Supplementary Appendix - Modelling the Conditional Distribution of Financial Returns with Asymmetric Tails.

Stephen Thiele

Queensland University of Technology

June 26, 2019

1 Sufficient Conditions for Invertibility of the AST DCS model

The condition in Assumption 1.(c) is implied by $E|\phi + [\kappa - \kappa^*(1 - 2I_{(y_t \leq 0)})][(I_{(y_t \leq \mu)})\frac{\partial s_t^l}{\partial \lambda_t} + (1 - I_{(y_t \leq \mu)})\frac{\partial s_t^r}{\partial \lambda_t}]| < 1$. Given the bounds $-C_{l/r} = -\frac{(\nu_{l/r}+3)(\nu_{l/r}+1)}{4\nu_{l/r}} \leq \frac{\partial s_t^{l/r}}{\partial \lambda_t} \leq 0$, and the fact that the condition is trivially satisfied at the upper bound, it is sufficient to satisfy the following $E|\phi + (\kappa^* - \kappa)(C_l I_{(y_t \leq \mu)} + C_r I_{(y_t > \mu)}) - 2\kappa^* C_r I_{(y_t > 0)}|$. The parameter space in then partitioned into two parts with $\mu \leq 0$ and $\mu > 0$, leading to the following conditions.

Case 1: $\mu > 0$

$$|\phi + (\kappa^* - \kappa)C_l|(\alpha - M_1) + |\phi + (\kappa^* - \kappa)C_l - 2\kappa^* C_r|M_1$$

$$+ |\phi + (\kappa^* - \kappa)C_r - 2\kappa^* C_r|(1 - \alpha) < 1$$

Case 2: $\mu \leq 0$

$$|\phi + (\kappa^* - \kappa)C_l|\alpha + |\phi + (\kappa^* - \kappa)C_r|M_2 + |\phi + (\kappa^* - \kappa)C_r - 2\kappa^* C_r|(1 - \alpha - M_2) < 1$$

where M_1 and M_2 are the probabilities that $0 < y_t \leq \mu$ and $\mu < y_t \leq 0$ respectively. These can be obtained directly from the CDF of the AST distribution given in Section 3 of Zhu and Galbraith (2010).

2 Proofs for the Univariate AST DCS model

2.1 AST DCS - Definitions and Notation

Theoretically, it will be convenient to work with the infinitely lived process

$$\lambda_{t+1} = \frac{\delta}{1-\phi} + \kappa \sum_{j=0}^{\infty} \phi^j s_{t-j}$$

In practice we are forced to use a finite initial value, denoted $\tilde{\lambda}_0$, which leads to the process $\{\tilde{\lambda}_t\}_{t\in\mathbb{N}}$. It will be shown below that the use of this initial value doesn't influence the asymptotic properties of the estimator. Throughout I will write the log likelihood as

$$L(y|\theta) = T^{-1} \sum_{t=1}^{T} ln\ f(y_t|\theta) = T^{-1} \sum_{t=1}^{T} l(y_t|\theta)$$
$$l(y_t|\theta) = I_{(y_t \leq \mu)} l^l(y_t|\theta) + (1 - I_{(y_t \leq \mu)}) l^r(y_t|\theta)$$

where $I_{(y_t \leq \mu)}$ is an indicator function that equals 1 if $y_t \leq \mu$ and 0 otherwise. Though this is based on the conditional density, dependence on the information set $\mathcal{F}_{t-1}$ will be suppressed throughout. Likewise, the score in Section 2.2 will be condensed down to $s_t = I_{g(y,\theta)} s_t^l + (1 - I_{g(y,\theta)}) s_t^r$. Values of the likelihood and score that are functions of the finite process $\{\tilde{\lambda}_t\}$ will also be denoted with a tilde (*i.e.*- $\tilde{s}_t$ and $\tilde{L}(y|\theta)$). Let θ_i refer to the i^{th} element of $\theta = \{\mu, \nu_l, \nu_r, \alpha, \delta, \phi, \kappa, \kappa^*\}$. The subscript θ_0 refers to the true parameter value while $\lambda_t(\theta_0) = \lambda_{0t}$. Use is made throughout of constants denoted C_i, for $i = 1, 2, 3, ...$ It should be noted that these do not represent the same constants across different results.

Weak derivatives are used throughout the following results as the functions at hand are differentiable almost everywhere. The functions in question generally have the form $g(y, \theta) = I_{(y_t \leq \mu)} g^l(y, \theta) + (1 - I_{(y_t \leq \mu)}) g^r(y, \theta)$. I define the weak derivative as

$$\frac{\partial g(y,\theta)}{\partial \theta} = I_{(y_t \leq \mu)} \frac{\partial g^l(y,\theta)}{\partial \theta} + (1 - I_{(y_t \leq \mu)}) \frac{\partial g^r(y,\theta)}{\partial \theta}$$

These derivatives are used to establish Lipschitz continuity as given in Davidson (1994, Theorem 21.10)

$$\left|g(y,\theta') - g(y,\theta)\right| \leq B\left|\theta' - \theta\right|, \quad \text{a.s.}$$

where the Lipschitz constant B can take the form $\left|\frac{\partial g(y,\theta)}{\partial \theta}\right|$. As this result need only hold almost surely, the lack of derivatives on sets of measure zero do not affect the results. To simplify many expressions, I denote $sgn(-y_t) \equiv h_t$. Finally, $|A| = (tr(A'A))^{1/2}$ denotes the Euclidean norm of a vector or matrix.

2.2 Preliminary Results - Consistency

The first lemma uses Lemma 7a) from Zhu (2012) and then follows similar arguments to those in equation (61) of the same paper.

Lemma 2.1. *For some positive constant* $C_1 > 1$, $E|ln(1 + C_1(y_t - \mu)^2)| < \infty$ *under Assumptions 1.(a) and 1.(b).*

Proof. Lemma 7a) from Zhu (2012) states that for some $\epsilon > 0$, there exists a positive finite constant M_0 such that $|ln(x)| \leq M_0(1 + x^\epsilon + x^{-\epsilon})$ for $x > 0$.[1] This can be used to bound the desired quantity as follows.

$$\begin{aligned}
E|ln(1 + C_1(y_t - \mu)^2)| &\leq E|ln(1 + 2C_1\mu^2 + 2C_1y_t^2)| \\
&= E|ln(1 + 2C_1\mu^2 + 2C_1y_t^2)I_{(y_t^2 \leq 1)} + ln(1 + 2C_1\mu^2 + 2C_1y_t^2)I_{(y_t^2 > 1)}| \\
&\leq 1 + 2C_1(1 + \bar{\mu}^2) + E|ln(1 + 2C_1\mu^2 + 2C_1y_t^2)I_{(y_t^2 > 1)}| \\
&\leq 1 + 2C_1(1 + \bar{\mu}^2) + (1 + 2C_1\bar{\mu}^2) + ln(2C_1) + 2E|ln(y_t^2)| \\
&\leq C_2 + 2M_0E(1 + y_t^{2\epsilon} + y_t^{-2\epsilon}) < \infty
\end{aligned}$$

where the last inequality holds for $\epsilon < \underline{\nu}/2$. □

[1] This follows from $x^\epsilon|ln(x)| \to 0$ as $x \to 0^+$ and $x^{-\epsilon}|ln(x)| \to 0$ as $x \to +\infty$.

Lemma 2.2. *Under Assumption 1.(a)-(c), $\{\lambda_{0t}\}$ is strictly stationary and ergodic, for all $\theta_0 \in \Theta$. This implies that ϵ_t and the estimated e_t are also strictly stationary and ergodic. Furthermore, any measurable function of ϵ_t will be measurable.*

Proof. First note that at θ_0, s_{0t} is a bounded *i.i.d.* sequence that is not a function of λ_{0t}. For the basic model with $\kappa_0^* = 0$, the conditions for stationarity and ergodicity of λ_{0t} as given in Brandt (1986) amount to

$$Elog|\phi| < 0 \quad Elog^+|s_0| < \infty$$

The first is met trivially given the parameter restriction in Assumption 1.(b) and the second follows directly from the fact that the score is naturally bounded. Hence, $\{\lambda_{0t}\}$ is strictly stationary and ergodic. The model with leverage (*i.e.* - $\kappa_0^* \neq 0$) no longer fits the linear stochastic recurrence equation of Brandt (1986), nor does it satisfy the Lipschitz conditions required in Bougerol (1993). In this case, the existence of a stationary and ergodic solution can be established by defining the following measurable maps for finite m

$$\lambda_{0t,m} = \begin{cases} \varrho_0 & m = 0 \\ \delta_0 + \phi_0 \lambda_{t-1,m} + \kappa_0 s(z_{t-1}) + \kappa_0^* sgn(g(z_{t-1}, \lambda_{0t-1,m}))s(z_{t-1}) & m \geq 1 \end{cases}$$

where ϱ_0 is an initializing constant. Iterating backwards yields

$$\lambda_{t,m} = \delta_0 \sum_{i=0}^{m} \phi_0^i + \sum_{i=0}^{m} \phi_0^i \left[(\kappa_0 + \kappa_0^* sgn(g(z_{t-1}, .., z_{t-1-m}, \varrho_0)))\, s(z_{t-1-i}) \right] + \phi_0^i \varrho_0$$

As $|\phi_0| < 1$, $0 \leq \kappa, \kappa^* \leq \bar{\kappa}$ and $|s(z_{t-1})| \leq C < \infty$ by Assumption 1.(b), this sequence converges almost surely on a closed segment of $\mathbb{R}$ as $m \to \infty$. By Proposition 2.6 of Straumann and Mikosch (2006), the process $\{\lambda_{0t}\}$ is stationary and ergodic and λ_{0t} is a measurable function of $(z_{t-1}, z_{t-2}, ...)$. It follows that ϵ_t is also stationary and ergodic by Assumption 1.(a). The estimated residuals $e_t = \epsilon_t + \mu_0 - \mu$ are a measurable function of ϵ_t and so strictly stationary and ergodic (Theorem 3.5.8 of Stout(1974)). □

Lemma 2.3. *Given Assumption 1.(b). regarding the compactness of the parameter space* Θ*, the time-varying scale is bounded by* $0 < \underline{\sigma} \leq \sigma_t \leq \bar{\sigma} < \infty$.

Proof. Note that the score is bounded by $s_l = -\frac{\nu+3}{2\underline{\nu}} \leq s_t \leq \frac{\bar{\nu}+3}{2} = s_u$. I further define $s_m = max(s_l, s_u)$. I then get

$$
\begin{aligned}
|\lambda_t| &= \left| \sum_{j=0}^{\infty} \phi^j (\delta + \kappa s_{t-1-j} + \kappa^* h_{t-1-j} s_{t-1-j}) \right| \\
&\leq \sum_{j=0}^{\infty} |\phi^j (\delta + \kappa s_{t-1-j} + \kappa^* h_{t-1-j} s_{t-1-j})| \\
&\leq \sum_{j=0}^{\infty} |\phi^j \delta| + \sum_{j=0}^{\infty} \phi^j \kappa |s_{t-1-j}| + \sum_{j=0}^{\infty} \phi^j \kappa^* |s_{t-1-j}| \leq \frac{|\bar{\delta}|}{1-|\bar{\phi}|} + \frac{2|\bar{\kappa} s_m|}{1-|\bar{\phi}|} < \infty
\end{aligned}
$$

relying on the restriction $|\phi| < \bar{\phi} < 1$. The result directly follows. □

Lemma 2.4. *Under Assumption 1.(a)-(c),* $E\left[\sup_{\theta\in\Theta} |\frac{\partial \lambda_t}{\partial \theta}|\right] < \infty \quad a.s.$.

Proof. For some x_{it} we have

$$
\begin{aligned}
\left|\frac{\partial \lambda_t}{\partial \theta_i}\right| &= \left| x_{i,t} + \left(\phi + (\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \lambda_{t-1}}\right) \frac{\partial \lambda_{t-1}}{\partial \theta_i} \right| \\
&= \left| x_{i,t} + \sum_{j=1}^{\infty} \prod_{l=1}^{j} \left(\phi + (\kappa + \kappa^* h_{t-l}) \frac{\partial s_{t-l}}{\partial \lambda_{t-l}}\right) x_{i,t-j} \right| \\
&\leq |x_{i,t}| + \sum_{j=1}^{\infty} \prod_{l=1}^{j} \left|\phi + (\kappa + \kappa^* h_{t-l}) \frac{\partial s_{t-l}}{\partial \lambda_{t-l}}\right| |x_{i,t-j}|
\end{aligned}
$$

By Assumption 1.(c), $sup_{\theta\in\Theta} \prod_{l=0}^{j} \left|\phi + (\kappa + \kappa^* h_{t-l}) \frac{\partial s_{t-l}}{\partial \lambda_{t-l}}\right| \rightarrow 0$ a.s. at an exponential rate as $j \rightarrow \infty$ following Lemma 2.4 of Straumann and Mikosch (2006). Given this, proof of the stated result requires $\sup_{\theta\in\Theta} E\left[log^+ x_{i,t}\right] < \infty$ as per Lemma 2.1 of the same paper.

Defining $a_t = \left(\frac{|y_t - \mu|}{2\sqrt{\nu_l}\alpha^* \sigma_t}\right)$ and $b_t = \left(\frac{|y_t - \mu|}{2\sqrt{\nu_r}(1-\alpha^*)\sigma_t}\right)$ and again using weak derivatives where appropriate, it is straightforward to show that under Assumption 1.(b)

$$
\begin{aligned}
|x_{1,t}| &= \left|\kappa \frac{\partial s_{t-1}(y_{t-1}|\theta,\sigma_{t-1})}{\partial \mu}\right| \leq \bar{\kappa}\left|\frac{(\nu_l+3)(\nu_l+1)}{\nu_l}\frac{a_{t-1}}{\left(1+a_{t-1}^2\right)^2}\right| \\
&\quad + \bar{\kappa}\left|\frac{(\nu_r+3)(\nu_r+1)}{\nu_r}\frac{b_{t-1}}{\left(1+b_{t-1}^2\right)^2}\right| \leq 2\bar{\kappa}\frac{(\bar{\nu}+3)(\bar{\nu}+1)}{\underline{\nu}} < \infty \\
|x_{2,t}| &= \left|\kappa \frac{\partial s_{t-1}(y_{t-1}|\theta,\sigma_{t-1})}{\partial \nu_l}\right| \leq \bar{\kappa}\left|\frac{(\nu_r+3)(\nu_r+1)}{\nu_r}\frac{b_{t-1}^2}{\left(1+b_{t-1}^2\right)^2}\frac{\partial \alpha^*}{\partial \nu_l}\right| \\
&\quad + \bar{\kappa}\left|\frac{3}{2\nu_l^2}s_{t-1}^l + \frac{\nu_l+3}{2\nu_l}\left[\frac{a_{t-1}^2}{1+a_{t-1}^2} - (\nu_l+1)\frac{a_{t-1}^2}{\left(1+a_{t-1}^2\right)^2}\left(\alpha^{*2}2\nu_l\frac{\partial \alpha^*}{\partial \nu_l}\right)\right]\right| \\
&\leq \bar{\kappa}\left(\frac{3}{2\underline{\nu}^2} + \frac{\bar{\nu}+3}{2\bar{\nu}}\right)\bar{\kappa}\frac{(\bar{\nu}+3)(\bar{\nu}+1)^2}{\underline{\nu}}\left|\frac{\partial \alpha^*}{\partial \nu_l}\right| \\
&\leq C_1 + C_2\left|\frac{\partial \alpha^*}{\partial \nu_l}\right| \leq C_1 + C_2 C_3 < \infty \\
|x_{3,t}| &= \left|\kappa \frac{\partial s_{t-1}(y_{t-1}|\theta,\sigma_{t-1})}{\partial \nu_r}\right| \leq C_1 + C_2 C_3 < \infty \\
|x_{4,t}| &= \left|\kappa \frac{\partial s_{t-1}(y_{t-1}|\theta,\sigma_{t-1})}{\partial \alpha}\right| \leq \bar{\kappa}\left|\frac{(\nu_l+3)(\nu_l+1)}{\nu_l}\frac{a_{t-1}^2}{\left(1+a_{t-1}^2\right)^2}\frac{1}{\alpha^{*2}}\frac{\partial \alpha^*}{\partial \alpha}\right| \\
&\quad + \bar{\kappa}\left|\frac{(\nu_r+3)(\nu_r+1)}{\nu_r}\frac{b_{t-1}^2}{\left(1+b_{t-1}^2\right)^2}\frac{1}{(1-\alpha^*)^2}\frac{\partial \alpha^*}{\partial \alpha}\right| \\
&\leq 2\frac{(\bar{\nu}+3)(\bar{\nu}+1)^2}{\underline{\nu}}\left(1+\frac{1-\bar{\alpha}}{\underline{\alpha}}\frac{K(\bar{\nu}))}{K(\underline{\nu}))}\right)^2\frac{4}{\bar{\alpha}(1-\bar{\alpha})} < \infty \\
|x_{5,t}| &= |1| < \infty \\
|x_{6,t}| &= |\lambda_{t-1}| \leq \frac{|\bar{\delta}|}{1-|\bar{\phi}|} + \frac{|\bar{\kappa} s_m|}{1-|\bar{\phi}|} < \infty \\
|x_{7,t}| &= |s_{t-1}| \leq s_m < \infty \\
|x_{8,t}| &= |h_{t-1}s_{t-1}| \leq s_m < \infty
\end{aligned}
$$

which is sufficient to establish the stated result. □

Lemma 2.5. *Under Assumption 1.(a)-(c),* $E\left[\underset{\theta\in\Theta}{sup}\left|\frac{\partial l(y_t|\theta)}{\partial \theta}\right|\right] < \infty \quad a.s..$

Proof. I begin with

$$\left|\frac{\partial l(y_t|\theta)}{\partial \phi}\right| = \left|\left[I_{f(y,\theta)}\frac{\partial l^l(y_t|\theta)}{\partial \lambda_t} + (1 - I_{f(y,\theta)})\frac{\partial l^r(y_t|\theta)}{\partial \lambda_t}\right]\frac{\partial \lambda_t}{\partial \phi}\right|$$
$$\leq 2\,|s_m|\left|\frac{\partial \lambda_t}{\partial \phi}\right| < \infty$$

which follows from the fact that $|s_m|$ is bounded. Application of Lemma 2.4 then provides the result for ϕ. The proofs for the derivatives w.r.t δ, κ and κ^* are essentially the same as they only appear in the likelihood through λ_t, and so they are omitted. Turning to other parameters, it is useful to first bound the expectation of the partial derivatives (holding σ_t fixed). This can be done using the compact parameter space (Assumption 1.(b)) and the results in Lemma 2.1. I demonstrate this with one side of the distribution

$$\left|\frac{\partial l^l(y_t|\theta,\sigma_t)}{\partial \mu}\right| = \left|\frac{\nu_l+1}{2\alpha^*\sigma_t}\frac{a_t}{1+a_t^2}\right| \leq \frac{\nu_l+1}{2\alpha^*\sigma_t} \leq \frac{\bar{\nu}+1}{\underline{\sigma}}\left(1+\frac{1-\bar{\alpha}}{\underline{\alpha}}\frac{K(\bar{\nu})}{K(\underline{\nu})}\right) < \infty$$
$$\left|\frac{\partial l^l(y_t|\theta,\sigma_t)}{\partial \nu_l}\right| = \left|\frac{1}{M(\alpha,\nu_l,\nu_r)}\frac{\partial M(\alpha,\nu_l,\nu_r)}{\partial \nu_l} - \frac{1}{2}ln(1+a_t^2) + \frac{\nu_l+1}{2\nu_l}\frac{a_t^2}{1+a_t^2}\right|$$
$$\leq C_1 + C_2 + \frac{\bar{\nu}+1}{\underline{\nu}} < \infty$$
$$\left|\frac{\partial l^l(y_t|\theta,\sigma_t)}{\partial \nu_r}\right| = \left|\frac{1}{M(\alpha,\nu_l,\nu_r)}\frac{\partial M(\alpha,\nu_l,\nu_r)}{\partial \nu_r}\right| \leq C_1 < \infty$$
$$\left|\frac{\partial l^l(y_t|\theta,\sigma_t)}{\partial \alpha}\right| = \left|\frac{1}{M(\alpha,\nu_l,\nu_r)}\frac{\partial M(\alpha,\nu_l,\nu_r)}{\partial \alpha} + \frac{\nu_l+1}{\alpha^*}\frac{a_t^2}{1+a_t^2}\frac{\partial \alpha^*}{\partial \alpha}\right|$$
$$\leq C_1 + (\bar{\nu}+1)\left(1+\frac{1-\bar{\alpha}}{\underline{\alpha}}\frac{K(\bar{\nu})}{K(\underline{\nu})}\right)\frac{4}{\bar{\alpha}(1-\bar{\alpha})} < \infty$$

where the terms for ν_l, ν_r and α use the fact that $M()$ is a continuously differentiable function of bounded parameters. Proceeding in the case of ν_l I get

$$E\left|\frac{\partial l(y_t|\theta)}{\partial \nu_l}\right| = E\left|\frac{\partial l(y_t|\theta,\sigma_t)}{\partial \nu_l} + \left[I_{f(y,\theta)}\frac{\partial l^l(y_t|\theta)}{\partial \lambda_t} + (1-I_{f(y,\theta)})\frac{\partial l^r(y_t|\theta)}{\partial \lambda_t}\right]\frac{\partial \lambda_t}{\partial \nu_l}\right|$$
$$\leq C_2 + 2\,|s_m|\,E\left|\frac{\partial \lambda_t}{\partial \nu_l}\right| < \infty$$

Similar arguments are used to bound the expectations of the derivatives with respect to ν_r, α and μ which established the stated result. □

Lemma 2.6. *Under Assumption 1.(a)-(c),* $E\left[\sup_{\theta\in\Theta}|l(y_t|\theta)|\right] < \infty$.

Proof. This can be shown by dealing with each half of the density in turn.

$$\begin{aligned}
|l^l(y_t|\theta)| &= \left|-\lambda_t + ln(M(\alpha,\nu_l,\nu_r)) - \frac{\nu_l+1}{2} ln\left(1+\frac{1}{\nu_l}\left(\frac{y_t-\mu}{2\alpha^*\sigma_t}\right)^2\right)\right| \\
&\leq |\lambda_t| + |ln(M(\alpha,\nu_l,\nu_r))| + \left|\frac{\nu_l+1}{2}\right|\left|ln\left(1+\frac{1}{\nu_l}\left(\frac{y_t-\mu}{2\alpha^*\sigma_t}\right)^2\right)\right| \\
&\leq C_1 + C_2 + C_3 ln(1+C_4(y_t-\mu)^2) < \infty
\end{aligned}$$

where the final inequality comes from Lemma 2.1 and the bounded constance $\{C_1, C_2, C_3, C_4\}$ exist due to Lemma 2.3 and the compactness of Θ (Assumption 1.(b)) as previously demonstrated. Similar arguments apply to $l^r(y_t|\theta)$. Finally, noting that $|I_{g(y,\theta)}l^l(y_t|\theta) + (1 - I_{g(y,\theta)})l^r(y_t|\theta)| \leq |l^l(y_t|\theta)| + |l^r(y_t|\theta)|$, the stated result is obtained. □

The initialisation can be ignored in the asymptotic arguments based on the following lemma

Lemma 2.7. *Under Assumption 1.(a)-(c), for all finite initial values* $\tilde{\lambda}_0$, $\sup_{\theta\in\Theta}|\tilde{L}(y,\theta,\tilde{\lambda}_0) - L(y,\theta,\lambda_0)| \xrightarrow{p} 0$ *as* $T \to \infty$.

Proof. The first part of the proof is similar to that of Lemma 2.4 above. Note that

$$\begin{aligned}
\frac{\partial l(y_t|\theta,\tilde{\lambda}_t)}{\partial\tilde{\lambda}_0} = s_t\frac{\partial\tilde{\lambda}_t}{\partial\tilde{\lambda}_0} &= s_t\left[\left(\phi + (\kappa+\kappa^* h_{t-1})\frac{\partial\tilde{s}_{t-1}}{\partial\tilde{\lambda}_{t-1}}\right)\frac{\partial\tilde{\lambda}_{t-1}}{\partial\tilde{\lambda}_0}\right] \\
&\leq s_m\prod_{j=0}^{t-1}\left(\phi + (\kappa+\kappa^* h_{t-j})\frac{\partial\tilde{s}_{t-1-j}}{\partial\tilde{\lambda}_{t-1-j}}\right) \xrightarrow{a.s.} 0 \quad \text{as } t\to\infty
\end{aligned}$$

for all $\theta\in\Theta$ by Lemma 2.1 of Straumann and Mikosch(2006). This relies on

$$\begin{aligned}
&\sup_{\theta\in\Theta} E\left[log\left|\phi + (\kappa+\kappa^* h_{t-1})\frac{\partial\tilde{s}_{t-1}}{\partial\tilde{\lambda}_{t-1}}\right|\right] < 0 \\
&\Longrightarrow \left(\phi + (\kappa+\kappa^* h_{t-j})\frac{\partial\tilde{s}_{t-1-j}}{\partial\tilde{\lambda}_{t-1-j}}\right)^j \xrightarrow{a.s.} 0 \quad \text{as } j\to\infty
\end{aligned}$$

which is guaranteed by Assumption 1.(c). It follows that

$$T^{-1}\sum_{t=1}^{T}\frac{\partial l(y_t|\theta,\tilde{\lambda}_t)}{\partial\tilde{\lambda}_0}\overset{a.s.}{\to}0 \quad \text{as } T\to\infty$$

This gives $\frac{\partial\tilde{L}(y|\theta)}{\partial\tilde{\lambda}_0}\overset{a.s.}{\to}0$ as $t\to\infty$ for all $\theta\in\Theta$. It follows that

$$\left|\tilde{L}(y|\theta)-L(y|\theta)\right|\leq\left|\frac{\partial\tilde{L}(y|\theta)}{\partial\tilde{\lambda}_0}\right|\left|\tilde{\lambda}_0-\lambda_0\right|\overset{a.s.}{\to}0 \quad \text{as } T\to\infty$$

□

2.3 Proofs for AST DCS consistency

Proof of Theorem 1a) (Consistency). Consistency is established by verifying the conditions of Theorem 2.1 in Newey and McFadden (1994). Lemma 2.7 enables us to ignore the initialisation and simply work with the process with an infinite past. Condition *ii*) (compact parameter space) is satisfied by Assumption 1.(b). The continuity *iii*) and uniform convergence of the objective function *iv*) can be shown using the uniform LLN of Andrews (1992, Theorem 3). This ULLN requires the Lipschitz condition given in Lemma 2.5 and pointwise convergence of the objective function. The latter is shown with the strong law of large numbers for stationary and ergodic processes (Stout (1974), Theorem 3.5.7) using Lemma 2.2 and Lemma 2.6.

The identification in condition *i*) can be established in two parts. Firstly, the results on identification of the static model in Proposition 1 of Zhu and Galbraith (2010) carry over to the present case. It is therefore only necessary to show that the time-varying scale $\sigma_t(\theta)$ is unique in the parameters θ. This can be proven by contradiction along similar lines to Weiss (1986) and Lumsdaine (1996). I start with

$$ln\left(\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}\right)=0 \implies (\theta-\theta_0)'\frac{\partial\lambda_t}{\partial\theta}\bigg|_{\theta=\theta^*}=0$$

where the derivative is taken over all y_t where it exists[2]. Assuming that $(\theta - \theta_0) = \gamma \neq 0$, I obtain the following

$$\gamma_1 \frac{\partial \lambda_t}{\partial \mu} = \gamma_1 \left[(\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \mu} + \left(\phi + (\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \lambda_{t-1}} \right) \frac{\partial \lambda_{t-1}}{\partial \mu} \right]$$
$$\implies \gamma_1 (\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \mu} = 0 \quad \text{a.s}$$
$$\gamma_2 \frac{\partial \lambda_t}{\partial \nu_l} = \gamma_2 \left[(\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \nu_l} + \left(\phi + (\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \lambda_{t-1}} \right) \frac{\partial \lambda_{t-1}}{\partial \nu_l} \right]$$
$$\implies \gamma_2 (\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \nu_l} = 0 \quad \text{a.s}$$
$$\gamma_3 \frac{\partial \lambda_t}{\partial \nu_r} = \gamma_3 \left[(\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \nu_r} + \left(\phi + (\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \lambda_{t-1}} \right) \frac{\partial \lambda_{t-1}}{\partial \nu_r} \right]$$
$$\implies \gamma_3 (\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \nu_r} = 0 \quad \text{a.s}$$
$$\gamma_4 \frac{\partial \lambda_t}{\partial \alpha} = \gamma_4 \left[(\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \alpha} + \left(\phi + (\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \lambda_{t-1}} \right) \frac{\partial \lambda_{t-1}}{\partial \alpha} \right]$$
$$\implies \gamma_4 (\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \alpha} = 0 \quad \text{a.s}$$
$$\gamma_5 \frac{\partial \lambda_t}{\partial \delta} = \gamma_5 \left[1 + \left(\phi + (\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \lambda_{t-1}} \right) \frac{\partial \lambda_{t-1}}{\partial \delta} \right] \implies \gamma_5 = 0$$
$$\gamma_6 \frac{\partial \lambda_t}{\partial \phi} = \gamma_6 \left[\lambda_{t-1} + \left(\phi + (\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \lambda_{t-1}} \right) \frac{\partial \lambda_{t-1}}{\partial \phi} \right] \implies \gamma_6 \lambda_{t-1} = 0 \quad \text{a.s}$$
$$\gamma_7 \frac{\partial \lambda_t}{\partial \kappa} = \gamma_7 \left[s_{t-1} + \left(\phi + (\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \lambda_{t-1}} \right) \frac{\partial \lambda_{t-1}}{\partial \kappa} \right] \implies \gamma_7 s_{t-1} = 0 \quad \text{a.s}$$
$$\gamma_8 \frac{\partial \lambda_t}{\partial \kappa^*} = \gamma_8 \left[h_{t-1} s_{t-1} + \left(\phi + (\kappa + \kappa^* h_{t-1}) \frac{\partial s_{t-1}}{\partial \lambda_{t-1}} \right) \frac{\partial \lambda_{t-1}}{\partial \kappa} \right] \implies \gamma_8 h_{t-1} s_{t-1} = 0 \quad \text{a.s}$$

However, the random variables that are the final terms of each equation are clearly non-degenerate, implying that $\frac{\partial \lambda_t}{\partial \theta}\gamma = 0 \iff \gamma = 0$ which leads to a contradiction. Therefore, $ln(\sigma_t(\theta_0)/\sigma_t(\theta)) = 0$ iff $\theta = \theta_0$.

As all conditions of Theorem 2.1 in Newey and McFadden (1994) are satisfied, $\hat{\theta} \xrightarrow{a.s.} \theta_0$ as $T \to \infty$. □

[2] The set over which it does not has probability zero. Hence the condition is met "almost everywhere", which is enough to establish identification.

2.4 Preliminary Results - Normality

Lemma 2.8. *Under Assumption 1.(a)-(c),* $\left|\frac{\partial^2 s_t}{\partial \lambda_t^2}\right| \leq 2\left|\frac{\partial s_t}{\partial \lambda_t}\right| \leq \frac{2(\nu_m+3)(\nu_m+1)}{\nu_m} < \infty$.

Proof. Starting with

$$s_t^l = C^l \left[\frac{a_t^2}{1+a_t^2} - \frac{1}{(\nu_l+1)}\right]$$

which gives

$$\begin{aligned}
\left|\frac{\partial s_t^l}{\partial \lambda_t}\right| &= \left|2C^l \frac{a_t^2}{\left(1+a_t^2\right)^2}\right| \leq 2C^l = \frac{(\nu_l+3)(\nu_l+1)}{\nu_l} \\
\left|\frac{\partial^2 s_t^l}{\partial \lambda_t^2}\right| &= \left|4C^l \left[\frac{a_t^2}{\left(1+a_t^2\right)^2}\left(1-\frac{2}{1+a_t^2}\right)\right]\right| \\
&\leq 4C^l = \frac{2(\nu_l+3)(\nu_l+1)}{\nu_l}
\end{aligned}$$

Similar arguments apply to s_t^r which implies that

$$\begin{aligned}
\left|\frac{\partial s_t}{\partial \lambda_t}\right| &\leq max\left(\frac{(\nu_l+3)(\nu_l+1)}{\nu_l}, \frac{(\nu_r+3)(\nu_r+1)}{\nu_r}\right) = \frac{(\bar{\nu}+3)(\bar{\nu}+1)}{\underline{\nu}} < \infty \\
\left|\frac{\partial^2 s_t^l}{\partial \lambda_t^2}\right| &\leq max\left(\frac{2(\nu_l+3)(\nu_l+1)}{4\nu_l}, \frac{2(\nu_r+3)(\nu_r+1)}{4\nu_r}\right) = \frac{2(\bar{\nu}+3)(\bar{\nu}+1)}{\underline{\nu}} < \infty
\end{aligned}$$

□

Lemma 2.9. *Under Assumption 1.(a)-(c),* $E\left[\sup_{\theta\in\Theta}\left|\frac{\partial \lambda_t}{\partial \theta_i}\frac{\partial \lambda_t}{\partial \theta_j}\right|\right] < \infty$.

Proof. Squaring the initial expression from Lemma 2.4 gives

$$\begin{aligned}\left|\frac{\partial\lambda_t}{\partial\theta_i}\frac{\partial\lambda_t}{\partial\theta_j}\right| &= \left|x_{i,t}x_{j,t} + x_{i,t}\left(\phi + (\kappa+\kappa^* h_{t-1})\frac{\partial s_{t-1}}{\partial\lambda_{t-1}}\right)\frac{\partial\lambda_{t-1}}{\partial\theta_j}\right. \\ &\quad \left. + x_{i,t}\left(\phi + (\kappa+\kappa^* h_{t-1})\frac{\partial s_{t-1}}{\partial\lambda_{t-1}}\right)\frac{\partial\lambda_{t-1}}{\partial\theta_j}\right| \\ &\quad + \left|\left(\phi + (\kappa+\kappa^* h_{t-1})\frac{\partial s_{t-1}}{\partial\lambda_{t-1}}\right)^2\frac{\partial\lambda_{t-1}}{\partial\theta_i}\frac{\partial\lambda_{t-1}}{\partial\theta_j}\right| \\ &= \left|\tilde{x}_{ij,t} + \sum_{p=1}^{\infty}\prod_{l=0}^{p}\left(\phi + (\kappa+\kappa^* h_{t-l})\frac{\partial s_{t-l}}{\partial\lambda_{t-l}}\right)^2\tilde{x}_{ij,t-p}\right|\end{aligned}$$

Once again relying on Lemma 2.1 of Straumann and Mikosch (2006) and Assumption 1.(c), leaving the need to bound $\sup_{\theta\in\Theta}E[log^+\tilde{x}_{ij,t}] < \infty$. The this can be done simply using Lemma 2.8 and the result that $|x_{i,t}|$ is naturally bounded for $i = 1, 2, .., 8$. □

I again note that θ_i refers to the i^{th} element of $\theta = \{\mu, \nu_l, \nu_r, \alpha, \delta, \phi, \kappa, \kappa^*\}$.

Lemma 2.10. *Under Assumption 1.(a)-(c),* $E\left[\sup_{\theta\in\Theta}\left|\frac{\partial^2\lambda_t}{\partial\theta_i\partial\theta_j}\right|\right] < \infty$.

Proof. For all $i, j = \{1, 2, 3, 4, 5, 6, 7, 8\}$,

$$\begin{aligned}\left|\frac{\partial^2\lambda_t}{\partial\theta_i\partial\theta_j}\right| &\leq \left|\frac{\partial x_{i,t}}{\partial\theta_j}\right| + \left|\left(\frac{\partial}{\partial\theta_j}\left[\phi + (\kappa+\kappa^* h_{t-1})\frac{\partial s_{t-1}}{\partial\lambda_{t-1}}\right]\right)\frac{\partial\lambda_{t-1}}{\partial\theta_i}\right| \\ &\quad + \left|\left[\phi + (\kappa+\kappa^* h_{t-1})\frac{\partial s_{t-1}}{\partial\lambda_{t-1}}\right]\frac{\partial^2\lambda_{t-1}}{\partial\theta_i\partial\theta_j}\right| \\ &= |x^*_{ij,t}| + \left|\left[\phi + (\kappa+\kappa^* h_{t-1})\frac{\partial s_{t-1}}{\partial\lambda_{t-1}}\right]\frac{\partial^2\lambda_{t-1}}{\partial\theta_i\partial\theta_j}\right| \\ &\leq |x^*_{ij,t}| + \sum_{p=1}^{\infty}\prod_{l=0}^{p}\left|\phi + (\kappa+\kappa^* h_{t-l})\frac{\partial s_{t-l}}{\partial\lambda_{t-l}}\right||x^*_{ij,t-p}|\end{aligned}$$

As in Lemma 2.4, $\sup_{\theta\in\Theta}\prod_{l=0}^{\infty}\left|\theta + (\kappa+\kappa^* h_{t-l})\frac{\partial s_{t-l}}{\partial\lambda_{t-l}}\right| \to 0 \quad a.s.$ by Lemma 2.4 of Straumann and Mikosch (2006) and Assumption 1.(c), which leaves only to show $\sup_{\theta\in\Theta}E[log^+x^*_{ij,t}] < \infty$ to establish the stated result. I begin by bounding $E\left|\frac{\partial x_{i,t}}{\partial\theta_j}\right|$. For $\{i, j\} = \{5-8, j\}$, this can be done directly using the Lemmas 2.4 and 2.8, together with expressions for $x_{i,t}$ obtained in the former. The remaining terms for $\{i, j\} = \{1-4, j\}$ can be bounded using the results in Lemma 2.4 and very similar arguments to those used in to bound $x_{i,t}$, where use is made of the compact parameter space and resulting bounds on λ_t from Lemma 2.3. To bound the second component of $E|x^*_{ij,t}|$, note that

$$E\left|\left(\frac{\partial}{\partial\theta_j}\left[\phi+\kappa\frac{\partial s_{t-1}}{\partial\lambda_{t-1}}\right]\right)\frac{\partial\lambda_{t-1}}{\partial\theta_i}\right| \leq \sqrt{E\left[\left|\frac{\partial}{\partial\theta_j}\left[\phi+\kappa\frac{\partial s_{t-1}}{\partial\lambda_{t-1}}\right]\right|^2\right]}\sqrt{E\left[\left|\frac{\partial\lambda_{t-1}}{\partial\theta_i}\right|^2\right]}$$

where the second component on the right is bounded by Lemma 2.9. The first component with $j=\{5,6,7,8\}$ can be bounded directly using Lemmas 2.4 and 2.8. For $j=\{1,2,3,4\}$, I need to also bound $E\left|\frac{\partial s_{t-1}(\theta,\sigma_{t-1})}{\partial\lambda_{t-1}\partial\theta_j}\right|^2$. This can be done using very similar arguments to those of Lemma 2.4, though the proofs are long and omitted here to preserve space. Taken together, this gives

$$\begin{aligned}
E\left|\frac{\partial^2 s_{t-1}}{\partial\lambda_{t-1}\partial\theta_j}\right|^2 &= E\left|\frac{\partial^2 s_{t-1}(y_{t-1}|\theta,\sigma_{t-1})}{\partial\lambda_{t-1}\partial\theta_j}+\frac{\partial^2 s_{t-1}}{\partial\lambda_{t-1}^2}\frac{\partial\lambda_{t-1}}{\partial\theta_i}\right|^2\\
&\leq 2E\left|\frac{\partial^2 s_{t-1}(y_{t-1}|\theta,\sigma_{t-1})}{\partial\lambda_{t-1}\partial\theta_j}\right|^2+2E\left[\left|\frac{\partial^2 s_{t-1}}{\partial\lambda_{t-1}^2}\right|^2\left|\frac{\partial\lambda_{t-1}}{\partial\theta_i}\right|^2\right]\\
&\leq 2E\left|\frac{\partial^2 s_{t-1}(y_{t-1}|\theta,\sigma_{t-1})}{\partial\lambda_{t-1}\partial\theta_j}\right|^2+2C_1E\left|\frac{\partial\lambda_{t-1}}{\partial\theta_i}\right|^2<\infty
\end{aligned}$$

by Jensen's inequality and Lemma 2.8. This is sufficient to find bounds for $j=\{1,2,3,4\}$. □

Lemma 2.11. *Under Assumption 1.(a)-(c),* $E\left[\sup_{\theta\in\Theta}\left|\left(\frac{\partial l(y_t|\theta)}{\partial\theta}\right)\left(\frac{\partial l(y_t|\theta)}{\partial\theta}\right)'\right|\right]<\infty$.

Proof. For $i,j=\{1,2,3,4,5,6,7,8\}$, each element of this matrix is bounded by

$$\begin{aligned}
E\left[\sup_{\theta\in\Theta}\left|\frac{\partial l(y_t|\theta)}{\partial\theta_i}\frac{\partial l(y_t|\theta)}{\partial\theta_j}\right|\right] &\leq E\left[\sup_{\theta\in\Theta}\left|\frac{\partial l(y_t|\theta,\lambda_t)}{\partial\theta_i}\frac{\partial l(y_t|\theta,\lambda_t)}{\partial\theta_j}\right|\right]\\
&+E\left[\sup_{\theta\in\Theta}\left|\frac{\partial l(y_t|\theta,\lambda_t)}{\partial\theta_i}\frac{\partial l(y_t|\theta)}{\partial\lambda_t}\frac{\partial\lambda_t}{\partial\theta_j}\right|\right]\\
&+E\left[\sup_{\theta\in\Theta}\left|\frac{\partial l(y_t|\theta,\lambda_t)}{\partial\theta_j}\frac{\partial l(y_t|\theta)}{\partial\lambda_t}\frac{\partial\lambda_t}{\partial\theta_i}\right|\right]\\
&+E\left[\sup_{\theta\in\Theta}\left|\frac{\partial^2 l(y_t|\theta)}{\partial\lambda_t^2}\left(\frac{\partial\lambda_t}{\partial\theta_i}\frac{\partial\lambda_t}{\partial\theta_j}\right)\right|\right]\\
&=H_1+H_2+H_3+H_4
\end{aligned}$$

H_1 is bounded by results in Lemma 2.5, H_2 and H_3 are bounded by Lemmas 2.5 and 2.4

together with the bounds on s_t, while H_4 is bounded by Lemma 2.9 and the bounds on s_t. □

Lemma 2.12. *Under Assumption 1.(a)-(c),* $E\left[\sup_{\theta\in\Theta}\left|\frac{\partial^2 l(y_t|\theta)}{\partial\theta\partial\theta'}\right|\right] < \infty$.

Proof. A full derivation for all 21 elements of the matrix is not possible in the space here. However, the result can be established using previous lemmas and results therein. A sketch off the proof is as follows. For $i, j = \{1,2,3,4,5,6,7,8\}$, each element of this matrix is bounded by

$$
\begin{aligned}
E\left|\sup_{\theta\in\Theta}\frac{\partial^2 l(y_t|\theta)}{\partial\theta_i\partial\theta_j}\right| &\le E\left[\sup_{\theta\in\Theta}\left|\left(\frac{\partial}{\partial\lambda_t}\frac{\partial l(y_t|\theta,\sigma_t)}{\partial\theta_i}\right)\frac{\partial\lambda_t}{\partial\theta_j}\right|\right] + E\left[\sup_{\theta\in\Theta}\left|\left(\frac{\partial}{\partial\theta_j}\frac{\partial l(y_t|\theta,\sigma_t)}{\partial\lambda_t}\right)\frac{\partial\lambda_t}{\partial\theta_i}\right|\right] \\
&+ E\left[\sup_{\theta\in\Theta}\left|\frac{\partial^2 l(y_t|\theta)}{\partial\lambda_t^2}\left(\frac{\partial\lambda_t}{\partial\theta_i}\frac{\partial\lambda_t}{\partial\theta_j}\right)\right|\right] + E\left[\sup_{\theta\in\Theta}\left|\frac{\partial l(y_t|\theta)}{\partial\lambda_t}\frac{\partial^2\lambda_t}{\partial\theta_i\partial\theta_j}\right|\right] \\
&+ E\left[\sup_{\theta\in\Theta}\left|\frac{\partial^2 l(y_t|\theta,\sigma_t)}{\partial\theta_i\partial\theta_j}\right|\right] \\
&= I_1 + I_2 + I_3 + I_4 + I_5
\end{aligned}
$$

I_1 can be bounded[3] from results in Lemma 2.5 together with Lemma 2.4, I_2 can be bounded from Lemma 2.4 and results in that proof on the bounds of $x_{i,t}$, I_3 can be bounded from Lemmas 2.8 and 2.9 and I_4 can be bounded from Lemmas 2.8 and 2.10.I_5 relates to the static model and can be shown to be bounded in much the same way as was done in Lemma 2.5. Inspection of that proof show that I_5 will involve the derivatives of functions (for the lhs of the likelihood) of the form $g(\theta_{2-4}) + h(\theta_{1-5})p(a_t)$, where $p(a_t)$ will be $a_t/(1+a_t^2)$,$a_t^2/(1+a_t^2)$ or $ln(1+a_t)$ and $g(), h()$ are continuous functions of the shape parameters. These once again lead to naturally bounded functions of a_t (and b_t for the rhs) and continuous functions of the bounded shape parameters. □

Lemma 2.13. *Under Assumption 1,* $E\left[\left.\frac{\partial l(y_t|\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right] = 0$.

Proof. The only departure from the standard proof is that $\int f'(y|\theta)dy = \partial/\partial\theta\int f(y|\theta)dy$ cannot be established through continuity of the function $f(y_t|\theta)$. I proceed in bounding the density function by $|f(y_t|\theta)| \le |f^l(y_t|\theta)| + |f^r(y_t|\theta)| \le 2C_1C_2$ where $|1/\sigma_t| \le C_1$ (Lemma 2.3),

[3] Note that in each case, part of term has been shown to be naturally bounded.

$M(\alpha, \nu_l, \nu_r) \leq C_2$ (from Assumption 2) and use is made of $1/(1+x)^p < 1$ when $x > 0$. From this follows

$$\left| \frac{f(y_t|\theta + e_j h) - f(y_t|\theta)}{h} \right| \leq \frac{4C_1C_2}{h}$$

for $0 < h < \bar{h} < \infty$, where e_j is a vector with its j^{th} element as one and zeros elsewhere. Application of the dominated convergence theorem gives the desired result. □

2.5 Proof of Normality

Proof of Theorem 1b) (Asymptotic Normality of AST DCS) Asymptotic normality is established by verifying the conditions of Theorem 7.2 in Newey and McFadden (1994), which explicitly allow for non-smooth objective functions. Consistency was already established in the section above. Condition i) which requires $E\left[\frac{\partial l(y_t|\theta_0)}{\partial \theta_0}\right] = 0$ is given by Lemma 2.13. Condition iii) which requires that θ_0 is an interior point of a compact parameter space is given under Assumption 1.(d). For condition iv), I need to apply a central limit theorem to the scores evaluated at θ_0. Lemma 2.11 established that the scores have a finite second moment (denoted A_0 below). By Lemma 2.2 and condition i) above I know that the sequence $\{\frac{\partial l(y_t|\theta_0)}{\partial \theta_0}\}$ is a stationary and ergodic martingale difference sequence. I can now apply Theorem 19.1 of Billingsley (1999) to get

$$A_0^{-1/2} \frac{1}{\sqrt{T}} \sum_{t}^{[Tr]} \frac{l(y_t|\theta_0)}{\partial \theta_0} \to W(r) \tag{1}$$

which satisfies condition iv). Conditions ii) and v) are verified through Theorem 7.3 in Newey and McFadden (1994). I define $g(y, \theta^*) = \frac{\partial l(y|\theta)}{\partial \theta}\Big|_{\theta=\theta^*}$. The first condition to be satisfied requires that with probability one,

$$r(y, \theta) = |g(y, \theta) - g(y, \theta_0) - \Delta(y)(\theta - \theta_0)| \,/\, |(\theta - \theta_0)| \to 0 \quad , \theta \to \theta_0 \tag{2}$$

for some function $\Delta(y)$. Given that the result must only hold almost surely, I define $\Delta(y) = \frac{\partial^2 l(y|\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_0}$ and use the following result

$$g(y,\theta) - g(y,\theta_0) = \left(\frac{\partial^2 l(y|\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\tilde{\theta}} \right)(\theta-\theta_0) \quad \text{a.s.}$$

form some $\tilde{\theta} \in [\theta, \theta_0]$. Substituting these expressions into equation(2) gives

$$\begin{aligned} r(y,\theta) &= \left| \left(\frac{\partial^2 l(y|\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\tilde{\theta}} - \frac{\partial^2 l(y|\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_0} \right)(\theta-\theta_0) \right| / \left|(\theta-\theta_0)\right| \\ &= \left| \frac{\partial^2 l(y|\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\tilde{\theta}} - \frac{\partial^2 l(y|\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_0} \right| \to 0 \quad ,\theta \to \theta_0 \quad \text{a.s.} \end{aligned}$$

following the continuous mapping theorem which allows for discontinuity on sets of measure zero (see van der Vaart (1998), pg 7). The second condition to be satisfied is that $E[sup_{\|\theta-\theta_0\|<\epsilon}\, r(y,\theta)] < \infty$ for some $\epsilon > 0$. This is can be shown directly with Lemma 2.5 and Lemma 2.12. The third condition is to show that $T^{-1}\sum_{t=1}^{T} \frac{\partial^2 l(y_t|\theta)}{\partial\theta\partial\theta'} \xrightarrow{p} E\left[\frac{\partial^2 l(y_t|\theta)}{\partial\theta\partial\theta'}\right]$. This can be achieved through application of the ergodic theorem (Stout (1974), Theorem 3.5.7) using Lemma 2.2 and Lemma 2.12. The conditions for Theorem 7.3 of Newey and McFadden (1994) are therefore satisfied, which in turn satisfies conditions ii) and v) of the Theorem 7.3. Finally, the asymptotic first order condition can be shown to hold with a proof directly following Ruppert and Carroll (1980) and Komunjer (2005). This provides asymptotic normality results as in Theorem 7.2 Newey and McFadden (1994), with the covariance matrix collapsing down to the inverse of the information matrix by the usual information matrix equality[4] The result regarding consistent estimation of the information matrix comes from the dominated convergence theorem, where a dominating function can be formed from the results in Lemma 2.5. □

[4]This is obtained by the dominated convergence theorem as in Lemma 2.13 by noting that $\frac{\partial^2 f(y_t|\theta)}{\partial\theta_i\partial\theta_j} = lim_{h\to\infty}\left|\frac{f(y_t|\theta+e_i h)-2f(y_t|\theta)+f(y_t|\theta-e_j h)}{h^2}\right|$ and a dominating function can be formed from the results in Lemma 2.5.

2.6 Proof of Asymptotic Properties under Misspecification

Proof of Theorem 2 (CAN of QMLE for AST DCS under Misspecification) Part a) of this theorem is established under Theorem 3.4 of White (1994) while part b) is again established under Theorem 7.2 of Newey and McFadden (1994). The proof is essentially the same as for Theorem 1 above with $\tilde{\theta}$ taking the place of θ_0 and q taking the place of $\underline{\nu}$ when bounding $E|y^\epsilon|$ for some $\epsilon > 0$. Stationarity and ergodicity of the process $\{\lambda_t\}_{t \in \mathbb{Z}}$ is now guaranteed directly by Assumption 2.(c) and no longer just $|\phi| < 1$. The result in Lemma 2.13 now follows from the fact that the pseudo-true parameter $\tilde{\theta}$ is the unique maximiser of the expected pseudo-likelihood. The remaining results still hold as the proofs there were constructed to leverage only the function form of the objective function and never the true distribution of errors. The final point regarding the estimation of $\tilde{B}$ follows from standard results such as Lemma 4.3 of Newey and McFadden (1994) with the use of an appropriate law of large numbers, such as the ergodic theorem of Stout (1974, Theorem 3.5.7). $\square$

3 Additional Simulation Results

Table A1 presents an alternative measure of central tendency to to the mean estimate results presented in Section 3 of the main paper. The results show that all parameters are essentially median unbiased with sample sizes of $T = 1000$ or more.

Table A1: **Median of AST DCS Model Estimates**

T	ϕ, α, ν_R	μ	α	ν_L	ν_R	δ	ϕ	κ
500	0.95, 0.6, 10	0.084	0.598	5.086	11.898	0.062	0.938	0.054
1000		0.091	0.598	5.051	10.862	0.056	0.944	0.050
2000		0.095	0.599	5.026	10.421	0.053	0.947	0.050
500	0.95, 0.6, 30	0.094	0.600	5.122	37.652	0.061	0.939	0.054
1000		0.098	0.600	5.041	33.011	0.056	0.944	0.050
2000		0.099	0.600	5.028	31.981	0.053	0.947	0.050
500	0.99, 0.6, 10	0.052	0.599	5.232	11.838	0.093	0.981	0.054
1000		0.082	0.600	5.149	10.858	0.070	0.986	0.049
2000		0.092	0.600	5.052	10.420	0.060	0.988	0.050
500	0.95, 0.3, 10	0.085	0.297	5.160	11.226	0.060	0.939	0.054
1000		0.091	0.298	5.063	10.588	0.056	0.944	0.050
2000		0.094	0.299	5.027	10.264	0.053	0.947	0.050
500	0.95, 0.6, 7	0.079	0.596	6.245	7.948	0.062	0.938	0.054
1000		0.089	0.597	5.046	7.467	0.056	0.944	0.050
2000		0.094	0.598	5.018	7.209	0.053	0.947	0.050
500	0.95, 0.6, 3	0.083	0.598	5.176	3.101	0.062	0.938	0.054
1000		0.090	0.599	5.102	3.045	0.056	0.944	0.050
2000		0.094	0.599	5.045	3.024	0.053	0.947	0.050

Note: Other parameters set at $\{\mu, \nu_L, \delta, \kappa\} = \{0.1, 5, 0.05, 0.05\}$. Based on 10,000 replications per setting. MSE times 1000 is reported for all parameters except ν_L and ν_R.

Table A2 presents an alternative to coverage results presented in Section 3 in the main paper. Here I replace the standard errors estimated for each parameter via finite differences of the likelihood with the standard deviation of the Monte Carlo estimates. The aim is to examine the validity of the asymptotic approximation in reasonable sample sizes, abstracting from the issue of obtaining accurate estimates of the standard errors. The results show that coverage properties are very good in this case.

Table A2: **Coverage of AST DCS Model Alternative 95% Confidence Intervals**

T	ϕ, α, ν_R	μ	α	ν_L	ν_R	δ	ϕ	κ
500	0.95, 0.6, 10	0.93	0.94	0.99	0.96	0.95	0.95	0.96
1000		0.94	0.95	0.96	0.98	0.94	0.94	0.95
2000		0.96	0.96	0.95	0.98	0.95	0.95	0.95
500	0.95, 0.6, 30	0.94	0.95	0.99	0.89	0.95	0.95	0.96
1000		0.96	0.96	0.98	0.93	0.95	0.95	0.95
2000		0.96	0.95	0.95	0.96	0.94	0.94	0.95
500	0.99, 0.6, 10	0.98	0.95	0.99	0.90	0.97	0.97	0.94
1000		0.99	0.95	0.95	0.96	0.94	0.94	0.95
2000		0.99	0.95	0.95	0.99	0.93	0.92	0.95
500	0.95, 0.3, 10	0.93	0.98	0.98	0.97	0.95	0.95	0.96
1000		0.94	0.95	0.99	0.98	0.95	0.95	0.95
2000		0.95	0.96	0.97	0.95	0.94	0.95	0.95
500	0.95, 0.6, 7	0.93	0.93	0.98	0.98	0.95	0.95	0.96
1000		0.93	0.94	0.95	0.97	0.95	0.95	0.95
2000		0.93	0.94	0.95	0.96	0.94	0.94	0.95
500	0.95, 0.6, 3	0.93	0.94	0.99	0.99	0.95	0.95	0.96
1000		0.95	0.95	0.94	0.95	0.95	0.95	0.95
2000		0.96	0.93	0.95	0.95	0.94	0.94	0.95

Note: Other parameters set at $\{\mu, \nu_L, \delta, \kappa\} = \{0.1, 5, 0.05, 0.05\}$. Based on 10,000 replications per setting. MSE times 1000 is reported for all parameters except ν_L and ν_R.

4 Full Sample Estimates - AST DCS and Restricted Models

This section presents in-sample estimation results for the S&P 500, DJIA, NASDAQ, Kospi and Bovespa equity market indices. The results show estimation for the AST DCS, AT DCS, ST DCS and T DCS models all with leverage, together with some model diagnostics and model comparison statistics.

Table A3: **Parameter Estimates for AST DCS and Restricted Models - S&P 500**

	μ	ν_l	ν_r	α	δ	ϕ	κ	κ^*	lnL	BIC	$LB^2(20)$	KS
AST DCS	0.066	7.74	105.2	0.519	-0.001	0.987	0.043	0.056	-5004.6	10075.7	0.191	0.003
	(0.026)	(1.12)	(235.6)	(0.011)	(0.001)	(0.002)	(0.005)	(0.004)				
AT DCS	0.030	6.94	330.0	-	-0.002	0.985	0.042	0.055	-5006.7	10071.5	0.048	0.004
	(0.013)	(0.92)	(161.5)		(0.001)	(0.003)	(0.005)	(0.004)				
ST DCS	0.106	10.75	-	0.545	-0.003	0.989	0.044	0.057	-5011.1	10080.5	0.009	0.000
	(0.021)	(1.63)		(0.009)	(0.001)	(0.003)	(0.006)	(0.005)				
T DCS	0.0177	10.28	-	-	-0.004	0.984	0.041	0.055	-5024.5	10098.3	0.163	0.000
	(0.013)	(1.49)			(0.001)	(0.003)	(0.006)	(0.005)				

Note: AST is the asymmetric skew t distribution. Special cases are the asymmetric t (AT) with $\alpha = 0.5$, skewed t (ST) with $\nu_L = \nu_R$ and Student's t with $\alpha = 0.5, \nu_L = \nu_R$. Values in parenthesis are standard errors. lnL is the log likelihood value. BIC is defined so that a lower value is preferred. $LB^2(20)$ is a p-value from a Ljung-Box test applied to the scores s_t with lag length of 20. KS is the p-value from a bootstrap Kolmogorov-Smirnov test to allow for estimated parameters in the distribution.

Table A4: **Parameter Estimates for AST DCS and Restricted Models - DJIA**

	μ	ν_l	ν_r	α	δ	ϕ	κ	κ^*	lnL	BIC	$LB^2(20)$	KS
AST DCS	0.069	7.15	24.62	0.517	-0.002	0.986	0.050	0.053	-4881.2	9828.1	0.671	0.011
	(0.009)	(1.04)	(12.54)	(0.007)	(0.001)	(0.003)	(0.006)	(0.005)				
AT DCS	0.039	6.56	39.16	-	-0.002	0.984	0.049	0.052	-4882.4	9822.4	0.224	0.027
	(0.000)	(0.83)	(10.79)		(0.001)	(0.002)	(0.006)	(0.004)				
ST DCS	0.104	9.74	-	0.539	-0.004	0.987	0.050	0.054	-4886.3	9830.0	0.014	0.002
	(0.021)	(1.36)		(0.009)	(0.001)	(0.003)	(0.006)	(0.005)				
T DCS	0.030	9.36	-	-	-0.004	0.983	0.048	0.052	-4896.0	9841.2	0.379	0.004
	(0.012)	(1.26)			(0.001)	(0.003)	(0.006)	(0.005)				

Note: AST is the asymmetric skew t distribution. Special cases are the asymmetric t (AT) with $\alpha = 0.5$, skewed t (ST) with $\nu_L = \nu_R$ and Student's t with $\alpha = 0.5, \nu_L = \nu_R$. Values in parenthesis are standard errors. lnL is the log likelihood value. BIC is defined so that a lower value is preferred. $LB^2(20)$ is a p-value from a Ljung-Box test applied to the scores s_t with lag length of 20. KS is the p-value from a bootstrap Kolmogorov-Smirnov test to allow for estimated parameters in the distribution.

Table A5: **Parameter Estimates for AST DCS and Restricted Models - NASDAQ**

	μ	ν_l	ν_r	α	δ	ϕ	κ	κ^*	lnL	BIC	$LB^2(20)$	KS
AST DCS	0.105	9.59	34.69	0.541	-0.003	0.993	0.055	0.037	-5600.4	11266.6	0.366	0.013
	(0.000)	(2.22)	(19.45)	(0.006)	(0.001)	(0.002)	(0.006)	(0.004)				
AT DCS	0.017	7.32	330.0	-	-0.000	0.991	0.051	0.035	-5607.9	11273.0	0.058	0.026
	(0.015)	(1.12)	(65.17)		(0.001)	(0.002)	(0.006)	(0.004)				
ST DCS	0.144	14.20	-	0.559	-0.001	0.993	0.057	0.035	-5603.4	11264.1	0.457	0.004
	(0.025)	(2.77)		(0.009)	(0.001)	(0.002)	(0.007)	(0.004)				
T DCS	0.005	14.41	-	-	-0.000	0.990	0.053	0.032	-5625.4	11300.0	0.087	0.002
	(0.015)	(2.87)			(0.001)	(0.002)	(0.007)	(0.004)				

Note: AST is the asymmetric skew t distribution. Special cases are the asymmetric t (AT) with $\alpha = 0.5$, skewed t (ST) with $\nu_L = \nu_R$ and Student's t with $\alpha = 0.5, \nu_L = \nu_R$. Values in parenthesis are standard errors. lnL is the log likelihood value. BIC is defined so that a lower value is preferred. $LB^2(20)$ is a p-value from a Ljung-Box test applied to the scores s_t with lag length of 20. KS is the p-value from a bootstrap Kolmogorov-Smirnov test to allow for estimated parameters in the distribution.

Table A6: **Parameter Estimates for AST DCS and Restricted Models - Kospi**

	μ	ν_l	ν_r	α	δ	ϕ	κ	κ^*	lnL	BIC	$LB^2(20)$	KS
AST DCS	-0.020	6.89	48.13	0.503	0.000	0.991	0.064	0.020	-5394.3	10854.3	0.272	0.384
	(0.000)	(1.17)	(44.05)	(0.006)	(0.001)	(0.003)	(0.006)	(0.004)				
AT DCS	-0.021	6.74	48.85	-	-0.000	0.991	0.063	0.019	-5394.4	10846.4	0.115	0.614
	(0.014)	(0.94)	(9.56)		(0.001)	(0.003)	(0.007)	(0.004)				
ST DCS	0.034	11.01	-	0.531	-0.001	0.991	0.066	0.019	-5403.0	10863.5	0.084	0.093
	(0.025)	(1.80)		(0.009)	(0.001)	(0.003)	(0.007)	(0.004)				
T DCS	-0.036	10.34	-	-	-0.001	0.990	0.065	0.019	-5408.7	10866.7	0.138	0.097
	(0.015)	(1.57)			(0.001)	(0.003)	(0.007)	(0.004)				

Note: AST is the asymmetric skew t distribution. Special cases are the asymmetric t (AT) with $\alpha = 0.5$, skewed t (ST) with $\nu_L = \nu_R$ and Student's t with $\alpha = 0.5, \nu_L = \nu_R$. Values in parenthesis are standard errors. lnL is the log likelihood value. BIC is defined so that a lower value is preferred. $LB^2(20)$ is a p-value from a Ljung-Box test applied to the scores s_t with lag length of 20. KS is the p-value from a bootstrap Kolmogorov-Smirnov test to allow for estimated parameters in the distribution.

Table A7: **Parameter Estimates for AST DCS and Restricted Models - Bovespa**

	μ	ν_l	ν_r	α	δ	ϕ	κ	κ^*	lnL	BIC	$LB^2(20)$	KS
AST DCS	0.018	10.00	35.59	0.502	0.008	0.983	0.047	0.028	-6990.8	14047.3	0.314	0.737
	(0.065)	(2.23)	(24.51)	(0.015)	(0.002)	(0.005)	(0.006)	(0.004)				
AT DCS	0.008	9.90	37.04	-	0.008	0.983	0.047	0.028	-6990.8	14039.1	0.308	0.614
	(0.025)	(1.84)	(8.18)		(0.002)	(0.005)	(0.006)	(0.004)				
ST DCS	0.099	14.96	-	0.524	0.068	0.983	0.047	0.029	-6993.8	14045.1	0.556	0.526
	(0.051)	(3.01)		(0.011)	(0.002)	(0.005)	(0.006)	(0.004)				
T DCS	-0.004	14.37	-	-	0.008	0.982	0.046	0.029	-6996.4	14042.1	0.336	0.395
	(0.025)	(2.77)			(0.002)	(0.005)	(0.006)	(0.004)				

Note: AST is the asymmetric skew t distribution. Special cases are the asymmetric t (AT) with $\alpha = 0.5$, skewed t (ST) with $\nu_L = \nu_R$ and Student's t with $\alpha = 0.5, \nu_L = \nu_R$. Values in parenthesis are standard errors. lnL is the log likelihood value. BIC is defined so that a lower value is preferred. $LB^2(20)$ is a p-value from a Ljung-Box test applied to the scores s_t with lag length of 20. KS is the p-value from a bootstrap Kolmogorov-Smirnov test to allow for estimated parameters in the distribution.

5 Out-of-sample Portfolio Construction - Frank Copula

This section replicates the results reported in Section 4.3 of the main paper with a Frank copula instead of the Clayton copula. The main difference between these two dependence models is that the Frank copula is symmetric, allowing for tail dependence for both gains and losses. The Clayton copula only allows for tail dependence amongst losses. These results are reported to show that the out-of-sample results appear reasonably robust to the copula specification, as the reported management fees (and statistical significance) are qualitatively and quantitatively very similar.

Table A8: **Out-of-sample Management Fees - Frank Copula**

	FTSE	S&P	DJIA	NASDAQ	Kospi	Bovespa
			Panel A: AST DCS vs ST DCS			
FTSE	-	0.84	0.73	2.34*	3.16*	1.86*
S&P	0.21	-	-0.67	0.62*	2.48*	2.55*
DJIA	0.19	0.77	-	-0.07	2.13*	1.24*
NASDAQ	0.02	0.32	0.50	-	2.19*	1.92*
Kospi	0.02	0.08	0.08	0.08	-	2.74*
Bovespa	0.07	0.04	0.17	0.06	0.03	-
			Panel B: AT DCS vs ST DCS			
FTSE	-	0.53	1.29	2.56*	4.03*	4.78*
S&P	0.38	-	-1.07	2.47*	4.63*	6.57*
DJIA	0.13	0.83	-	1.59	4.60*	6.16*
NASDAQ	0.10	0.09	0.15	-	8.98*	4.60*
Kospi	0.10	0.08	0.06	0.01	-	4.99*
Bovespa	0.00	0.00	0.00	0.04	0.01	-

Note: AST is the asymmetric skew t distribution. Special cases are the asymmetric t (AT) with $\alpha = 0.5$ and the skewed t (ST) with $\nu_L = \nu_R$. Upper triangle values are the management fees (percent per year) that a CRRA investor would pay to invest according to the AST/AT DCS model over the ST DCS model. Lower triangle provides the p-value for superior predictive ability of the ST DCS model. * indicates significance at a 10% level.

6 References - Supplementary Appendix

Andrews DW. 1992. Generic Uniform Convergence. *Econometric Theory* **8**: 241 - 257. DOI:http://dx.doi.org/10.1017/S0266466600012780

Billingsley P. 1999. *Convergence at Probability Measures*. Wiley: New York.

Bougerol P. 199). Kalman Filtering with Random Coefficients and Contractions. *Siam Journal of Control and Optimization* **31**(4): 942-959. DOI:10.1137/0331041

Brandt A. 1986. The Stochastic Equation Yn+1 = An Yn + Bn with Stationary Coefficients. *Advances in Applied Probability* **18**: 211 - 220. DOI:10.2307/1427243

Davidson J. 1994. *Stochastic Limit Theory*, Oxford University Press: New York.

Komunjer I. 2005. Quasi-Maximum Likelihood Estimation for Conditional Quantiles. *Journal of Econometrics* **128**: 137-164. DOI:10.1016/j.jeconom.2004.08.010

Lumsdaine RL. 1996. Consistency and Asymptotic Normality of the Quasi-maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models. *Econometrica* **64**: 575-96. DOI: 10.2307/2171862

Newey WK, McFadden DL. 1994. Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, Vol. 4, Engle RF, McFadden DL (eds). Elsevier Science: Amsterdam; 2113-2247.

Ruppert D, Carroll RJ. 1980. Trimmed Least Squares Estimation in the Linear Model. *Journal of the American Statistical Association* **75**: 828-838. DOI: 10.1080/01621459.1980.10477560

Stout WF. 1974. *Almost Sure Convergence, Probability and Mathematical Statistics*. Academic Press: New York.

Straumann D, Mikosch T. 2006. Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: a Stochastic recurrence Equations Approach. *Annals of Statistics* **34**: 2449-2495. DOI: 10.1214/009053606000000803

Vaart, A. W. van der (1998). *Asymptotic Statistics*, Cambridge University Press: Cambridge.

Weiss AA. 1986. Asymptotic Theory for ARCH Models: Estimation and Testing. *Econometric Theory* **2**: 107-131. DOI: http://dx.doi.org/10.1017/S0266466600011397

Zhu D. 2012. Asymmetric Parametric Distributions and a New Class of Asymmetric Generalized t-Distribution. SSRN Working Paper - 2427545.

Zhu D, Galbraith JW. 2010. A Generalized Asymmetric Student-t Distribution with Application to Financial Econometrics. *Journal of Econometrics* **157**: 297-305. DOI:10.1016/j.jeconom.2010.0