Supplemental Appendix to

A Stochastic Recurrence Equations Approach for Score Driven Correlation Models

Francisco Blasques[a], André Lucas[a] and Erkki Silde[a,b]

[(a)] VU University Amsterdam and Tinbergen Institute

[(b)] Duisenberg School of Finance

B Student's t updating recursion driven by i.i.d. noise

We consider the general model $y_t \sim p(y_t), y_t = h(f_t)u_t, h(f_t) \in \mathbb{R}^{n,n}$ and $u_t \sim p(u_t)$ i.i.d., which implies the relationship $p(y_t) = |h(f_t)|^{-1} p(h(f_t)^{-1}u_t)$. We are able to model the time-variation in dependence by parameterizing L_t in terms the dynamic factors $\rho(f_t)$.

The most general distributional form we consider is the multivariate Student's t distribution,the density of which defined by

$$p(\mathbf{y}_t|\nu, \Sigma_t) = \frac{\Gamma\left[(\nu+k)/2\right]}{\Gamma(\nu/2)[(\nu-2)\pi]^{k/2}\left|\mathbf{\Sigma}_t\right|^{1/2}} \left[1 + \frac{1}{\nu-2}\mathbf{y}_t'\mathbf{\Sigma}_t{}^{-1}\mathbf{y}_t\right]^{-(\nu+k)/2}, \tag{B.1}$$

which has the additional closure property of y_t and u_t being in the same class of distributions. This definition of the t-density implies $E[\mathbf{y}_t] = 0$ and $\mathrm{Var}[\mathbf{y}_t] = \Sigma_t$, i.e. the shape parameter ν affects only the tail thickness without having a direct influence on the variance.

We consider a multivariate Student's t density in equation (B.1). Theorem 1 in Creal et al. (2011) gives the following expression for the information matrix,

$$\mathcal{I}_{\rho,t} = \frac{1}{(\nu+4)(1-\rho_t^2)^2}\left((\nu+2)(1+\rho_t^2) - 2\rho_t^2\right), \tag{B.2}$$

and the score

$$\begin{aligned}\nabla_{\rho,t} = \frac{1}{(1-\rho_t^2)^2}\frac{1}{\nu-2+\epsilon_t^2+\eta_t^2}\Big\{\nu\left((1+\rho_t^2)(y_{1t}y_{2t}-\rho_t) - \rho_t(y_{1t}^2+y_{2t}^2-2)\right) + \\ (1+\rho_t^2)\left(2y_{1t}y_{2t} - \rho_t(\epsilon_t^2+\eta_t^2-2)\right) - 2\rho_t\left(y_{1t}^2+y_{1t}^2-(\epsilon_t^2+\eta_t^2-2)\right)\Big\}.\end{aligned} \tag{B.3}$$

Next we write the score entirely in terms of the independent noise $u_t = (\epsilon_t, \eta_t)^\top$ such that $\mathbb{E}[u_t u_t^\top] = \mathrm{I}_2$. However this decomposition is not unique. Two prominent choices are:

1. Cholesky root, obtained by setting $\psi(\rho) = \arcsin(\rho)$ in equation (12) of the paper:

$$h(f) = \begin{pmatrix} 1 & 0 \\ \rho(f) & \sqrt{1-\rho(f)^2}\end{pmatrix},$$

$$\begin{aligned}\nabla_{\rho,t} = \frac{1}{1-\rho_t^2}\frac{1}{\nu-2+\epsilon_t^2+\eta_t^2}\Bigg\{\nu\left[\sqrt{1-\rho_t^2}\epsilon_t\eta_t - \rho_t(\eta_t^2-1)\right] + \\ 2\left[\sqrt{1-\rho_t^2}\epsilon_t\eta_t + \rho_t\left(\frac{1}{2}(\epsilon_t^2-\eta_t^2)-1\right)\right]\Bigg\}.\end{aligned}$$

2. Symmetric root, obtained by setting $\psi(\rho) = 1/2 \arcsin(\rho)$ in equation (12) of the paper:

$$h(f) = \begin{pmatrix} \frac{1}{2}(\sqrt{1+\rho(f)} + \sqrt{1-\rho(f)}) & \frac{1}{2}(\sqrt{1+\rho(f)} - \sqrt{1-\rho(f)}) \\ \frac{1}{2}(\sqrt{1+\rho(f)} - \sqrt{1-\rho(f)}) & \frac{1}{2}(\sqrt{1+\rho(f)} + \sqrt{1-\rho(f)}) \end{pmatrix},$$

$$\nabla_{\rho,t} = \frac{1}{1-\rho_t^2} \frac{1}{\nu - 2 + \epsilon_t^2 + \eta_t^2} \left\{ \nu \left[\epsilon_t \eta_t - \frac{1}{2} \rho_t (\epsilon_t^2 + \eta_t^2 - 2) \right] + 2 \left[\epsilon_t \eta_t - \rho_t \right] \right\}.$$

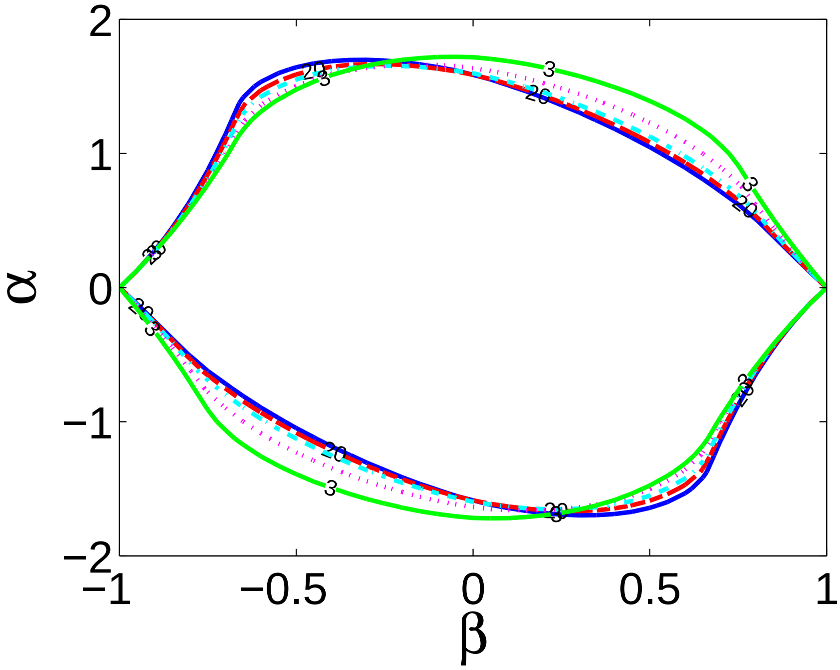
Notice from the above how the limiting case $\nu \to \infty$ reduces to updating corresponding to the normal distribution.

Appropriate scalings and transformations of the above then yield closed-form expressions for the updating equation. Also note that reparameterizing the correlation parameter by the Fisher transformation $\rho_t = \tanh(f_t)$, result in multiplying the score by a factor $(1 - \rho_t^2)$, and thus the information matrix by a factor $(1 - \rho_t^2)^2$ by an application of the chain rule.

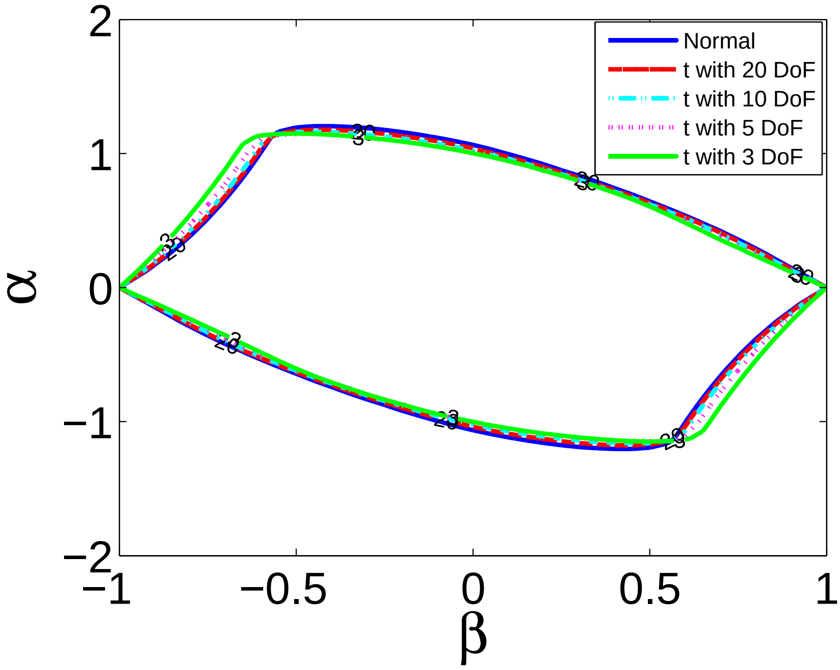
C Further numerical results

Figure C.1 plots the SE regions for the Student's t case with different degrees of freedom. For the symmetric root case (panel (a)), in the relevant first quadrant lower degrees of freedom result in larger regions. The opposite holds for the Cholesky decomposition; see panel (b).

Figure C.2 plots the results for $\psi(\rho) = k_\psi \arcsin(\rho)$. The left panel gives the result for the symmetric matrix root $k_\psi = 1/2$. The right panel is for the Cholesky decomposition, $k_\psi = 1$. Each panel presents 5 different regions. The outer region is based on the numerical evaluation of the original condition (8), with the infimum over ψ replaced by the choice $\psi(\rho) = \arcsin(\rho)/2$. The next region is obtained a numerical evaluation of (8) after applying Jensen's inequality, interchanging the expectations and the log operator. The next region follows after applying the triangle inequality, see the second line of equation (12). The final two regions are obtained after applying the Cauchy-Schwarz, or a second triangle inequality; see equations (12) and (13).

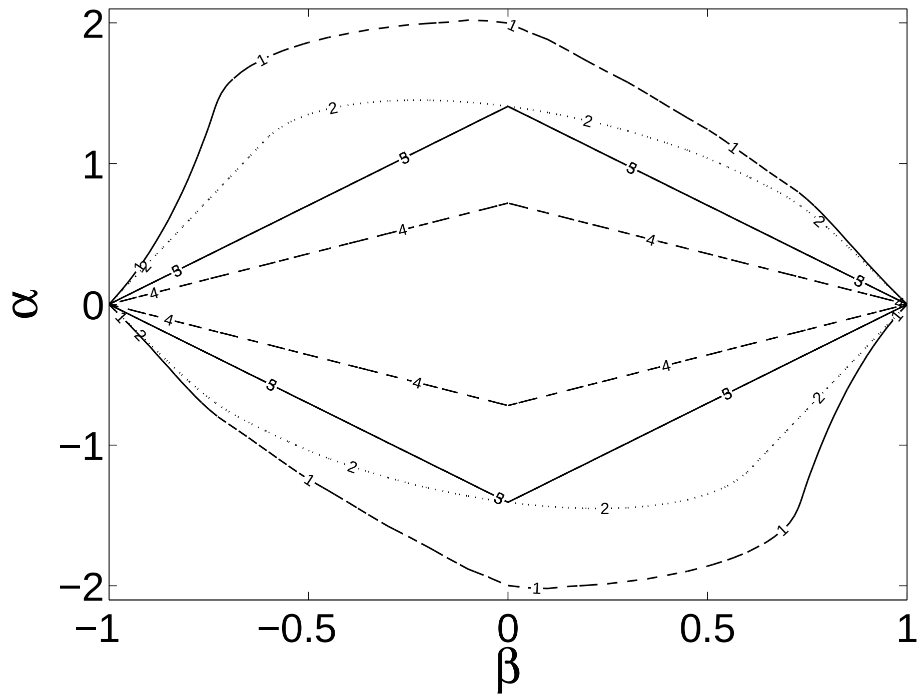


(a) Symmetric root ($k_\psi = 1/2$)

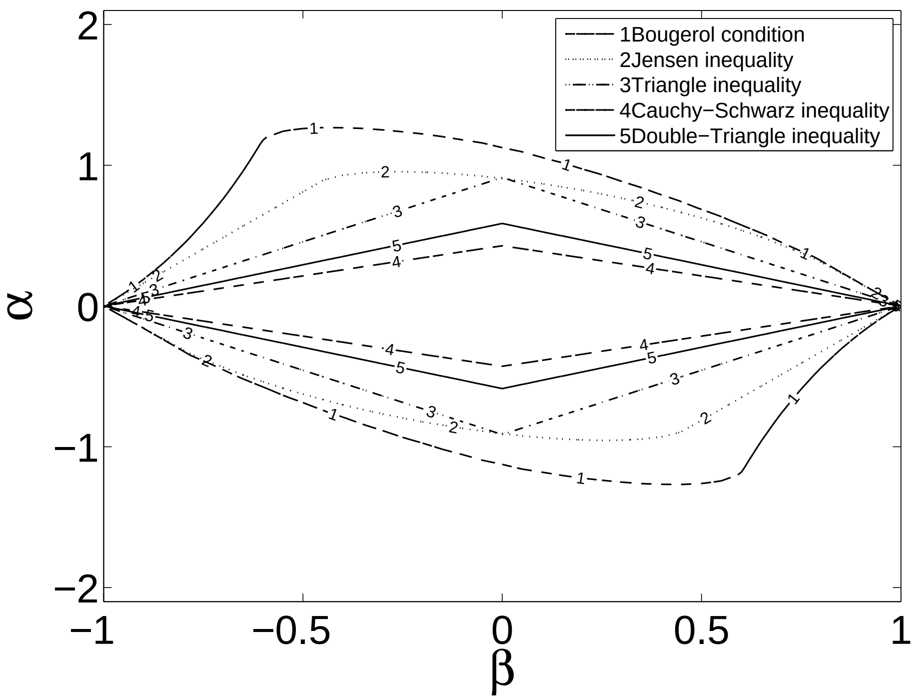


(b) Cholesky root ($k_\psi = 1$)

Figure C.1: Stationarity and ergodicity sufficiency regions for different Student's t degrees of freedom (DoF) and $a = 1/2$



(a) Symmetric root ($k_\psi = 1/2$)



(b) Cholesky root ($k_\psi = 1$)

Figure C.2: Stationarity and ergodicity sufficiency regions for the normal distribution using unit scaling ($S(f) \equiv 1$) and the stricter inequalities in equation (12).

D Primitive conditions for asymptotic properties of MLE

D.1 Bivariate Gaussian model with $\delta \cdot \tanh$ link function and unit scaling

We set $\rho_t = \delta \cdot \tanh(f_t)$, with $0 < \delta < 1$ a fixed, user defined constant. We obtain

$$\frac{\partial \rho_t}{\partial f_t} = \delta \left(1 - \tanh(f_t)^2\right) = (\delta^2 - \rho_t^2)/\delta, \tag{D.1}$$

which reaches its maximum of $\delta < 1$ at $f_t = \rho_t = 0$. We also set $S(f_t) \equiv 1$ and use the normal distribution for the disturbances. The log likelihood function is given by

$$L_T(\theta, f_1) \;=\; \sum_{t=1}^{T} \ell_t(\theta, f_1)$$

with

$$\ell_t(\theta, f_1) = -\log 2\pi - \tfrac{1}{2}\log(1-\rho_t^2) - \frac{1}{2(1-\rho_t^2)}\left(y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t} y_{2,t}\right).$$

D.1.1 Consistency

The high-level assumptions required for the consistency of the ML estimator are:

A1. $\{y_t\}_{t\in\mathbb{Z}}$ is SE and $\{f_t(\theta, f_1)\}_{t\in\mathbb{N}}$ converges e.a.s. to an SE process $\{f_t(\theta)\}_{t\in\mathbb{Z}}$;

A2. $\mathbb{E}\sup_{\theta\in\Theta} |\ell_t(\theta, f_1)| < \infty$;

A3. θ_0 is the unique maximizer of $\mathbb{E}\ell_t(\theta, f_1)$.

The first part of A1 is implied by the conditions of Lemma 1 and an application of Krengel's Ergodic Theorem to the observation equation in (1). The second part of A1 is implied by the conditions of Lemma 3. The moment condition (i) in Lemma 3 is implied by a compact parameter space and

$$\begin{aligned}
&\mathbb{E}\sup_{\theta\in\Theta}\left|\frac{\partial \rho_t}{\partial f_t}\left(\frac{\rho_t}{1-\rho_t^2} - \frac{\rho_t \cdot \left(y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t} y_{2,t}\right)}{(1-\rho_t^2)^2} + \frac{y_{1,t} y_{2,t}}{1-\rho_t^2}\right)\right| \\
&\quad\le \frac{\delta^2}{1-\delta^2} + \mathbb{E}\frac{\delta^2 \cdot \left(y_{1,t}^2 + y_{2,t}^2 + 2\delta |y_{1,t} y_{2,t}|\right)}{(1-\delta^2)^2} + \delta\mathbb{E}\frac{|y_{1,t} y_{2,t}|}{1-\delta^2} < \infty,
\end{aligned}$$

using the score expression from equation (D.2) below, where boundedness is implied by requiring $\{y_t\}_{t\in\mathbb{Z}}$ to be SE and $\mathbb{E}|y_{i,t} y_{j,t}|$ to be finite for $i = 1, 2$ and $j = 1, 2$. The second condition in Lemma 3 restricts the parameter space, in particular α and β.

A2 is implied by selecting requiring $\mathbb{E}|y_{i,t}y_{j,t}| < \infty$ for $i = 1, 2$ and $j = 1, 2$, since

$$\begin{aligned}\mathbb{E}\sup_{\theta\in\Theta}|\ell_t(\theta, f_1)| &\leq \frac{1}{2}\mathbb{E}\sup_{\theta\in\Theta}|\log(1-\rho_t^2)| + \mathbb{E}\sup_{\theta\in\Theta}\Big|\frac{1}{2(1-\rho_t^2)}\big(y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t}\big)\Big| \\ &\leq \frac{1}{2}|\log(1-\delta^2)| + \frac{1}{2(1-\delta^2)}\mathbb{E}\big|y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t}\big|\end{aligned}$$

and $\mathbb{E}|y_{i,t}y_{j,t}| < \infty$ implies

$$\mathbb{E}\big|y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t}\big| \leq \mathbb{E}|y_{1,t}|^2 + \mathbb{E}|y_{2,t}|^2 + 2\delta\mathbb{E}|y_{1,t}y_{2,t}| < \infty.$$

A3 is implied by $\alpha \neq 0$ and $\partial q_t(y_t, f_t)/\partial y \neq 0$ for every f_t and almost every y_t; see the global identification conditions and Theorem 3 in Blasques et al. (2014) for a line of proof that can easily be extended to the case of a bivariate $y_t = (y_{1,t}, y_{2,t})'$.

'Intermediate conditions' for consistency are thus

A1$'$. Conditions of Lemma 1;

A2$'$. $0 < \delta < 1$;

A3$'$. $\mathbb{E}|y_{i,t}y_{j,t}| < \infty$ for $i = 1, 2$ and $j = 1, 2$;

A4$'$. $\alpha \neq 0$;

A5$'$. $\partial q_t(y_t, f_t)/\partial y \neq 0$ for every f_t and almost every y_t.

Finally, we note that in the Gaussian unit-scaling case conditions A3$'$ and A5$'$ hold trivially by inspection of equations (1) and (4). Furthermore, the conditions of Lemma 1 hold easily on a parameter space whose size depends on the choice of δ. Hence, we are left with A2$'$ and A4$'$ for as the sole 'primitive conditions' required for consistency. Both conditions are directly controlled by the researcher:

A2$'$. $0 < \delta < 1$;

A4$'$. $\alpha \neq 0$.

D.1.2 Asymptotic Normality

In this Gaussian setting each element of the score vector is given by

$$\begin{aligned}\nabla_j\ell_t(\theta, f_1) = \frac{\partial\ell_t(\theta, f_1)}{\partial\theta_j} &= \frac{\partial\rho_t}{\partial f_t}\frac{\partial f_t(\theta, f_1)}{\partial\theta_j}\rho_t(1-\rho_t^2)^{-1} \\ &\quad - \frac{\partial\rho_t}{\partial f_t}\frac{\partial f_t(\theta, f_1)}{\partial\theta_j}\rho_t(1-\rho_t^2)^{-2}\big(y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t}\big) \\ &\quad + y_{1,t}y_{2,t}\frac{\partial\rho_t}{\partial f_t}\frac{\partial f_t(\theta, f_1)}{\partial\theta_j}(1-\rho_t^2)^{-1} \qquad \text{(D.2)}\end{aligned}$$

Furthermore, the elements of the log likelihood's second derivative are given by

$$
\begin{aligned}
\nabla_{ij}^2 \ell_t(\theta, f_1) \quad = \quad & \frac{\partial^2 \rho_t}{\partial f_t^2} \frac{\partial f_t(\theta, f_1)}{\partial \theta_i} \frac{\partial f_t(\theta, f_1)}{\partial \theta_j} \rho_t (1-\rho_t^2)^{-1} \\
& + \frac{\partial \rho_t}{\partial f_t} \frac{\partial^2 f_t(\theta, f_1)}{\partial \theta_j \partial \theta_i} \rho_t (1-\rho_t^2)^{-1} \\
& + \left(\frac{\partial \rho_t}{\partial f_t}\right)^2 \frac{\partial f_t(\theta, f_1)}{\partial \theta_i} \frac{\partial f_t(\theta, f_1)}{\partial \theta_j} (1+\rho_t^2)(1+\rho_t^2)^{-2} \\
& - \frac{\partial^2 \rho_t}{\partial f_t^2} \frac{\partial f_t(\theta, f_1)}{\partial \theta_i} \frac{\partial f_t(\theta, f_1)}{\partial \theta_j} \rho_t (1-\rho_t^2)^{-2} \left(y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t} y_{2,t}\right) \\
& - \frac{\partial \rho_t}{\partial f_t} \frac{\partial^2 f_t(\theta, f_1)}{\partial \theta_j \partial \theta_i} \rho_t (1-\rho_t^2)^{-2} \left(y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t} y_{2,t}\right) \\
& - \left(\frac{\partial \rho_t}{\partial f_t}\right)^2 \frac{\partial f_t(\theta, f_1)}{\partial \theta_i} \frac{\partial f_t(\theta, f_1)}{\partial \theta_j} (1+3\rho_t^2)(1-\rho_t^2)^{-3} \left(y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t} y_{2,t}\right) \\
& + 2 \left(\frac{\partial \rho_t}{\partial f_t}\right)^2 \frac{\partial f_t(\theta, f_1)}{\partial \theta_i} \frac{\partial f_t(\theta, f_1)}{\partial \theta_j} \rho_t (1-\rho_t^2)^{-2} y_{1,t} y_{2,t} \\
& + y_{1,t} y_{2,t} \frac{\partial^2 \rho_t}{\partial f_t^2} \frac{\partial f_t(\theta, f_1)}{\partial \theta_i} \frac{\partial f_t(\theta, f_1)}{\partial \theta_j} (1-\rho_t^2)^{-1} \\
& + y_{1,t} y_{2,t} \frac{\partial \rho_t}{\partial f_t} \frac{\partial^2 f_t(\theta, f_1)}{\partial \theta_j \partial \theta_i} (1-\rho_t^2)^{-1} \\
& + 2 y_{1,t} y_{2,t} \left(\frac{\partial \rho_t}{\partial f_t}\right)^2 \frac{\partial f_t(\theta, f_1)}{\partial \theta_i} \frac{\partial f_t(\theta, f_1)}{\partial \theta_j} \rho_t (1-\rho_t^2)^{-2}. \qquad \text{(D.3)}
\end{aligned}
$$

In addition to the consistency conditions, Theorem 2 (Asymptotic Normality) imposes the following high-level assumptions:[5]

AN1. $\mathbb{E}|\dot{\ell}_t(\theta_0, f_1)|^2 < \infty$;

AN2. $\mathbb{E}\sup_{\theta\in\Theta} |\ddot{\ell}_t(\theta, f_1)| < \infty$.

Note that ρ_t, $(1-\rho_t^2)^{-1}$, $\partial\rho_t/\partial f_t$, and $\partial^2\rho_t/\partial f_t^2$ are all uniformly (in f_t) bounded. Therefore, AN1 is implied by $\mathbb{E}\Big|\frac{\partial f_t(\theta_0,f_1)}{\partial \theta_h}\Big|^2 < \infty$, $h = 1, \ldots, n_\theta$ (all elements of θ) and $\mathbb{E}|\frac{\partial f_t(\theta_0,f_1)}{\partial \theta_h} y_{i,t} y_{j,t}|^2 < \infty$ since by (Blasques et al., 2014a, Lemma SA.2) , there exists a constant $c > 0$, such that

$$
\begin{aligned}
\mathbb{E}|\nabla_j \ell_t(\theta_0, f_1)|^2 \leq\; & \delta^2(1-\delta^2)^{-1} c\, \mathbb{E}\Big|\frac{\partial f_t(\theta_0, f_1)}{\partial \theta_j}\Big|^2 \\
& + \delta^2(1-\delta^2)^{-2} c\, \mathbb{E}\Big|\frac{\partial f_t(\theta_0, f_1)}{\partial \theta_j} \left(y_{1,t}^2 + y_{2,t}^2 + 2\delta|y_{1,t} y_{2,t}|\right)\Big|^2 \\
& + \delta(1-\delta^2)^{-1} c\, \mathbb{E}\Big|y_{1,t} y_{2,t} \frac{\partial f_t(\theta_0, f_1)}{\partial \theta_j}\Big|^2 < \infty,
\end{aligned}
$$

[5]This presupposes the SE nature of $\partial f_t(\theta, f_1)/\partial\theta$ and $\partial^2 f_t(\theta, f_1)/\partial\theta\partial\theta'$. The conditions for this are implied by the conditions for the existence of moments of these same quantities, which we establish at the end of the proof.

where the boundedness of the second term is obtained by a second application of (Blasques et al., 2014a, Lemma SA.2)

$$\mathbb{E}\Big|\frac{\partial f_t(\theta_0,f_1)}{\partial \theta_j}\big(y_{1,t}^2+y_{2,t}^2+2\delta|y_{1,t}y_{2,t}|\big)\Big|^2 \le c\,\mathbb{E}\Big|\frac{\partial f_t(\theta_0,f_1)}{\partial \theta_j}y_{1,t}^2\Big|^2+ c\,\mathbb{E}\Big|\frac{\partial f_t(\theta_0,f_1)}{\partial \theta_j}y_{2,t}^2\Big|^2+2c\,\mathbb{E}\Big|\frac{\partial f_t(\theta_0,f_1)}{\partial \theta_j}y_{1,t}y_{2,t}\Big|^2<\infty$$

AN2 is implied by the following additional conditions

$$\mathbb{E}\sup_{\theta\in\Theta}\Big|\frac{\partial f_t(\theta,f_1)}{\partial\theta_h}\Big|^2<\infty, \qquad \mathbb{E}\sup_{\theta\in\Theta}\Big|\frac{\partial f_t(\theta,f_1)}{\partial\theta_h}y_{i,t}y_{j,t}\Big|^2<\infty,$$

$$\mathbb{E}\sup_{\theta\in\Theta}\Big|\frac{\partial^2 f_t(\theta,f_1)}{\partial\theta_h\partial\theta_{h'}}\Big|<\infty, \qquad \mathbb{E}\sup_{\theta\in\Theta}\Big|\frac{\partial^2 f_t(\theta,f_1)}{\partial\theta_h\partial\theta_{h'}}y_{i,t}y_{j,t}\Big|<\infty.$$

This follows directly by inspecting the expression for the second derivative of the log likelihood contribution in (D.3) and an application of norm sub-additivity inequalities.

Summing up, the additional 'intermediate conditions' used for establishing asymptotic normality are

AN1′ $\mathbb{E}\Big|\frac{\partial f_t(\theta_0,f_1)}{\partial\theta_h}\Big|^2<\infty$;

AN2′ $\mathbb{E}|\frac{\partial f_t(\theta_0,f_1)}{\partial\theta_h}y_{i,t}y_{j,t}|^2<\infty$;

AN3′ $\mathbb{E}\sup_{\theta\in\Theta}\Big|\frac{\partial f_t(\theta,f_1)}{\partial\theta_h}\Big|^2<\infty$;

AN4′ $\mathbb{E}\sup_{\theta\in\Theta}|\frac{\partial f_t(\theta,f_1)}{\partial\theta_h}y_{i,t}y_{j,t}|^2<\infty$;

AN5′ $\mathbb{E}\sup_{\theta\in\Theta}\Big|\frac{\partial^2 f_t(\theta,f_1)}{\partial\theta_h\partial\theta_{h'}}\Big|<\infty$;

AN6′ $\mathbb{E}\sup_{\theta\in\Theta}|\frac{\partial^2 f_t(\theta,f_1)}{\partial\theta_h\partial\theta_{h'}}y_{i,t}y_{j,t}|<\infty$.

Clearly, AN1′ and AN2′ are implied by AN3′ and AN4′. Furthermore, since y_t has moments of arbitrary order due to the normality assumption, then by application of a generalized Holder's inequality we can substitute AN4′ and AN6′ by $\mathbb{E}\sup_{\theta\in\Theta}|\frac{\partial f_t(\theta,f_1)}{\partial\theta_h}|^{2+d}<\infty$ and $\mathbb{E}\sup_{\theta\in\Theta}|\frac{\partial^2 f_t(\theta,f_1)}{\partial\theta_h\partial\theta_{h'}}|^{1+d}<\infty$ for some $d>0$. As a result, the set of 'intermediate conditions' for asymptotic normality can be reduced to

AN1″ $\mathbb{E}\sup_{\theta\in\Theta}|\frac{\partial f_t(\theta,f_1)}{\partial\theta_h}|^{2+d}<\infty$ for some $d>0$;

AN2″ $\mathbb{E}\sup_{\theta\in\Theta}|\frac{\partial^2 f_t(\theta,f_1)}{\partial\theta_h\partial\theta_{h'}}|^{1+d}<\infty$ for some $d>0$.

Finally, we obtain the desired moments for the derivative processes by application of Proposition SA.2 from the supplemental appendix to Blasques et al. (2014a). These conditions also imply the SE nature of these derivative processes. Before we formulate the conditions, we spell out the derivatives in terms of their primitives. We have

$$\frac{\partial f_{t+1}}{\partial \theta_j} = \frac{\partial \omega}{\partial \theta_j} + \frac{\partial \alpha}{\partial \theta_j} q(y_t, f_t) + \frac{\partial \beta}{\partial \theta_j} f_t + \left(\beta + \alpha \frac{\partial q(y_t, f_t)}{\partial f_t}\right) \frac{\partial f_t(\theta, f_1)}{\partial \theta_j},$$

where $q(y_t, f_t)$ is defined as the score with respect to the time-varying parameter (the right-hand side in (D.2), but with the factors $\partial f_t(\theta, f_1)/\partial \theta_j$ dropped). Similarly, we have for the second derivatives

$$\begin{aligned}\frac{\partial^2 f_{t+1}}{\partial \theta_i \partial \theta_j} = & \frac{\partial^2 \omega}{\partial \theta_i \partial \theta_j} + \frac{\partial^2 \alpha}{\partial \theta_i \partial \theta_j} q(y_t, f_t) + \frac{\partial^2 \beta}{\partial \theta_i \partial \theta_j} f_t + \frac{\partial \alpha}{\partial \theta_j} \frac{\partial q(y_t, f_t)}{\partial f_t} \frac{\partial f_t(\theta, f_1)}{\partial \theta_i} + \frac{\partial \beta}{\partial \theta_j} \frac{\partial f_t(\theta, f_1)}{\partial \theta_i} + \\ & \left(\frac{\partial \beta}{\partial \theta_i} + \frac{\partial \alpha}{\partial \theta_i} \frac{\partial q(y_t, f_t)}{\partial f_t} + \alpha \frac{\partial^2 q(y_t, f_t)}{\partial f_t^2} \frac{\partial f_t(\theta, f_1)}{\partial \theta_i}\right) \frac{\partial f_t(\theta, f_1)}{\partial \theta_j} + \\ & \left(\beta + \alpha \frac{\partial q(y_t, f_t)}{\partial f_t}\right) \frac{\partial^2 f_t(\theta, f_1)}{\partial \theta_i \partial \theta_j}.\end{aligned}$$

Following Proposition SA.2 from the supplemental appendix to Blasques et al. (2014a) and assuming fixed initial conditions $\partial f_1(\theta, f_1)/\partial \theta$ and $\partial^2 f_1(\theta, f_1)/\partial \theta \partial \theta'$, the conditions become

AN1a$''$ $\mathbb{E} \sup_{(\theta,f) \in \Theta \times \mathcal{F}} \left|\beta + \alpha \frac{\partial q(y_t, f)}{\partial f_t}\right|^{2+d} < 1$ for some $d > 0$;

AN1b$''$ $\mathbb{E} \sup_{\theta \in \Theta} \left|\frac{\partial \omega}{\partial \theta_j} + \frac{\partial \alpha}{\partial \theta_j} q(y_t, f_1) + \frac{\partial \beta}{\partial \theta_j} f_1\right|^{2+d} < \infty$ for some $d > 0$;

AN2a$''$ $\mathbb{E} \sup_{(\theta,f) \in \Theta \times \mathcal{F}} \left|\beta + \alpha \frac{\partial q(y_t, f)}{\partial f_t}\right|^{1+d} < 1$ for some $d > 0$;

AN2b$''$ $\mathbb{E} \sup_{\theta \in \Theta} \Bigg| \frac{\partial^2 \omega}{\partial \theta_i \partial \theta_j} + \frac{\partial^2 \alpha}{\partial \theta_i \partial \theta_j} q(y_t, f_1) + \frac{\partial^2 \beta}{\partial \theta_i \partial \theta_j} f_1 + \frac{\partial \alpha}{\partial \theta_j} \frac{\partial q(y_t, f_1)}{\partial f_t} \frac{\partial f_1(\theta, f_1)}{\partial \theta_i} + \frac{\partial \beta}{\partial \theta_j} \frac{\partial f_1(\theta, f_1)}{\partial \theta_i} +$

$$\left(\frac{\partial \beta}{\partial \theta_i} + \frac{\partial \alpha}{\partial \theta_i} \frac{\partial q(y_t, f_1)}{\partial f_t} + \alpha \frac{\partial^2 q(y_t, f_1)}{\partial f_t^2} \frac{\partial f_1(\theta, f_1)}{\partial \theta_i}\right) \frac{\partial f_1(\theta, f_1)}{\partial \theta_j} \Bigg|^{1+d} < \infty$$

for some fixed initial condition $\partial f_1(\theta, f_1)/\partial \theta$ and $d > 0$.

By (Blasques et al., 2014a, Lemma SA.2) we obtain via similar lines as before that AN1b$''$ is satisfied if $\mathbb{E}|y_{i,t} y_{j,t}|^{2+d} < \infty$, which trivially holds due to the conditional Gaussianity of $y_{i,t}$. The expressions for $\partial q(y_t, f_1)/\partial f_t$ and $\partial^2 q(y_t, f_1)/\partial f_t^2$ are highly cumbersome. Looking at the expression in equation (D.3), however, we can see that both these derivatives are polynomials in ρ_t, $\partial^j \rho_t / \partial f_t^j$ for $j = 1, 2, 3$, $(1 - \rho_t^2)^{-1}$, and $(y_{i,t} y_{j,t})$. Given the fact that all of these have bounded moments under the assumption of conditional normality and $0 < \delta < 1$, AN2b$''$ is automatically satisfied by the previous sets of conditions.

Since AN1a$''$ implies AN2a$''$, we are finally left with AN1a$''$ as the sole additional condition needed to obtain asymptotic normality. Building on the consistency conditions, we thus obtain the asymptotic normality of the MLE from:

A1$'$. Conditions of Lemma 1;

A2′. $0 < \delta < 1$;

A4′. $\alpha \neq 0$;

AN1a″ $\mathbb{E}\sup_{(\theta,f)\in\Theta\times\mathcal{F}} \left|\beta + \alpha\frac{\partial q(y_t,f)}{\partial f_t}\right|^{2+d} < 1$ for some $d > 0$;

Finally, we note that Condition AN1a″ restricts the parameter space in a very similar way as do the contraction conditions in Lemma 1, and we already showed how the size of these regions depend on $0 < \delta < 1$. Hence, we are left as before with two conditions that are directly controlled by the researcher:

A2′. $0 < \delta < 1$;

A4′. $\alpha \neq 0$.

D.2 Bivariate Student's t model with $\delta \cdot \tanh$ link function and unit scaling

As previously, we set $\rho_t = \delta \cdot \tanh(f_t)$, with $0 < \delta < 1$ a fixed, user defined constant. We also set $S(f_t) \equiv 1$ and again have the result from equation (D.1). For the Student's t model with $\nu < \infty$ degrees of freedom, we follow the same lines as for the Gaussian case (i.e., the case $\nu \to \infty$) in Appendix D.1. The likelihood contribution at time t for the Student's t model is given by

$$\begin{aligned}\ell_t(\theta, f_1) =& \log\left(\tfrac{1}{2}\nu\right) - \log(2\pi\,(\nu-2)) - \tfrac{1}{2}\log(1-\rho_t^2) \\ & - \tfrac{1}{2}(\nu+2)\log\left(1 + \frac{1}{(\nu-2)(1-\rho_t^2)}(y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t})\right),\end{aligned} \tag{D.4}$$

with $2 < \underline{\nu} \leq \nu \leq \overline{\nu} < \infty$. The score is given by

$$q(y_t, f_t) = \frac{\rho_t}{1-\rho_t^2}\frac{\partial \rho_t}{\partial f_t} - (\nu+2)\frac{\rho_t\frac{y_{1,t}^2+y_{2,t}^2-2\rho_t y_{1,t}y_{2,t}}{(\nu-2)(1-\rho_t^2)^2} - \frac{y_{1,t}y_{2,t}}{(\nu-2)(1-\rho_t^2)}}{1+\frac{y_{1,t}^2+y_{2,t}^2-2\rho_t y_{1,t}y_{2,t}}{(\nu-2)(1-\rho_t^2)}}\frac{\partial \rho_t}{\partial f_t}. \tag{D.5}$$

Note that we have parameterized the correlation matrix in this case by assuming $\nu > 2$, such that variances exist. Alternatively, we could interpret the scaling matrix in the data generating process as a pure scaling matrix and drop the requirement that second moments exist ($\nu > 2$). The arguments below would continue to apply to that case as well, with the number of required moments of the data dropping to the existence of an arbitrarily small moment. To see this, it is important to note that for $0 < \delta < 1$ we have $0 \leq \rho_t^2 \leq \delta^2 < 1$

such that the correlation matrix is positive definite and therefore

$$\sup_{y_i,y_j\in\mathbb{R}} \frac{|y_i y_j|}{1+(\nu-2)^{-1}(1-\rho_t^2)^{-1}(y_1^2+y_2^2-2\rho_t y_1 y_2)} \leq c < \infty, \tag{D.6}$$

for some constant c.

D.2.1 Consistency

The high-level assumptions required for the consistency of the ML estimator are:

B1. $\{y_t\}_{t\in\mathbb{Z}}$ is SE and $\{f_t(\theta, f_1)\}_{t\in\mathbb{N}}$ converges e.a.s. to an SE process $\{f_t(\theta)\}_{t\in\mathbb{Z}}$;

B2. $\mathbb{E}\sup_{\theta\in\Theta}|\ell_t(\theta, f_1)| < \infty$;

B3. θ_0 is the unique maximizer of $\mathbb{E}\ell_t(\theta, f_1)$.

The first part of B1 is implied by the conditions of Lemma 1 and an application of Krengel's theorem to the observation equation (1), and possible non-degenerate regions were presented in the paper. For the second part of B1, we require a compact parameter space and

B1a′. $\mathbb{E}\sup_{\theta\in\Theta}\log\sup_{f\in\mathcal{F}}\left|\beta+\alpha\frac{\partial q(y_t,f)}{\partial f_t}\right| < 0$;

B1b′. $\mathbb{E}\sup_{\theta\in\Theta}|q(y_t, f_1)| < \infty$.

Given the score expression in (D.5) and assuming $0 < \delta < 1$ such that (D.6) applies, we have

$$\begin{aligned}\mathbb{E}\sup_{\theta\in\Theta}|q(y_t, f_1)| <& \frac{\delta^2}{1-\delta^2} + \mathbb{E}\sup_{\theta\in\Theta}\delta\,(\nu+2)\left(\frac{\rho_t}{\left(1-\rho_t^2\right)^2} + \frac{|y_{1,t}y_{t,2}|}{(\nu-2)(1-\rho_t^2)+y_{1,t}^2+y_{2,t}^2-2\rho_t y_{1,t}y_{2,t}}\right)\\ \leq& \frac{\delta^2}{1-\delta^2} + \delta\,(\overline{\nu}+2)\left(\frac{\delta}{\left(1-\delta^2\right)^2}+c\right) < \infty,\end{aligned}$$

for some constant c. To obtain the second term in the first line, we have used the fact that $x/(1+x) \leq 1$ for $x \geq 0$.

B2 follows by first noting that $\rho_t^2 \leq \delta^2$, such that

$$\begin{aligned}\mathbb{E}\sup_{\theta\in\Theta}|\ell_t(\boldsymbol{\theta}, f_1)| \leq& \sup_{\theta\in\Theta}\left(\left|\log\tfrac{1}{2}\nu\right| + |\log(2\pi\,(\nu-2))| + \tfrac{1}{2}\left|\log(1-\delta^2)\right|\right)\\ &+ \mathbb{E}\sup_{\theta\in\Theta}\frac{\nu+2}{2}\log\left(1+\frac{y_{1,t}^2+y_{2,t}^2-2\rho_t y_{1,t}y_{2,t}}{(\nu-2)(1-\delta^2)}\right).\end{aligned} \tag{D.7}$$

The assumptions $0 < \delta < 1$ and $\mathbb{E}|y_{i,t}y_{j,t}|^{\epsilon} < \infty$ for some $\epsilon > 0$ then imply $\mathbb{E}\sup_{\theta\in\Theta}|\ell_t(\boldsymbol{\theta}, f_1)| < \infty$. An $\epsilon = 2$ moment of $y_{i,t}y_{j,t}$ is implied by the existence of second moments in the data generating process (1).[6]

B3 is satisfied using precisely the same arguments as for the Gaussian case.

Consistency is thus ensured by the following primitive conditions:

B1′. Conditions of Lemma 1 and Lemma 3;

B2′. $0 < \delta < 1$;

B3′. $\alpha \neq 0$.

B4′. $2 < \underline{\nu} \leq \nu \leq \overline{\nu} < \infty$.

D.2.2 Asymptotic Normality

For the Student's t setting, each element of the score vector is given by

$$\begin{aligned}\nabla_j \ell_t(\theta, f_1) =& \frac{\partial \rho_t}{\partial f_t}\frac{\partial f_t(\theta, f_1)}{\partial \theta_j}\rho_t(1-\rho_t^2)^{-1} + \left(\frac{1}{\nu} - \frac{1}{\nu-2}\right)\frac{\partial \nu}{\partial \theta_j} \\ &- \frac{1}{2}\frac{\partial \nu}{\partial \theta_j}\log\left(1 + \frac{y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t}}{(\nu-2)(1-\rho_t^2)}\right) \\ &+ \frac{\nu+2}{2(\nu-2)}\frac{\partial \nu}{\partial \theta_j}\frac{y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t}}{(\nu-2)(1-\rho_t^2) + y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t}} \\ &- (\nu+2)\frac{\partial \rho_t}{\partial f_t}\frac{\partial f_t(\theta, f_1)}{\partial \theta_j}\frac{\rho_t(1-\rho_t^2)^{-1}(y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t})}{(\nu-2)(1-\rho_t^2) + y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t}} \\ &+ (\nu+2)\frac{\partial \rho_t}{\partial f_t}\frac{\partial f_t(\theta, f_1)}{\partial \theta_j}\frac{y_{1,t}y_{2,t}}{(\nu-2)(1-\rho_t^2) + y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t}}.\end{aligned} \tag{D.8}$$

Again, we have to satisfy the high-level conditions

BN1. $\mathbb{E}|\dot{\ell}_t(\theta_0, f_1)|^2 < \infty$;

BN2. $\mathbb{E}\sup_{\theta\in\Theta}|\ddot{\ell}_t(\theta, f_1)| < \infty$.

Using equation (D.8), (Blasques et al., 2014a, Lemma SA.2) and the $\delta \cdot \tanh$ parameterization with $0 < \delta < 1$, we can substitute BN1 by the following conditions for all elements of the derivative processes:

BN1a′. $\mathbb{E}\left|\frac{\partial f_t(\theta_0, f_1)}{\partial \theta_j}\right|^2 < \infty$;

BN1b′. $\mathbb{E}\left|\log\left(1 + \frac{y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t}}{(\nu-2)(1-\rho_t^2)}\right)\right|^2 < \infty$;

BN1c′. $\mathbb{E}\left|\frac{y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t}}{(\nu-2)(1-\rho_t^2) + y_{1,t}^2 + y_{2,t}^2 - 2\rho_t y_{1,t}y_{2,t}}\right|^2 < \infty$;

[6] For the case of a pure scaling matrix, the requirement follows due to the existence of a moment $0 < \epsilon < \nu$ under the Student's t assumption.

BN1d′. $\mathbb{E}\left|\frac{\partial f_t(\theta_0,f_1)}{\partial\theta_j}\frac{(y_{1,t}^2+y_{2,t}^2-2\rho_t y_{1,t}y_{2,t})}{(\nu-2)(1-\rho_t^2)+y_{1,t}^2+y_{2,t}^2-2\rho_t y_{1,t}y_{2,t}}\right|^2<\infty$;

BN1e′. $\mathbb{E}\left|\frac{\partial f_t(\theta_0,f_1)}{\partial\theta_j}\frac{y_{1,t}y_{2,t}}{(\nu-2)(1-\rho_t^2)+y_{1,t}^2+y_{2,t}^2-2\rho_t y_{1,t}y_{2,t}}\right|^2<\infty$;

BN1b′ is directly implied by the existence of second moments of y_t. BN1a′ implies BN1c′, BN1d′, and BN1e′ due to equation (D.6). Note that conditions for the existence of second moments of the derivative process will also imply the e.a.s. convergence to an SE process of the first and second order derivative processes using Proposition SA.2 from Blasques et al. (2014a).

Rather than spelling out the second derivative of the likelihood in detail, we draw attention to the expression in equation (D.8). From this and using (D.6), it becomes clear that BN2 is satisfied if

BN1f′. $\mathbb{E}\sup_{\theta\in\Theta}\left|\frac{\partial f_t(\theta,f_1)}{\partial\theta_j}\right|^2<\infty$;

BN1g′. $\mathbb{E}\sup_{\theta\in\Theta}\left|\frac{\partial^2 f_t(\theta,f_1)}{\partial\theta_i\partial\theta_j}\right|^2<\infty$.

BN1f′ implies BN1a′.

Similar as for the Gaussian case, we can use the results for general stochastic recurrence equations in Proposition SA.2 by Blasques et al. (2014a). The expressions for the derivatives of $f_t(\theta,f_0)$ are similar to the ones for the Gaussian case, but with additional terms to account for the parameter ν, in particular terms involving $\partial q(y_t,f_t)/\partial\nu$ and $\partial^2 q(y_t,f_t)/\partial\nu^2$, for example

$$\frac{\partial f_{t+1}}{\partial\theta_j}=\frac{\partial\omega}{\partial\theta_j}+\frac{\partial\alpha}{\partial\theta_j}q(y_t,f_t)+\alpha\frac{\partial q(y_t,f_t)}{\partial\nu}\frac{\partial\nu}{\partial\theta_j}+\frac{\partial\beta}{\partial\theta_j}f_t+\left(\beta+\alpha\frac{\partial q(y_t,f_t)}{\partial f_t}\right)\frac{\partial f_t(\theta,f_1)}{\partial\theta_j}.$$

Using equation (D.6) once more, the uniform boundedness ensures that BN1a′–BN1g′ as well as BN2 are satisfied if

BN1″ $\mathbb{E}\sup_{(\theta,f)\in\Theta\times\mathcal{F}}\left|\beta+\alpha\frac{\partial q(y_t,f_t)}{\partial f_t}\right|^2<1$.

BN1″ restricts the parameter space, in particular the permissible values of α and β, in a similar way as the conditions in Lemma 1 and Lemma 3 in the paper and result in similar regions as the one shown in the paper.

E Complementary results to the empirical application

E.1 FTSE: Diagnostic checks on the univariate models in the GJR family

First, we use the GJR family

$$\sigma^2(t) = c + g_1\sigma^2(t-1) + \ldots + g_P\sigma^2(t-P) + a_1 y(t-1)^2 + \ldots + a_Q y(t-Q)^2$$
$$+ L_1\,1[y(t-1) < 0]y(t-1)^2 + \ldots + L_Q\,1[y(t-Q) < 0]y(t-q)^2$$

The GJR(1,2) specification is preferred as it gives the log-likelihood of -5121.31 as opposed to -5127 of the GJR(1,1) model and -5129 of the EGARCH(1,2) model. Adding more lags is neither significant, nor do the diagnostics change.

Table E.1: GJR(1,2) Conditional Variance Model for FTSE 100 Index

Parameter	Value	Standard Error	t- Statistic
Constant	0.0372671	0.008906	4.18449
GARCH1	0.874143	0.0129657	67.4199
ARCH1	0.00225888	0.0132958	0.169894
ARCH2	0.0755335	0.0188271	4.01196
Leverage1	0.0816841	0.0189078	4.32013
DoF	7.18691	0.956046	7.51732

Table E.2: Results of Ljung-Box Test for Remaining ARCH effects (H0:none)

lags	Result Hypothesis	pValue	Statistic	Crit Val
10	0	0.7749	6.4643	18.3070
15	0	0.8144	10.0840	24.9958
20	0	0.8900	12.6963	31.4104
25	0	0.8866	16.8668	37.6525
30	0	0.7492	24.4944	43.7730

Table E.3: Results of Engle's Test for Remaining ARCH effects (H0:none)

lags	Result Hypothesis	pValue	Statistic	Crit Val
10	0	0.7826	6.3780	18.3070
15	0	0.8018	10.2799	24.9958
20	0	0.8865	12.7830	31.4104
25	0	0.8919	16.7129	37.6525
30	0	0.7266	24.9676	43.7730

E.2 Athex Composite Index: Diagnostic checks on the univariate models in the EGARCH family

Here the EGARCH family

$$\begin{aligned}\log[\sigma^2(t)] = c + g_1 \log[\sigma^2(t-1)] + \ldots + g_P \log[\sigma^2(t-P)] \\ + a_1(|y(t-1)| - E|y(t-1)|) + a_Q(|y(t-Q)| - E|y(t-Q)|) \\ + L_1 y(t-1) + \ldots + L_Q y(t-Q)\end{aligned}$$

is chosen. EGARCH(1,2) is preferred as it gives a log-likelihood of -4041.02 as opposed to -4046.13 of the EGARCH(1,1) model and -4052.14 of the GJR(1,2) model. Adding more lags is neither signficant, nor do the diagnostics change.

Table E.4: EGARCH(1,2) Conditional Variance Model for Athex Composite Index

Parameter	Value	Standard Error	t- Statistic
Constant	-0.000106474	0.00194011	-0.0548802
GARCH1	0.98243	0.00240827	407.941
ARCH1	-0.0192319	0.0413664	-0.464917
ARCH2	0.143411	0.0429705	3.33744
Leverage1	-0.130415	0.010867	-12.0011
DoF	12.8521	2.86693	4.48289

Table E.5: Results of Ljung-Box Test for Remaining ARCH effects (H0:none)

lags	Result Hypothesis	pValue	Statistic	Crit Val
10	0	0.8610	5.4257	18.3070
15	0	0.5087	14.2230	24.9958
20	0	0.5506	18.5588	31.4104
25	1	0.0261	40.4590	37.6525
30	1	0.0399	44.8484	43.7730

Table E.6: Results of Engle's Test for Remaining ARCH effects (H0:none)

lags	Result Hypothesis	pValue	Statistic	Crit Val
10	0	0.8653	5.3680	18.3070
15	0	0.5097	14.2092	24.9958
20	0	0.5547	18.4975	31.4104
25	1	0.0390	38.7546	37.6525
30	0	0.0546	43.3479	43.7730

E.3 Robustness check: Modeling both marginals by an EGARCH(1,2) model

The results in this section show that reasonable leave our results practically unaltered. We model both series by an EGARCH(1,2). We omit the robustness check of modeling both series by as for the Athex Composite Index, as the GJR binding positivity result in an inferior model fit (a likehood decrease of more than 10 points).

Table E.7: Both EGARCH(1,2) marginals. Full Estimation Results.

	EWMA1	EWMA2	EWMA3	t(∞)GAS (a=0)	t(∞)GAS (a=0.5)	t(∞)GAS (a=1)	t(5)GAS (a=0)	t(5)GAS (a=0.5)	t(5)GAS (a=1)	t($\cdot$)GAS (a=1)
λ	5	5	8.6648	∞	∞	∞	5	5	5	8.9598
			0.00742							0.8788
c		0.0002	0.0010	0.0115	0.0112	0.0109	0.0089	0.0089	0.0088	0.0112
		0.0003	0.0003	0.0448	0.0451	0.0358	0.0111	0.0113	0.0067	0.0071
A				0.0254	0.0281	0.0310	0.0334	0.0315	0.0296	0.0356
				0.0481	0.0472	0.0416	0.0172	0.0144	0.0089	0.0100
B	0.9771	0.9769	0.9761	0.9757	0.9763	0.9770	0.9798	0.9799	0.9801	0.9766
	0.0069	0.0048	0.0019	0.1098	0.0914	0.0723	0.0237	0.0240	0.0142	0.0141
Log-likelihood In-sample	−7845	−7844	−7812	−7886	−7886	−7886	−7838	−7838	−7838	−7803
AIC	15691	15692	15629	15779	15779	15778	15682	15682	15682	15615
BIC	15697	15704	15647	15797	15797	15796	15700	15700	15700	15639
# estimated parameters	1	2	3	3	3	3	3	3	3	4
Reject Zero Const Corr at 0.05 level	Yes	Yes	No	No	No	No	No	No	No	No
Reject Const Corr at 0.05 level	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No

Notes: The second line underneath each parameter estimate depicts HAC sandwich standard errors. For volatility, EGARCH(1,2) marginals are used for both series.

Table E.8: Location of parameter estimates with respect to the SE contraction region

	t(∞)GAS(1,1) (a=0)	t(∞)GAS(1,1) (a=0.5)	t(∞)GAS(1,1) (a=1)	t(5)GAS(1,1) (a=0)	t(5)GAS(1,1) (a=0.5)	t(5)GAS(1,1) (a=1)	t(9)GAS(1,1) (a=1)
Inside SE region?							
$k_\psi = 1$ (Cholesky)	Yes	Yes	No	No	No	No	No
$k_\psi = 1/2$ (Symmetric)	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: For volatility, EGARCH(1,2) marginals are used for both series.