right] = (X^*_{t-1}\otimes X^*_{t-1})\mathrm{vec}\left[\frac{\partial\theta_{t-1}}{\partial\psi^*}\left(\frac{\partial\theta_{t-1}}{\partial\psi^*}\right)'\right]+[(W_{2,2}u_{t-1})\otimes X^*_{t-1}]\mathrm{vec}\left(\frac{\partial\theta_{t-1}}{\partial\psi^*}\right) \tag{45}$$

$$+\left\{\begin{bmatrix}0\\ \mathrm{sgn}(-\epsilon_{t-1})\end{bmatrix}\otimes X^*_{t-1}\right\}\mathrm{vec}\left(\frac{\partial\theta_{t-1}}{\partial\psi^*}\right)+(X^*_{t-1}\otimes W_{2,2}u_{t-1})\mathrm{vec}\left[\left(\frac{\partial\theta_{t-1}}{\partial\psi^*}\right)'\right]$$

$$+(W_{2,2}\otimes W_{2,2})\mathrm{vec}(u_{t-1}u'_{t-1})+\left\{\begin{bmatrix}0\\ \mathrm{sgn}(-\epsilon_{t-1})\end{bmatrix}\otimes W_{2,2}\right\}\mathrm{vec}(u_{t-1})$$

$$+\left\{X^*_{t-1}\otimes\begin{bmatrix}0\\ \mathrm{sgn}(-\epsilon_{t-1})\end{bmatrix}\right\}\mathrm{vec}\left[\left(\frac{\partial\theta_{t-1}}{\partial\psi^*}\right)'\right]+\left\{W_{2,2}\otimes\begin{bmatrix}0\\ \mathrm{sgn}(-\epsilon_{t-1})\end{bmatrix}\right\}\mathrm{vec}(u'_{t-1})$$

$$+\mathrm{vec}\left\{\begin{bmatrix}0\\ \mathrm{sgn}(-\epsilon_{t-1})\end{bmatrix}\begin{bmatrix}0\\ \mathrm{sgn}(-\epsilon_{t-1})\end{bmatrix}'\right\}$$

Eqs. (43) to (45) indicate that the condition for the existence of the unconditional mean of the gradient is that the maximum modulus of eigenvalues of $E(X^*_{t-1}\otimes X^*_{t-1})$ is less than one. To validate this statement, firstly, the expectations of Eqs. (43) to (45), that are conditional on $\mathcal{F}_{t-2}$, are written. The law of iterated expectations, and Eqs. (8), (9), and (11) are used, to show that the unconditional means of all terms of Eqs. (43) to (45) are finite. *QED*

5. Stationarity and ergodicity of $\mathbf{G_t(\hat{\Theta})'}$

In Propositions 10(a-b), the conditions of e.a.s. convergence of $G_t(\hat{\Theta})'$ to the unique strictly stationary and ergodic sequence $G_t(\Theta_0)'$ for the Beta-t-QVAR(1) models are presented. Stationarity and ergodicity of $G_t(\hat{\Theta})'$ are implied by the stationarity and ergodicity of $\partial\hat{\theta}_t/\partial\Theta'$ to the unique strictly stationary and ergodic sequence $\partial\theta_t(\Theta_0)/\Theta'$. For the proofs, results from Brandt (1986), Elton (1990), and Alsmeyer (2003) are used. We also refer to Straumann and Mikosch (2006), and Blasques et al. (2017, 2018).

Proposition 10(a): For Beta-t-QVAR(1) and Beta-t-QVAR(1)-M, Eqs. (15) and (31) are:

$$\frac{\partial\theta_t}{\partial\psi_{i,j}} = X_{t-1}\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}} + W_{i,j}u_{t-1} \tag{46}$$

$$\frac{\partial\theta_t}{\partial\phi_{i,j}} = X_{t-1}\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}} + W_{i,j}\theta_{t-1} \tag{47}$$

for $i, j = 1, 2$. Vector $G_t(\hat{\Theta})'$ converges e.a.s. to the unique strictly stationary and ergodic sequence $G_t(\Theta_0)'$, i.e. $||G_t(\hat{\Theta})' - G_t(\Theta_0)'||_1 \xrightarrow{e.a.s.} 0$ for $t \to \infty$, when the following hold:

(i) $E(\ln^+ ||W_{i,j}u||_1) < \infty$ and $E(\ln^+ ||W_{i,j}\theta||_1) < \infty$ for $i, j = 1, 2$, where respectively

$$||W_{i,j}u||_1 \equiv \sup\{||W_{i,j}u_1x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\} \tag{48}$$

$$||W_{i,j}\theta||_1 \equiv \sup\{||W_{i,j}\theta_1x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\} \tag{49}$$

(ii) $E(\ln^+ ||X + W_{i,j}(\partial u/\partial u_j)||_1) < \infty$ and $E(\ln^+ ||X + W_{i,j}(\partial\theta/\partial\theta_j)||_1) < \infty$ for $i, j = 1, 2$, where respectively

$$||X + W_{i,j}(\partial u/\partial u_j)||_1 \equiv \sup\{||[X_2 + W_{i,j}(\partial u_2/\partial u_{j,1})]x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\} \tag{50}$$

$$||X + W_{i,j}(\partial\theta/\partial\theta_j)||_1 \equiv \sup\{||[X_2 + W_{i,j}(\partial\theta_2/\partial\theta_{j,1})]x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\} \tag{51}$$

(iii) the following Lyapunov exponents are negative:

$$\inf_{n\geq 1}\left\{n^{-1}E\left[\ln\left|\left|\prod_{t=1}^{n}\frac{\partial\left(\frac{\partial\theta_t}{\partial\psi_{i,j}}\right)}{\partial\left(\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}}\right)}\right|\right|_1\right]\right\} \tag{52}$$

$$= \inf_{n\geq 1}\left\{n^{-1}E\left(\ln\left|\left|\prod_{t=1}^{n}\left[X_{t-1} + W_{i,j}\left(\frac{\partial u_{t-1}}{\partial u_{j,t-2}}\right)\right]\right|\right|_1\right)\right\} < 0$$

$$\inf_{n\geq 1}\left\{n^{-1}E\left[\ln\left|\left|\prod_{t=1}^{n}\frac{\partial\left(\frac{\partial\theta_t}{\partial\phi_{i,j}}\right)}{\partial\left(\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}}\right)}\right|\right|_1\right]\right\} \tag{53}$$

$$= \inf_{n\geq 1}\left\{n^{-1}E\left(\ln\left|\left|\prod_{t=1}^{n}\left[X_{t-1} + W_{i,j}\left(\frac{\partial\theta_{t-1}}{\partial\theta_{j,t-2}}\right)\right]\right|\right|_1\right)\right\} < 0$$

for $i, j = 1, 2$.

(iv) $W_{i,j}u_{t-1}$ and $W_{i,j}\theta_{t-1}$ are strictly stationary and ergodic.

(v) $X_t + W_{i,j}(\partial u_t/\partial u_{j,t-1})$ and $X_t + W_{i,j}(\partial\theta_t/\partial\theta_{j,t-1})$ are strictly stationary and ergodic.

Proof: It follows from the proofs of Brandt (1986), Elton (1990), and Alsmeyer (2003). (i) and (ii) hold for the Beta-t-QVAR models of this paper. (iii) is a maintained assumption, which can be verified empirically. (iv) is due to the properties of the score functions of Beta-t-QVAR. (v) is due to the following arguments: By using the properties of the score functions (Section 5.2), $X_t + W_{i,j}(\partial u_t/\partial u_{j,t-1})$ and $X_t + W_{i,j}(\partial\theta_t/\partial\theta_{j,t-1})$ are $\mathcal{F}$-measurable functions of $(\epsilon_1, \ldots, \epsilon_t)$. Variables $X_t + W_{i,j}(\partial u_t/\partial u_{j,t-1})$ and $X_t + W_{i,j}(\partial\theta_t/\partial\theta_{j,t-1})$ are strictly stationary and ergodic, because ϵ_t is strictly stationary and ergodic (White 2001, Theorem 3.35). *QED*

Proposition 10(b): For Beta-t-QVAR(1)-lev and Beta-t-QVAR(1)-M-lev, Eqs. (33) to (35) are:

$$\frac{\partial\theta_t}{\partial\psi_{i,j}} = X^*_{t-1}\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}} + W_{i,j}u_{t-1} \tag{54}$$

$$\frac{\partial\theta_t}{\partial\phi_{i,j}} = X^*_{t-1}\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}} + W_{i,j}\theta_{t-1} \tag{55}$$

$$\frac{\partial\theta_t}{\partial\psi^*} = X^*_{t-1}\frac{\partial\theta_{t-1}}{\partial\psi^*} + \mathrm{sgn}(-\epsilon_{t-1})W_{2,2}u_{t-1} + \begin{bmatrix} 0 \\ \mathrm{sgn}(-\epsilon_{t-1}) \end{bmatrix} \tag{56}$$

respectively, for $i,j = 1,2$. Vector $G_t(\hat{\Theta})'$ converges e.a.s. to the unique strictly stationary and ergodic sequence $G_t(\Theta_0)'$, i.e. $||G_t(\hat{\Theta})' - G_t(\Theta_0)'||_1 \xrightarrow{e.a.s.} 0$ for $t \to \infty$, when:

(i) We assume $E(\ln^+ ||W_{i,j}u||_1) < \infty$ for $i,j = 1,2$, $E(\ln^+ ||W_{i,j}\theta||_1) < \infty$ for $i,j = 1,2$, and $E(\ln^+ ||\mathrm{sgn}(-\epsilon)W_{2,2}u + [0, \mathrm{sgn}(-\epsilon)]'||_1) < \infty$, where respectively

$$||W_{i,j}u||_1 \equiv \sup\{||W_{i,j}u_1x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\} \tag{57}$$

$$||W_{i,j}\theta||_1 \equiv \sup\{||W_{i,j}\theta_1x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\} \tag{58}$$

$$||\mathrm{sgn}(-\epsilon)W_{2,2}u + [0, \mathrm{sgn}(-\epsilon)]'||_1 \tag{59}$$

$$\equiv \sup\{||\mathrm{sgn}(-\epsilon_1)W_{2,2}u_1 + [0, \mathrm{sgn}(-\epsilon_1)]'x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\}$$

(ii) $E(\ln^+ ||X^* + W_{i,j}(\partial u/\partial u_j)||_1) < \infty$ for $i,j = 1,2$, $E(\ln^+ ||X^* + W_{i,j}(\partial\theta/\partial\theta_j)||_1) < \infty$ for $i,j = 1,2$, and $E(\ln^+ ||X^* + \mathrm{sgn}(-\epsilon)W_{2,2}(\partial u/\partial u_j)||_1) < \infty$, where respectively

$$||X^* + W_{i,j}(\partial u/\partial u_j)||_1 \equiv \sup\{||[X^*_2 + W_{i,j}(\partial u_2/\partial u_{j,1})]x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\} \tag{60}$$

$$||X^* + W_{i,j}(\partial\theta/\partial\theta_j)||_1 \equiv \sup\{||[X_2^* + W_{i,j}(\partial\theta_2/\partial\theta_{j,1})]x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\} \tag{61}$$

$$||X^* + \text{sgn}(-\epsilon)W_{2,2}(\partial u/\partial u_j)||_1 \tag{62}$$

$$\equiv \sup\{||[X_2^* + \text{sgn}(-\epsilon_1)W_{2,2}(\partial u_2/\partial u_{j,1})]x||_1 : x \in \mathbb{R}^2, ||x||_1 \leq 1\}$$

(iii) the following Lyapunov exponents are negative:

$$\inf_{n\geq 1}\left\{n^{-1}E\left[\ln\left|\left|\prod_{t=1}^{n}\frac{\partial\left(\frac{\partial\theta_t}{\partial\psi_{i,j}}\right)}{\partial\left(\frac{\partial\theta_{t-1}}{\partial\psi_{i,j}}\right)}\right|\right|_1\right]\right\} \tag{63}$$

$$= \inf_{n\geq 1}\left\{n^{-1}E\left(\ln\left|\left|\prod_{t=1}^{n}\left[X_{t-1}^* + W_{i,j}\left(\frac{\partial u_{t-1}}{\partial u_{j,t-2}}\right)\right]\right|\right|_1\right)\right\} < 0$$

$$\inf_{n\geq 1}\left\{n^{-1}E\left[\ln\left|\left|\prod_{t=1}^{n}\frac{\partial\left(\frac{\partial\theta_t}{\partial\phi_{i,j}}\right)}{\partial\left(\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}}\right)}\right|\right|_1\right]\right\} \tag{64}$$

$$= \inf_{n\geq 1}\left\{n^{-1}E\left(\ln\left|\left|\prod_{t=1}^{n}\left[X_{t-1}^* + W_{i,j}\left(\frac{\partial\theta_{t-1}}{\partial\theta_{j,t-2}}\right)\right]\right|\right|_1\right)\right\} < 0$$

$$\inf_{n\geq 1}\left\{n^{-1}E\left[\ln\left|\left|\prod_{t=1}^{n}\frac{\partial\left(\frac{\partial\theta_t}{\partial\psi^*}\right)}{\partial\left(\frac{\partial\theta_{t-1}}{\partial\psi^*}\right)}\right|\right|_1\right]\right\} \tag{65}$$

$$= \inf_{n\geq 1}\left\{n^{-1}E\left(\ln\left|\left|\prod_{t=1}^{n}\left[X_{t-1}^* + \text{sgn}(-\epsilon_{t-1})W_{2,2}\left(\frac{\partial u_{t-1}}{\partial u_{j,t-2}}\right)\right]\right|\right|_1\right)\right\} < 0$$

for $i, j = 1, 2$.

(iv) $W_{i,j}u_{t-1}$, $W_{i,j}\theta_{t-1}$, and $\text{sgn}(-\epsilon)W_{2,2}u + [0, \text{sgn}(-\epsilon)]'$ are strictly stationary and ergodic.

(v) $X_t^* + W_{i,j}(\partial u_t/\partial u_{j,t-1})$, $X_t^* + W_{i,j}(\partial \theta_t/\partial \theta_{j,t-1})$, and $X_t^* + \text{sgn}(-\epsilon_t)W_{2,2}(\partial u_t/\partial u_{j,t-1})$ are strictly stationary and ergodic.

Proof: It follows from the works of Brandt (1986), Elton (1990), and Alsmeyer (2003). (i) and (ii) hold for the Beta-t-QVAR models of this paper. (iii) is a maintained assumption, which can be verified empirically. (iv) is due to the properties of the score functions of Beta-t-QVAR. (v) is due to the following arguments: By using the properties of the score functions (Section 5.2), $X_t^* + W_{i,j}(\partial u_t/\partial u_{j,t-1})$, $X_t^* + W_{i,j}(\partial \theta_t/\partial \theta_{j,t-1})$, and $X_t^* + \text{sgn}(-\epsilon_t)W_{2,2}(\partial u_t/\partial u_{j,t-1})$ are $\mathcal{F}$-measurable functions of $(\epsilon_1, \ldots, \epsilon_t)$. Variables $X_t^* + W_{i,j}(\partial u_t/\partial u_{j,t-1})$, $X_t^* + W_{i,j}(\partial \theta_t/\partial \theta_{j,t-1})$, and $X_t^* + \text{sgn}(-\epsilon_t)W_{2,2}(\partial u_t/\partial u_{j,t-1})$ are strictly stationary and ergodic, because ϵ_t is strictly stationary and ergodic (White 2001, Theorem 3.35). *QED*

6. Stationarity and ergodicity of $H_t(\Theta_0)$

In this subsection, we present the conditions according to which $H_t(\Theta_0)$ is strictly stationary and ergodic. Firstly, we refer to Eqs. (37) and (40) for Beta-t-QVAR and Beta-t-QVAR-M:

$$\text{vec}\left[\frac{\partial \theta_t}{\partial \psi_{i,j}}\left(\frac{\partial \theta_t}{\partial \psi_{k,l}}\right)'\right] = (X_{t-1} \otimes X_{t-1})\text{vec}\left[\frac{\partial \theta_{t-1}}{\partial \psi_{i,j}}\left(\frac{\partial \theta_{t-1}}{\partial \psi_{k,l}}\right)'\right] \tag{66}$$

$$+[(W_{k,l}u_{t-1}) \otimes X_{t-1}]\text{vec}\left(\frac{\partial \theta_{t-1}}{\partial \psi_{k,l}}\right) + [X_{t-1} \otimes (W_{i,j}u_{t-1})]\text{vec}\left(\frac{\partial \theta_{t-1}}{\partial \psi_{k,l}}\right)' + \text{vec}\left(W_{i,j}u_{t-1}u_{t-1}'W_{k,l}'\right)$$

and

$$\text{vec}\left[\frac{\partial \theta_t}{\partial \phi_{i,j}}\left(\frac{\partial \theta_t}{\partial \phi_{k,l}}\right)'\right] = (X_{t-1} \otimes X_{t-1})\text{vec}\left[\frac{\partial \theta_{t-1}}{\partial \phi_{i,j}}\left(\frac{\partial \theta_{t-1}}{\partial \phi_{k,l}}\right)'\right] \tag{67}$$

$$+[(W_{k,l}\theta_{t-1}) \otimes X_{t-1}]\text{vec}\left(\frac{\partial \theta_{t-1}}{\partial \phi_{i,j}}\right) + [X_{t-1} \otimes (W_{i,j}\theta_{t-1})]\text{vec}\left[\left(\frac{\partial \theta_{t-1}}{\partial \phi_{k,l}}\right)'\right] + \text{vec}\left(W_{i,j}\theta_{t-1}\theta_{t-1}'W_{k,l}'\right)$$

Secondly, we refer to Eqs. (43) to (45) for Beta-t-QVAR-lev and Beta-t-QVAR-M-lev:

$$\text{vec}\left[\frac{\partial \theta_t}{\partial \psi_{i,j}}\left(\frac{\partial \theta_t}{\partial \psi_{k,l}}\right)'\right] = (X_{t-1}^* \otimes X_{t-1}^*)\text{vec}\left[\frac{\partial \theta_{t-1}}{\partial \psi_{i,j}}\left(\frac{\partial \theta_{t-1}}{\partial \psi_{k,l}}\right)'\right] \tag{68}$$

$$+[(W_{k,l}u_{t-1}) \otimes X_{t-1}^*]\text{vec}\left(\frac{\partial \theta_{t-1}}{\partial \psi_{k,l}}\right) + [X_{t-1}^* \otimes (W_{i,j}u_{t-1})]\text{vec}\left(\frac{\partial \theta_{t-1}}{\partial \psi_{k,l}}\right)' + \text{vec}\left(W_{i,j}u_{t-1}u_{t-1}'W_{k,l}'\right)$$

and

$$\operatorname{vec}\left[\frac{\partial\theta_t}{\partial\phi_{i,j}}\left(\frac{\partial\theta_t}{\partial\phi_{k,l}}\right)'\right]=(X^*_{t-1}\otimes X^*_{t-1})\operatorname{vec}\left[\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}}\left(\frac{\partial\theta_{t-1}}{\partial\phi_{k,l}}\right)'\right] \tag{69}$$

$$+[(W_{k,l}\theta_{t-1})\otimes X^*_{t-1}]\operatorname{vec}\left(\frac{\partial\theta_{t-1}}{\partial\phi_{i,j}}\right)+[X^*_{t-1}\otimes(W_{i,j}\theta_{t-1})]\operatorname{vec}\left[\left(\frac{\partial\theta_{t-1}}{\partial\phi_{k,l}}\right)'\right]+\operatorname{vec}\left(W_{i,j}\theta_{t-1}\theta'_{t-1}W'_{k,l}\right)$$

and

$$\operatorname{vec}\left[\frac{\partial\theta_t}{\partial\psi^*}\left(\frac{\partial\theta_t}{\partial\psi^*}\right)'\right]=(X^*_{t-1}\otimes X^*_{t-1})\operatorname{vec}\left[\frac{\partial\theta_{t-1}}{\partial\psi^*}\left(\frac{\partial\theta_{t-1}}{\partial\psi^*}\right)'\right]+[(W_{2,2}u_{t-1})\otimes X^*_{t-1}]\operatorname{vec}\left(\frac{\partial\theta_{t-1}}{\partial\psi^*}\right) \tag{70}$$

$$+\left\{\begin{bmatrix}0\\ \operatorname{sgn}(-\epsilon_{t-1})\end{bmatrix}\otimes X^*_{t-1}\right\}\operatorname{vec}\left(\frac{\partial\theta_{t-1}}{\partial\psi^*}\right)+(X^*_{t-1}\otimes W_{2,2}u_{t-1})\operatorname{vec}\left[\left(\frac{\partial\theta_{t-1}}{\partial\psi^*}\right)'\right]$$

$$+(W_{2,2}\otimes W_{2,2})\operatorname{vec}(u_{t-1}u'_{t-1})+\left\{\begin{bmatrix}0\\ \operatorname{sgn}(-\epsilon_{t-1})\end{bmatrix}\otimes W_{2,2}\right\}\operatorname{vec}(u_{t-1})$$

$$+\left\{X^*_{t-1}\otimes\begin{bmatrix}0\\ \operatorname{sgn}(-\epsilon_{t-1})\end{bmatrix}\right\}\operatorname{vec}\left[\left(\frac{\partial\theta_{t-1}}{\partial\psi^*}\right)'\right]+\left\{W_{2,2}\otimes\begin{bmatrix}0\\ \operatorname{sgn}(-\epsilon_{t-1})\end{bmatrix}\right\}\operatorname{vec}(u'_{t-1})$$

$$+\operatorname{vec}\left\{\begin{bmatrix}0\\ \operatorname{sgn}(-\epsilon_{t-1})\end{bmatrix}\begin{bmatrix}0\\ \operatorname{sgn}(-\epsilon_{t-1})\end{bmatrix}'\right\}$$

Finally, by using Brandt (1986), Elton (1990), and Alsmeyer (2003), conditions of e.a.s. convergence of $H_t(\hat{\Theta})$ to the unique strictly stationary and ergodic sequence $H_t(\Theta_0)$ for the Beta-t-QVAR(1) models can be obtained. We also refer to Straumann and Mikosch (2006), and Blasques et al. (2017, 2018).

7. Explicit formulas of the dependent variable $\mathbf{y_t}$

(i) Beta-t-QVAR(1):

$$\begin{aligned}y_t&=c+\beta_1\left(\phi_{1,1}\mu^\dagger_{t-1}+\phi_{1,2}\lambda^\dagger_{t-1}+\psi_{1,1}u_{\mu,t-1}+\psi_{1,2}u_{\lambda,t-1}\right)\\&+\exp\left(\omega+\phi_{2,1}\mu^\dagger_{t-1}+\phi_{2,2}\lambda^\dagger_{t-1}+\psi_{2,1}u_{\mu,t-1}+\psi_{2,2}u_{\lambda,t-1}\right)\epsilon_t\end{aligned} \tag{71}$$

(ii) Beta-t-QVAR(1)-M:

$$\begin{aligned} y_t &= c + \beta_1 \left(\phi_{1,1}\mu^{\dagger}_{t-1} + \phi_{1,2}\lambda^{\dagger}_{t-1} + \psi_{1,1}u_{\mu,t-1} + \psi_{1,2}u_{\lambda,t-1}\right) \\ &+\beta_2 \exp\left(\omega + \phi_{2,1}\mu^{\dagger}_{t-1} + \phi_{2,2}\lambda^{\dagger}_{t-1} + \psi_{2,1}u_{\mu,t-1} + \psi_{2,2}u_{\lambda,t-1}\right) \\ &+ \exp\left(\omega + \phi_{2,1}\mu^{\dagger}_{t-1} + \phi_{2,2}\lambda^{\dagger}_{t-1} + \psi_{2,1}u_{\mu,t-1} + \psi_{2,2}u_{\lambda,t-1}\right)\epsilon_t \end{aligned} \tag{72}$$

(iii) Beta-t-QVAR(1)-lev:

$$\begin{aligned} y_t &= c + \beta_1 \left(\phi_{1,1}\mu^{\dagger}_{t-1} + \phi_{1,2}\lambda^{\dagger}_{t-1} + \psi_{1,1}u_{\mu,t-1} + \psi_{1,2}u_{\lambda,t-1}\right) \\ &+ \exp\left[\omega + \phi_{2,1}\mu^{\dagger}_{t-1} + \phi_{2,2}\lambda^{\dagger}_{t-1} + \psi_{2,1}u_{\mu,t-1} + \psi_{2,2}u_{\lambda,t-1} + \psi^{*}\text{sgn}(-\epsilon_{t-1})(u_{\lambda,t}+1)\right]\epsilon_t \end{aligned} \tag{73}$$

(iv) Beta-t-QVAR(1)-M-lev:

$$\begin{aligned} y_t &= c + \beta_1 \left(\phi_{1,1}\mu^{\dagger}_{t-1} + \phi_{1,2}\lambda^{\dagger}_{t-1} + \psi_{1,1}u_{\mu,t-1} + \psi_{1,2}u_{\lambda,t-1}\right) \\ &+\beta_2 \exp\left[\omega + \phi_{2,1}\mu^{\dagger}_{t-1} + \phi_{2,2}\lambda^{\dagger}_{t-1} + \psi_{2,1}u_{\mu,t-1} + \psi_{2,2}u_{\lambda,t-1} + \psi^{*}\text{sgn}(-\epsilon_{t-1})(u_{\lambda,t}+1)\right] \\ &+ \exp\left[\omega + \phi_{2,1}\mu^{\dagger}_{t-1} + \phi_{2,2}\lambda^{\dagger}_{t-1} + \psi_{2,1}u_{\mu,t-1} + \psi_{2,2}u_{\lambda,t-1} + \psi^{*}\text{sgn}(-\epsilon_{t-1})(u_{\lambda,t}+1)\right]\epsilon_t \end{aligned} \tag{74}$$

8. Explicit formulas of volatility σ_t

(i) Beta-t-QVAR(1), and (ii) Beta-t-QVAR(1)-M:

$$\begin{aligned} &\ln\sigma_t^2 = 2\omega + 2\lambda_t^{\dagger} + \tfrac{\nu}{\nu-2} \\ &\lambda_t^{\dagger} = \phi_{2,1}\mu^{\dagger}_{t-1} + \phi_{2,2}\lambda^{\dagger}_{t-1} + \psi_{2,1}u_{\mu,t-1} + \psi_{2,2}u_{\lambda,t-1} \\ &u_{\mu,t} = \left[1 + \tfrac{(y_t-\mu_t)^2}{\nu\exp(2\lambda_t)}\right]^{-1}(y_t-\mu_t) \\ &u_{\lambda,t} = \tfrac{(\nu+1)(y_t-\mu_t)^2}{\nu\exp(2\lambda_t)+(y_t-\mu_t)^2} - 1 \end{aligned} \tag{75}$$

(iii) Beta-t-QVAR(1)-lev, and (iv) Beta-t-QVAR(1)-M-lev:

$$\begin{aligned} &\ln\sigma_t^2 = 2\omega + 2\lambda_t^{\dagger} + \tfrac{\nu}{\nu-2} \\ &\lambda_t^{\dagger} = \phi_{2,1}\mu^{\dagger}_{t-1} + \phi_{2,2}\lambda^{\dagger}_{t-1} + \psi_{2,1}u_{\mu,t-1} + \psi_{2,2}u_{\lambda,t-1} + \psi^{*}\text{sgn}(-\epsilon_{t-1})(u_{\lambda,t}+1) \\ &u_{\mu,t} = \left[1 + \tfrac{(y_t-\mu_t)^2}{\nu\exp(2\lambda_t)}\right]^{-1}(y_t-\mu_t) \\ &u_{\lambda,t} = \tfrac{(\nu+1)(y_t-\mu_t)^2}{\nu\exp(2\lambda_t)+(y_t-\mu_t)^2} - 1 \end{aligned} \tag{76}$$

References

Alsmeyer, G., 2003. On the Harris recurrence of iterated random Lipschitz functions and related convergence rate results. *Journal of Theoretical Probability* 16, 217–247. doi: 10.1023/A:1022290807360.

Blasques, F., Gorgi, P., Koopman, S.J., Wintenberger, O., 2018. Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics* 12, 1019–1052. doi: 10.1214/18-EJS1416.

Blasques, F., Koopman, S.J., Lucas, A., 2017. Maximum likelihood estimation for score-driven models. TI 2014-029/III Tinbergen Institute Discussion Paper. https://papers.tinbergen.nl/14029.pdf.

Brandt, A., 1986. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability* 18, 211–220. doi: 10.2307/1427243.

Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28, 777–795. doi: 10.1002/jae.1279.

Elton, J.H., 1990. A multiplicative ergodic theorem for Lipschitz maps. *Stochastic Processes and their Applications* 34, 39–47. doi: 10.1016/0304-4149(90)90055-W.

Harvey, A.C., 2013. *Dynamic Models for Volatility and Heavy Tails*. Cambridge University Press, Cambridge.

Straumann, D., Mikosch, T. 2006, Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics* 34, 2449–2495. doi: 10.1214/009053606000000803.