Supplemental Appendix to:

Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk)

Andrew J. Patton — Duke University

Johanna F. Ziegel — University of Bern

Rui Chen — Duke University

12 September 2018

This appendix contains three parts. Part 1 presents lemmas that provide further details on the proof of Theorem 2 presented in the main paper. Part 2 presents a detailed verification that the high level assumptions made in the theorems of the paper hold for the widely-used GARCH(1,1) process. Part 3 contains additional tables of analysis.

Appendix SA.1: Detailed proofs

Throughout this appendix, we suppress the subscript on $\hat{\boldsymbol{\theta}}_T$ for simplicity of presentation, and we denote the conditional distribution and density functions as F_t and f_t rather than $F_t(\cdot|\mathcal{F}_{t-1})$ and $f_t(\cdot|\mathcal{F}_{t-1})$.

In Lemmas 1 and 3 below, we will refer to the expected score, defined as:

$$\begin{aligned}\lambda(\boldsymbol{\theta}) &= \mathbb{E}\left[g_t(\boldsymbol{\theta})\right] \qquad (1)\\ &= \mathbb{E}\Bigg[\frac{1}{-e_t(\boldsymbol{\theta})}\left(\frac{F_t\left(v_t(\boldsymbol{\theta})\right)}{\alpha}-1\right)\nabla v_t(\boldsymbol{\theta})'+\\ &\frac{1}{e_t(\boldsymbol{\theta})^2}\left(\frac{F_t\left(v_t(\boldsymbol{\theta})\right)}{\alpha}v_t(\boldsymbol{\theta})-\frac{1}{\alpha}\mathbb{E}_{t-1}[Y_t 1\left\{Y_t\leq v_t(\boldsymbol{\theta})\right\}]-v_t(\boldsymbol{\theta})+e_t(\boldsymbol{\theta})\right)\nabla e_t(\boldsymbol{\theta})'\Bigg]\end{aligned}$$

Lemma 1 *Let*

$$\Lambda(\boldsymbol{\theta}^*) = \left.\frac{\partial\mathbb{E}\left[g_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \qquad (2)$$

Then under Assumptions 1-2,

$$\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0) = \left(\Lambda^{-1}(\boldsymbol{\theta}^0)+o_p(1)\right)\left(-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_t(\boldsymbol{\theta}^0)+o_p(1)\right) \qquad (3)$$

Proof of Lemma 1. Consider a mean-value expansion of $\lambda(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}^0$:

$$\lambda(\hat{\boldsymbol{\theta}}) = \lambda(\boldsymbol{\theta}^0) + \left.\frac{\partial \mathbb{E}\left[g_t(\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \tag{4}$$

$$= \Lambda(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \tag{5}$$

where $\boldsymbol{\theta}^*$ lies between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^0$, and noting that $\lambda(\boldsymbol{\theta}^0) = 0$ and the definition of $\Lambda(\boldsymbol{\theta}^*)$ given in the statement of the lemma. Proving the claim involves two results: (I) $\Lambda^{-1}(\boldsymbol{\theta}^*) = \Lambda^{-1}(\boldsymbol{\theta}^0) + o_p(1)$, and (II) $\sqrt{T}\lambda(\hat{\boldsymbol{\theta}}) = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\boldsymbol{\theta}^0) + o_p(1)$. Part (I) is easy to verify: Since $v_t(\boldsymbol{\theta})$ and $e_t(\boldsymbol{\theta})$ are twice continuously differentiable, and $e_t(\boldsymbol{\theta}^0) < 0$, $\Lambda(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ and $\Lambda(\boldsymbol{\theta})$ is non-singular in a neighborhood of $\boldsymbol{\theta}^0$. Then by the continuous mapping theorem, $\boldsymbol{\theta}^* \xrightarrow{p} \boldsymbol{\theta}^0 \Rightarrow \Lambda(\boldsymbol{\theta}^*)^{-1} \xrightarrow{p} \Lambda^{-1}(\boldsymbol{\theta}^0)$. Establishing (II) builds on Theorem 3 of Huber (1967) and Lemma A.1 of Weiss (1991), which extends Huber's conclusion to the case of non-*iid* dependent random variables. We are going to verify the conditions of Weiss's Lemma A.1. Since the other conditions are easily checked, we only need to show that $T^{-1/2}\sum_{t=1}^{T} g_t(\hat{\boldsymbol{\theta}}) = o_p(1)$, which we show in Lemma 2, and that his assumptions N3 and N4 hold, which we show in Lemmas 3-6. ■

Lemma 2 *Under Assumptions 1-2,* $T^{-1/2}\sum_{t=1}^{T} g_t(\hat{\boldsymbol{\theta}}) = o_p(1)$.

Proof of Lemma 2. Let $\{e_j\}_{j=1}^{p}$ be the standard basis of $\mathbb{R}^p$ and define

$$L_T^j(a) = T^{-1/2}\sum_{t=1}^{T} L_{FZ0}\left(Y_t, v_t(\hat{\boldsymbol{\theta}} + ae_j), e_t(\hat{\boldsymbol{\theta}} + ae_j); \alpha\right) \tag{6}$$

where a is a scalar. Let $G_T^j(a)$ (a scalar) be the right partial derivative of $L_T^j(a)$, that is

$$G_T^j(a) = T^{-1/2}\sum_{t=1}^{T}\left(\frac{\nabla_j v_t(\hat{\boldsymbol{\theta}} + ae_j)}{-e_t(\hat{\boldsymbol{\theta}} + ae_j)}(\frac{1}{\alpha}\mathbf{1}\left\{Y_t \le v_t(\hat{\boldsymbol{\theta}} + ae_j)\right\} - 1) + \right. \tag{7}$$
$$\left.\frac{\nabla_j e_t(\hat{\boldsymbol{\theta}} + ae_j)}{e_t(\hat{\boldsymbol{\theta}} + ae_j)^2}\left(\frac{1}{\alpha}\mathbf{1}\left\{Y_t \le v_t(\hat{\boldsymbol{\theta}} + ae_j)\right\}(v_t(\hat{\boldsymbol{\theta}} + ae_j) - Y_t) - v_t(\hat{\boldsymbol{\theta}} + ae_j) + e_t(\hat{\boldsymbol{\theta}} + ae_j)\right)\right)$$

$G_T^j(0) = \lim_{\xi_1 \to 0+} G_T^j(\xi_1)$ is the right partial derivative of $L_T(\boldsymbol{\theta})$ at $\hat{\boldsymbol{\theta}}$ in the direction θ_j, while $\lim_{\xi_2 \to 0+} G_T^j(-\xi_2)$ is the left partial derivative of $L_T(\boldsymbol{\theta})$ at $\hat{\boldsymbol{\theta}}$ in the direction θ_j. Because $L_T(\boldsymbol{\theta})$ achieves its minimum at $\hat{\boldsymbol{\theta}}$, and its left and right partial derivatives exist, its left derivative must be non-positive and its right derivative must be non-negative. Thus,

$$
\begin{aligned}
|G_T^j(0)| &\leq \lim_{\xi_1 \to 0+} G_T^j(\xi_1) - \lim_{\xi_2 \to 0+} G_T^j(-\xi_2) \\
&= T^{-1/2} \sum_{t=1}^{T} \left(\frac{\nabla_j v_t(\hat{\boldsymbol{\theta}})}{-e_t(\hat{\boldsymbol{\theta}})} \frac{1}{\alpha} \mathbf{1}\left\{ Y_t = v_t(\hat{\theta}) \right\} + \frac{\nabla_j e_t(\hat{\boldsymbol{\theta}})}{e_t(\hat{\boldsymbol{\theta}})^2} \frac{1}{\alpha} \left(v_t(\hat{\boldsymbol{\theta}}) - Y_t \right) \mathbf{1}\left\{ Y_t = v_t(\hat{\boldsymbol{\theta}}) \right\} \right) \quad (8) \\
&= T^{-1/2} \sum_{t=1}^{T} \frac{|\nabla_j v_t(\hat{\boldsymbol{\theta}})|}{-e_t(\hat{\theta})} \frac{1}{\alpha} \mathbf{1}\left\{ Y_t = v_t(\hat{\boldsymbol{\theta}}) \right\}
\end{aligned}
$$

The second term in the penultimate line vanishes as $\mathbf{1}\{Y_t = v_t(\hat{\boldsymbol{\theta}})\}(v_t(\hat{\boldsymbol{\theta}}) - Y_t)$ is always zero.

By Assumption 2(C), for all t, $|\nabla_j v_t(\hat{\boldsymbol{\theta}})| \leq \|\nabla v_t(\hat{\boldsymbol{\theta}})\| \leq V_1(\mathcal{F}_{t-1})$ and $\left|1/e_t(\hat{\boldsymbol{\theta}})\right| \leq H$, thus:

$$
|G_T^j(0)| \leq \frac{H}{\alpha} \left[T^{-1/2} \max_{1 \leq t \leq T} V_1(\mathcal{F}_{t-1}) \right] \left[\sum_{t=1}^{T} \mathbf{1}\left\{ Y_t = v_t(\hat{\boldsymbol{\theta}}) \right\} \right] \quad (9)
$$

H is finite by Assumption 2(C). Next note that for all $\epsilon > 0$,

$$
\Pr\left[T^{-1/2} \max_{1 \leq t \leq T} V_1(\mathcal{F}_{t-1}) > \epsilon \right] \leq \sum_{t=1}^{T} \Pr\left[V_1(\mathcal{F}_{t-1}) > \epsilon T^{1/2} \right] \leq \sum_{t=1}^{T} \frac{\mathbb{E}[V_1(\mathcal{F}_{t-1})^3]}{\epsilon^3 T^{3/2}} \to 0 \quad (10)
$$

with the latter inequality following from Markov's inequality. Since $\mathbb{E}[V_1(\mathcal{F}_{t-1})^3]$ is finite by assumption 2(D), we then have that $T^{-1/2} \max_{1 \leq t \leq T} V_1(\mathcal{F}_{t-1}) = o_p(1)$. Finally, by Assumption 2(G) we have $\sum_{t=1}^{T} \mathbf{1}\left\{ Y_t = v_t(\hat{\boldsymbol{\theta}}) \right\} = \mathcal{O}_{a.s.}(1)$. We therefore have $G_T^j(0) \xrightarrow{p} 0$. Since this holds for every j, we have $T^{-1/2} \sum_{t=1}^{T} g_t(\hat{\boldsymbol{\theta}}) = o_p(1)$.

■

The following three lemmas show each of the three parts of Assumption N3 of Weiss (1991) holds. In the proofs below we make repeated use of mean-value expansions, and we use $\boldsymbol{\theta}^*$ to denote a point on the line connecting $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^0$, and $\boldsymbol{\theta}^{**}$ to denote a point on the line connecting $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^0$. The particular point on the line can vary from expansion to expansion.

Lemma 3 *Under assumptions 1-2, Assumption N3(i) of Weiss (1991) holds:*

$$
\|\lambda_T(\boldsymbol{\theta})\| \geq a \|\boldsymbol{\theta} - \boldsymbol{\theta}^0\|, \textit{ for } \|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| \leq d_0.
$$

for T sufficiently large, where a and d_0 are strictly positive numbers.

Proof of Lemma 3. A mean-value expansion yields:

$$\lambda_T(\hat{\boldsymbol{\theta}}) = \lambda_T(\boldsymbol{\theta}^0) + \Lambda(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) = \Lambda_T(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \tag{11}$$

since $\lambda_T(\boldsymbol{\theta}^0) = 0$, where $\Lambda(\boldsymbol{\theta}) = \partial\mathbb{E}[g_t(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}$. Using the fact that

$$\frac{\partial\mathbb{E}[Y_t\mathbf{1}\{Y_t \le v_t(\boldsymbol{\theta})\}|\mathcal{F}_{t-1}]}{\partial\boldsymbol{\theta}} = \frac{\partial}{\partial\boldsymbol{\theta}}\left\{\int_{-\infty}^{v_t(\boldsymbol{\theta})} y f_t(y)dy\right\} = v_t(\boldsymbol{\theta})f_t(v_t(\boldsymbol{\theta}))\nabla v_t(\boldsymbol{\theta}) \tag{12}$$

we can write:

$$\begin{aligned}
\Lambda(\boldsymbol{\theta}) = \mathbb{E}\Bigg[&\left(\frac{\nabla^2 v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} + \frac{\nabla v_t(\boldsymbol{\theta})'\nabla e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} + \frac{\nabla e_t(\boldsymbol{\theta})'\nabla v_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2}\right)\left(\frac{F_t(v_t(\boldsymbol{\theta}))}{\alpha} - 1\right) \\
&+ \left(\nabla^2 e_t(\boldsymbol{\theta})\frac{1}{e_t(\boldsymbol{\theta})^2} + \frac{-2}{e_t(\boldsymbol{\theta})^3}\nabla e_t(\boldsymbol{\theta})'\nabla e_t(\boldsymbol{\theta})\right) \\
&\cdot\left(\left(\frac{F_t(v_t(\boldsymbol{\theta}))}{\alpha} - 1\right)v_t(\boldsymbol{\theta}) - \frac{1}{\alpha}\mathbb{E}\left[Y_t\mathbf{1}\{Y_t \le v_t(\boldsymbol{\theta})\}|\mathcal{F}_{t-1}\right] + e_t(\boldsymbol{\theta})\right) \\
&+ \frac{f_t(v_t(\boldsymbol{\theta}))}{-\alpha e_t(\boldsymbol{\theta})}\nabla' v_t(\boldsymbol{\theta})\nabla v_t(\boldsymbol{\theta}) \\
&+ \frac{1}{e_t(\boldsymbol{\theta})^2}\nabla' e_t(\boldsymbol{\theta})\nabla e_t(\boldsymbol{\theta})]\}\Bigg|\,\mathcal{F}_{t-1}\Bigg]
\end{aligned} \tag{13}$$

Evaluated at $\boldsymbol{\theta}^0$, the first two terms of Λ drop out because $F_t\left(v_t(\boldsymbol{\theta}^0)\right) = \alpha$ and $\frac{1}{\alpha}\mathbb{E}[Y_t\mathbf{1}\{Y_t \le v_t(\boldsymbol{\theta}^0)\}|\mathcal{F}_{t-1}] = e_t\left(\boldsymbol{\theta}^0\right)$. Define D as

$$D \equiv \Lambda(\boldsymbol{\theta}^0) = T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[\frac{f_t(v_t(\boldsymbol{\theta}^0))}{-\alpha e_t(\boldsymbol{\theta}^0)}\nabla v_t(\boldsymbol{\theta}^0)'\nabla v_t(\boldsymbol{\theta}^0) + \frac{1}{e_t(\boldsymbol{\theta}^0)^2}\nabla e_t(\boldsymbol{\theta}^0)'\nabla e_t(\boldsymbol{\theta}^0)\right] \tag{14}$$

Below we show that $\Lambda(\boldsymbol{\theta}^*) = D + O(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\|)$ by decomposing $\|\Lambda_T(\boldsymbol{\theta}^*) - D\|$ into four terms and showing that each is bounded by a $O(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\|)$ term.

First term: Using a mean-value expansion around $\boldsymbol{\theta}^0$ and Assumptions 2(C)-(D) we obtain:

$$\begin{aligned}
&\left\|\mathbb{E}\left[\left(\frac{\nabla^2 v_t(\boldsymbol{\theta}^*)}{-e_t(\boldsymbol{\theta}^*)} + \frac{\nabla v_t(\boldsymbol{\theta}^*)'\nabla e_t(\boldsymbol{\theta}^*)}{e_t(\boldsymbol{\theta}^*)^2} + \frac{\nabla e_t(\boldsymbol{\theta}^*)'\nabla v_t(\boldsymbol{\theta}^*)}{e_t(\boldsymbol{\theta}^*)^2}\right)\left(\frac{F_t(v_t(\boldsymbol{\theta}^*))}{\alpha} - 1\right)\right]\right\| \\
\le\ &\mathbb{E}\left[\left\|\left(HV_2(\mathcal{F}_{t-1}) + 2H^2V_1(\mathcal{F}_{t-1})H_1(\mathcal{F}_{t-1})\right)\left(\frac{f_t(v_t(\boldsymbol{\theta}^{**}))}{\alpha}\nabla v_t(\boldsymbol{\theta}^{**})(\boldsymbol{\theta}^* - \boldsymbol{\theta}^0)\right)\right\|\right] \qquad (15) \\
\le\ &\frac{K}{\alpha}\left\{H\mathbb{E}[V_1(\mathcal{F}_{t-1})^3]^{1/3}\mathbb{E}[V_2(\mathcal{F}_{t-1})^{3/2}]^{2/3} + 2H^2\mathbb{E}[V_1(\mathcal{F}_{t-1})^3]^{2/3}\mathbb{E}[H_1(\mathcal{F}_{t-1})^3]^{1/3}\right\}\|\boldsymbol{\theta}^* - \boldsymbol{\theta}^0\|
\end{aligned}$$

Second term: Again using a mean-value expansion around $\boldsymbol{\theta}^0$ and Assumptions 2(C)-(D):

$$\left\|\mathbb{E}\left[\left(\frac{1}{e_t(\boldsymbol{\theta}^*)^2}\nabla^2 e_t(\boldsymbol{\theta}^*) - \frac{2}{e_t(\boldsymbol{\theta}^*)^3}\nabla e_t(\boldsymbol{\theta}^*)'\nabla e_t(\boldsymbol{\theta}^*)\right)\right.\right.$$
$$\left.\left.\cdot\left(\left(\frac{F_t(v_t(\boldsymbol{\theta}^*))}{\alpha}-1\right)v_t(\boldsymbol{\theta}^*) - \frac{1}{\alpha}\mathbb{E}[Y_t\mathbf{1}\{Y_t \le v_t(\boldsymbol{\theta}^*)\}|\mathcal{F}_{t-1}] + e_t(\boldsymbol{\theta}^*)\right)\right]\right\|$$

$$\le\mathbb{E}[\| \left(H_2(\mathcal{F}_{t-1})H^2 + H_1(\mathcal{F}_{t-1})\cdot 2H^3\cdot H_1(\mathcal{F}_{t-1})\right) \tag{16}$$
$$\cdot\left(\left(F_t\left(v_t(\boldsymbol{\theta}^{**})\right)/\alpha - 1\right)\nabla v_t(\boldsymbol{\theta}^{**}) + \nabla e_t(\boldsymbol{\theta}^{**})\right)(\boldsymbol{\theta}^* - \boldsymbol{\theta}^0)\|]$$
$$\le \{(1/\alpha+1)(H^2\mathbb{E}[V_1(\mathcal{F}_{t-1})H_2(\mathcal{F}_{t-1})] + 2H^3\mathbb{E}[V_1(\mathcal{F}_{t-1})H_1(\mathcal{F}_{t-1})^2])$$
$$+ (H^2\cdot\mathbb{E}[H_1(\mathcal{F}_{t-1})H_2(\mathcal{F}_{t-1})] + 2H^3\mathbb{E}[H_1(\mathcal{F}_{t-1})^3])\}\|\boldsymbol{\theta}^* - \boldsymbol{\theta}^0\|$$

Third term:

$$\left\|\mathbb{E}\left[\frac{f_t(v_t(\boldsymbol{\theta}^*))}{-e_t(\theta^*)\alpha}\nabla v_t(\boldsymbol{\theta}^*)'\nabla v_t(\boldsymbol{\theta}^*) - \frac{f_t(v_t(\boldsymbol{\theta}^0))}{-e_t(\boldsymbol{\theta}^0)\alpha}\nabla v_t(\theta^0)'\nabla v_t(\boldsymbol{\theta}^0)\right]\right\|$$

$$= \frac{1}{\alpha}\left\|T^{-1}\sum_{t=1}^T\mathbb{E}\{\frac{f_t(v_t(\boldsymbol{\theta}^*))}{-e_t(\boldsymbol{\theta}^*)}\nabla v_t(\boldsymbol{\theta}^*)'\nabla v_t(\boldsymbol{\theta}^*) - \frac{f_t(v_t(\boldsymbol{\theta}^*))}{-e_t(\boldsymbol{\theta}^*)}\nabla v_t(\boldsymbol{\theta}^0)'\nabla v_t(\boldsymbol{\theta}^*)\right. \tag{17}$$
$$+ \frac{f_t(v_t(\boldsymbol{\theta}^*))}{-e_t(\boldsymbol{\theta}^*)}\nabla v_t(\boldsymbol{\theta}^0)'\nabla v_t(\boldsymbol{\theta}^*) - \frac{f_t(v_t(\boldsymbol{\theta}^0))}{-e_t(\boldsymbol{\theta}^*)}\nabla v_t(\boldsymbol{\theta}^0)'\nabla v_t(\boldsymbol{\theta}^*)$$
$$+ \frac{f_t(v_t(\boldsymbol{\theta}^0))}{-e_t(\boldsymbol{\theta}^*)}\nabla v_t(\boldsymbol{\theta}^0)'\nabla v_t(\boldsymbol{\theta}^*) - \frac{f_t(v_t(\boldsymbol{\theta}^0))}{-e_t(\boldsymbol{\theta}^0)}\nabla v_t(\boldsymbol{\theta}^0)'\nabla v_t(\boldsymbol{\theta}^*)$$
$$\left.+\frac{f_t(v_t(\boldsymbol{\theta}^0))}{-e_t(\boldsymbol{\theta}^0)}\nabla v_t(\boldsymbol{\theta}^0)'\nabla v_t(\boldsymbol{\theta}^*) - \frac{f_t(v_t(\boldsymbol{\theta}^0))}{-e_t(\boldsymbol{\theta}^0)}\nabla v_t(\boldsymbol{\theta}^0)'\nabla v_t(\boldsymbol{\theta}^0)\}\right\|$$
$$= \frac{1}{\alpha}\left\|T^{-1}\sum_{t=1}^T\mathbb{E}\{\frac{f_t(v_t(\boldsymbol{\theta}^*))}{-e_t(\boldsymbol{\theta}^*)}[\nabla^2 v_t(\boldsymbol{\theta}^{**})(\boldsymbol{\theta}^* - \boldsymbol{\theta}^0)]\nabla v_t(\theta^*)\right.$$
$$+ \frac{f_t(v_t(\boldsymbol{\theta}^*)) - f_t(v_t(\boldsymbol{\theta}^0))}{-e_t(\boldsymbol{\theta}^*)}\nabla v_t(\boldsymbol{\theta}^0)'\nabla v_t(\boldsymbol{\theta}^*)$$
$$+ \frac{f_t(v_t(\boldsymbol{\theta}^0))}{e_t(\boldsymbol{\theta}^{**})^2}(\boldsymbol{\theta}^* - \boldsymbol{\theta}^0)\nabla v_t(\boldsymbol{\theta}^0)'\nabla v_t(\boldsymbol{\theta}^*)$$
$$\left.+\frac{f_t(v_t(\boldsymbol{\theta}^0))}{-e_t(\boldsymbol{\theta}^0)}\nabla v_t(\boldsymbol{\theta}^0)'(\boldsymbol{\theta}^* - \boldsymbol{\theta}^0)^2 v_t(\boldsymbol{\theta}^{**})\}\right\|$$
$$\le \frac{1}{\alpha}T^{-1}\sum_{t=1}^T\mathbb{E}\{V_2(\mathcal{F}_{t-1})\left(KH\cdot V_1(\mathcal{F}_{t-1})\right) + KH\cdot V_1(\mathcal{F}_{t-1})^3$$
$$+ KH^2 H_1(\mathcal{F}_{t-1})V_1(\mathcal{F}_{t-1})^2 + KHV_1(\mathcal{F}_{t-1})V_2(\mathcal{F}_{t-1})\}\cdot\|\boldsymbol{\theta}^* - \boldsymbol{\theta}^0\|$$

Fourth term: The bound on this term follows similar steps to that of the third term:

$$\begin{aligned}
&\left\| T^{-1}\sum_{t=1}^{T}\mathbb{E}\{\frac{1}{e_t(\boldsymbol{\theta}^*)^2}\nabla e_t(\boldsymbol{\theta}^*)'\nabla e_t(\boldsymbol{\theta}^*) - \frac{1}{e_t(\boldsymbol{\theta}^0)^2}\nabla e_t(\boldsymbol{\theta}^0)'\nabla e_t(\boldsymbol{\theta}^0)\}\right\| \\
= \quad &\left\| T^{-1}\sum_{t=1}^{T}\mathbb{E}\{\frac{1}{e_t(\boldsymbol{\theta}^*)^2}\nabla e_t(\boldsymbol{\theta}^*)'\nabla e_t(\boldsymbol{\theta}^*) - \frac{1}{e_t(\boldsymbol{\theta}^*)^2}\nabla e_t(\boldsymbol{\theta}^0)'\nabla e_t(\boldsymbol{\theta}^*) \right. \qquad (18) \\
&+\frac{1}{e_t(\boldsymbol{\theta}^*)^2}\nabla e_t(\boldsymbol{\theta}^0)'\nabla e_t(\boldsymbol{\theta}^*) - \frac{1}{e_t(\boldsymbol{\theta}^0)^2}\nabla e_t(\boldsymbol{\theta}^0)'\nabla e_t(\boldsymbol{\theta}^*) \\
&\left.+\frac{1}{e_t(\boldsymbol{\theta}^0)^2}\nabla e_t(\boldsymbol{\theta}^0)'\nabla e_t(\boldsymbol{\theta}^*) - \frac{1}{e_t(\boldsymbol{\theta}^0)^2}\nabla e_t(\boldsymbol{\theta}^0)'\nabla e_t(\boldsymbol{\theta}^0)\}\right\| | \\
\leq \quad & T^{-1}\sum_{t=1}^{T}\{H^2\cdot\mathbb{E}[H_1(\mathcal{F}_{t-1})H_2(\mathcal{F}_{t-1})] + 2H^3\mathbb{E}\left[H_1(\mathcal{F}_{t-1})^3\right] + H^2\mathbb{E}[H_1(\mathcal{F}_{t-1})H_2(\mathcal{F}_{t-1})]\}\|\boldsymbol{\theta}^* - \boldsymbol{\theta}^0\|
\end{aligned}$$

Therefore, $\Lambda_T(\boldsymbol{\theta}^*) = D_T + O(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\|) \Rightarrow \|\Lambda_T(\boldsymbol{\theta}^*) - D_T\| \leq K\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\|$,where K is some constant $< \infty$, for T sufficiently large. By Assumption 2(E), D_T has eigenvalues bounded below by a positive constant, denoted as a, for T sufficiently large. Thus,

$$\begin{aligned}
\|\lambda_T(\hat{\boldsymbol{\theta}})\| &= \|\Lambda_T(\boldsymbol{\theta}^*)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right)\| \\
&= \|D_T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) - (D_T - \Lambda_T(\boldsymbol{\theta}^*))(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)\| \qquad (19) \\
&\geq \|D_T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)\| - \|(D_T - \Lambda_T(\boldsymbol{\theta}^*))(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)\| \\
&\geq (a - K\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\|)\cdot\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\|
\end{aligned}$$

The penultimate inequality holds by the triangle inequality, and the final inequality follows from Assumption 2(E) on the minimum eigenvalue of D_T. Thus, for T sufficiently large so that $a - K\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\| > 0$, the result follows. ■

Lemma 4 *Define*

$$\mu_t(\boldsymbol{\theta}, d) = \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d}\|g_t(\boldsymbol{\tau}) - g_t(\boldsymbol{\theta})\| \qquad (20)$$

Then under assumptions 1-2, Assumption N3(ii) of Weiss (1991) holds

$$\mathbb{E}[\mu_t(\boldsymbol{\theta}, d)] \leq bd, \text{ for } \|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| + d \leq d_0, d \geq 0 \qquad (21)$$

for T sufficiently large, where b, d,and d_0 are strictly positive numbers.

Proof of Lemma 4. In this proof, the strictly positive constant c and the mean-value expansion term, $\boldsymbol{\tau}^*$, can change from line to line. Pick d_0 such that for any $\boldsymbol{\theta}$ that satisfies

$\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\| \leq d_0$, all the conditions in Assumption 2(C) and 2(D) hold as well as $e_t(\boldsymbol{\theta}) \leq v_t(\boldsymbol{\theta}) \leq 0$. Let us expand $g_t(\boldsymbol{\theta})$ into six terms:

$$g_t(\boldsymbol{\theta}) = \frac{1}{\alpha}\frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})}\mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} + \frac{1}{\alpha}\frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2}\mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\} - \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} - \frac{1}{\alpha}\frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2}\mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\}Y_t + \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})} \tag{22}$$

We will bound $\mu_t(\boldsymbol{\theta}, d)$ by considering six terms, $\mu_t(\boldsymbol{\theta}, d)^{(i)}, i = 1, 2, \cdots, 6$, defined below. Each term is shown to be bounded by a constant times d.

First term:

$$\mu_t(\boldsymbol{\theta}, d)^{(1)} = \frac{1}{\alpha}\sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d}\left\|\frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})}\mathbf{1}\{Y_t \leq v_t(\boldsymbol{\tau})\} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})}\mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\}\right\| \tag{23}$$

Set $\boldsymbol{\tau}_1 = \arg\min_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} v_t(\boldsymbol{\tau})$ and $\boldsymbol{\tau}_2 = \arg\max_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} v_t(\boldsymbol{\tau})$. Since $v_t(\boldsymbol{\theta})$ and $e_t(\boldsymbol{\theta})$ are assumed to be twice continously differentiable, $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ exist. We want to take the indicator function out from the 'sup' operator. To this end, let us discuss what $\alpha \cdot \mu_t(\boldsymbol{\theta}, d)^{(1)}$ equals in two cases.

Case 1: $Y_t \leq v_t(\boldsymbol{\theta})$. (a) If $Y_t > v_t(\boldsymbol{\tau}_2)$, $\alpha \cdot \mu_t(\boldsymbol{\theta}, d)^{(1)} = \left\|\frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})}\right\|$. (b) If $Y_t < v_t(\boldsymbol{\tau}_1)$, $\alpha \cdot \mu_t(\boldsymbol{\theta}, d)^{(1)} = \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d}\left\|\frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})}\right\|$. (c) If $v_t(\boldsymbol{\tau}_1) \leq Y_t \leq v_t(\boldsymbol{\tau}_2)$,

$$\begin{aligned}\alpha \cdot \mu_t(\boldsymbol{\theta}, d)^{(1)} &= \max\left\{\sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d, Y_t \leq v(\boldsymbol{\tau})}\left\|\frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})}\right\|, \left\|\frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})}\right\|\right\} \\ &\leq \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d}\left\|\frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})}\right\| + \left\|\frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})}\right\|\end{aligned} \tag{24}$$

Case 2: $Y_t > v_t(\boldsymbol{\theta})$,

$$\begin{aligned}\alpha \cdot \mu_t(\boldsymbol{\theta}, d)^{(1)} &= \mathbf{1}\{Y_t \leq v(\boldsymbol{\tau}_2)\} \cdot \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d, Y_t \leq v(\boldsymbol{\tau})}\left\|\frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})}\right\| \\ &\leq \mathbf{1}\{Y_t \leq v(\boldsymbol{\tau}_2)\} \cdot \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d}\left\|\frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})}\right\|\end{aligned} \tag{25}$$

$\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\| + d \leq d_0$ implies that both $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ (which are in a d-neighborhood of $\boldsymbol{\theta}$) are in a d_0-neighborhood of $\boldsymbol{\theta}_0$, and so

$$\left\|\frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})}\right\| \leq \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d}\left\|\frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})}\right\| \leq \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\|\leq d_0}\left\|\frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})}\right\| \tag{26}$$

Thus,

$$
\begin{aligned}
&\alpha \cdot \mu_t(\boldsymbol{\theta}, d)^{(1)} \\
&\leq (\mathbf{1}\{v_t(\boldsymbol{\tau}_2) < Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\tau}_1) \leq Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\theta}) < Y_t \leq v_t(\boldsymbol{\tau}_2)\}) \qquad (27)\\
&\quad \cdot \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\| \leq d_0} \left\| \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| + \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\|,
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{E}_{t-1}[\mathbf{1}\{v_t(\boldsymbol{\tau}_2) < Y_t \leq v_t(\boldsymbol{\theta})\}] = &\int_{v_t(\boldsymbol{\tau}_2)}^{v_t(\boldsymbol{\theta})} f_t(y) dy \qquad (28)\\
&\leq K|v_t(\boldsymbol{\tau}_2) - v_t(\boldsymbol{\theta})| \leq KV_1(\mathcal{F}_{t-1})\|\boldsymbol{\tau}_2 - \boldsymbol{\theta}\| \leq KV_1(\mathcal{F}_{t-1})d
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{1}\{v_t(\boldsymbol{\theta}) < Y_t \leq v_t(\boldsymbol{\tau}_2)\} | \mathcal{F}_{t-1}\right] &\leq KV_1(\mathcal{F}_{t-1})d \qquad (29)\\
\text{and } \mathbb{E}\left[\mathbf{1}\{v_t(\boldsymbol{\tau}_1) < Y_t \leq v_t(\boldsymbol{\theta})\} | \mathcal{F}_{t-1}\right] &\leq KV_1(\mathcal{F}_{t-1})d
\end{aligned}
$$

Further

$$
\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\| \leq d} \left\| \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| \leq HV_1(\mathcal{F}_{t-1}) \qquad (30)
$$

and by the mean-value theorem,

$$
\frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} = \left\| \frac{\nabla^2 v_t(\boldsymbol{\tau}^*)}{-e_t(\boldsymbol{\tau}^*)} + \frac{\nabla' v_t(\boldsymbol{\tau}^*) \nabla e_t(\boldsymbol{\tau}^*)}{e_t(\boldsymbol{\tau}^*)^2} \right\| \cdot (\boldsymbol{\tau} - \boldsymbol{\theta}) \qquad (31)
$$

$$
\Rightarrow \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| \leq \left(HV_2(\mathcal{F}_{t-1}) + H^2 V_1(\mathcal{F}_{t-1}) H_1(\mathcal{F}_{t-1}) \right) \cdot d. \qquad (32)
$$

By Assumption 2(D), $\mathbb{E}[V_2(\mathcal{F}_{t-1})]$ and $\mathbb{E}[V_1(\mathcal{F}_{t-1})H_1(\mathcal{F}_{t-1})]$ are finite, so $\mathbb{E}[\mu_t(\boldsymbol{\theta}, d)^{(1)}] \leq cd$, where c is a strictly positive constant.

Second term: $\mu_t(\boldsymbol{\theta}, d)^{(2)} = \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\|$. It was shown in the derivations for the first term that $\mathbb{E}[\mu_t(\boldsymbol{\theta}, d)^{(2)}] \leq cd$, where c is a strictly positive constant.

Third term:

$$
\mu_t(\boldsymbol{\theta}, d)^{(3)} = \frac{1}{\alpha} \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{v_t(\boldsymbol{\tau}) \nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} \mathbf{1}\{Y_t \leq v_t(\boldsymbol{\tau})\} - \frac{v_t(\boldsymbol{\theta}) \nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\} \right\| \qquad (33)
$$

Similar to the first term, $\alpha \cdot \mu_t(\boldsymbol{\theta}, d)^{(3)}$ can be bounded by

$$\begin{aligned}
&\left(\mathbf{1}\{v_t(\boldsymbol{\tau}_2) < Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\tau}_1) \leq Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\theta}) < Y_t \leq v_t(\boldsymbol{\tau}_2)\}\right) \\
&\cdot \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\| \leq d_0} \left\| \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| + \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{v_t(\boldsymbol{\tau})\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\|
\end{aligned} \tag{34}$$

where

$$\mathbb{E}\left[\mathbf{1}\{v_t(\boldsymbol{\tau}_2) < Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\tau}_1) \leq Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\theta}) < Y_t \leq v_t(\boldsymbol{\tau}_2)\} | \mathcal{F}_{t-1}\right] \leq 3KV_1(\mathcal{F}_{t-1})d \tag{35}$$

and

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\| \leq d} \left\| \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \leq H \cdot H_1(\mathcal{F}_{t-1}) \tag{36}$$

where $e_t(\boldsymbol{\theta}) \leq v_t(\boldsymbol{\theta}) \leq 0$ is used, and by the mean-value theorem,

$$\begin{aligned}
&\frac{v_t(\boldsymbol{\tau})\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \\
=& \left\| \frac{\nabla' e_t(\boldsymbol{\tau}^*)\nabla v_t(\boldsymbol{\tau}^*)}{e_t(\boldsymbol{\tau}^*)^2} - \frac{2v_t(\boldsymbol{\tau}^*)\nabla' e_t(\boldsymbol{\tau}^*)\nabla e_t(\boldsymbol{\tau}^*)}{e_t(\boldsymbol{\tau}^*)^3} + \frac{v_t(\boldsymbol{\tau}^*)\nabla^2 e_t(\boldsymbol{\tau}^*)}{e_t(\boldsymbol{\tau}^*)^2} \right\| \cdot (\boldsymbol{\tau} - \boldsymbol{\theta})
\end{aligned} \tag{37}$$

$$\begin{aligned}
&\Rightarrow \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{v_t(\boldsymbol{\tau})\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \\
&\leq \left(H^2 V_1(\mathcal{F}_{t-1}) H_1(\mathcal{F}_{t-1}) + 2H^2 H_1(\mathcal{F}_{t-1})^2 + H \cdot H_2(\mathcal{F}_{t-1})\right) \cdot d
\end{aligned} \tag{38}$$

By Assumption 2(D), $\mathbb{E}[V_1(\mathcal{F}_{t-1})H_1(\mathcal{F}_{t-1})]$, $\mathbb{E}[H_1(\mathcal{F}_{t-1})^2]$, $\mathbb{E}[H_2(\mathcal{F}_{t-1})] < \infty$. Therefore, $\mathbb{E}[\mu_t(\boldsymbol{\theta}, d)^{(3)}] \leq cd$, where c is a strictly positive constant.

Fourth term: $\mu_t(\boldsymbol{\theta}, d)^{(4)} = \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{v_t(\boldsymbol{\tau})\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\|$. In the derivations for the third term we showed that $\mathbb{E}[\mu_t(\boldsymbol{\theta}, d)^{(4)}] \leq cd$, where c is a strictly positive constant.

Fifth term:

$$\mu_t(\boldsymbol{\theta}, d)^{(5)} = \frac{1}{\alpha} \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} \mathbf{1}\{Y_t \leq v_t(\boldsymbol{\tau})\} Y_t - \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\} Y_t \right\| \tag{39}$$

Similar to the first term, $\alpha \cdot \mu_t(\boldsymbol{\theta}, d)^{(5)}$ can be bounded by

$$\begin{aligned}
&\left(\mathbf{1}\{v_t(\boldsymbol{\tau}_2) < Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\tau}_1) \leq Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\theta}) < Y_t \leq v_t(\boldsymbol{\tau}_2)\}\right) \\
&\cdot |Y_t| \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\| \leq d_0} \left\| \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| + |Y_t| \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\|
\end{aligned} \tag{40}$$

where

$$\mathbb{E}\left[\mathbf{1}\{v_t(\boldsymbol{\tau}_2) < Y_t \le v_t(\boldsymbol{\theta})\}|Y_t|\ |\mathcal{F}_{t-1}\right] = \int_{v_t(\boldsymbol{\tau}_2)}^{v_t(\boldsymbol{\theta})} |y|f_t(y)dy \le K|v_t(\boldsymbol{\tau}_2)| \cdot |v_t(\boldsymbol{\tau}_2) - v_t(\boldsymbol{\theta})| \tag{41}$$
$$\le KV(\mathcal{F}_{t-1})V_1(\mathcal{F}_{t-1})\|\boldsymbol{\tau}_2 - \boldsymbol{\theta}\| \le KV(\mathcal{F}_{t-1})V_1(\mathcal{F}_{t-1})d$$

and similarly,

$$\mathbb{E}\left[\mathbf{1}\{v_t(\boldsymbol{\tau}_1) < Y_t \le v_t(\boldsymbol{\theta})\}|Y_t|\ |\mathcal{F}_{t-1}\right] \le KV(\mathcal{F}_{t-1})V_1(\mathcal{F}_{t-1})d \tag{42}$$
$$\text{and} \quad \mathbb{E}\left[\mathbf{1}\{v_t(\boldsymbol{\theta}) < Y_t \le v_t(\boldsymbol{\tau}_2)\}|Y_t|\ |\mathcal{F}_{t-1}\right] \le KV(\mathcal{F}_{t-1})V_1(\mathcal{F}_{t-1})d$$

Further

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\|\le d} \left\| \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \le H^2 H_1(\mathcal{F}_{t-1}) \tag{43}$$

and by the mean-value theorem,

$$\frac{\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} = \left\| -\frac{2\nabla' e_t(\boldsymbol{\tau}^*)\nabla e_t(\boldsymbol{\tau}^*)}{e_t(\boldsymbol{\tau}^*)^3} + \frac{\nabla^2 e_t(\boldsymbol{\tau}^*)}{e_t(\boldsymbol{\tau}^*)^2} \right\| \cdot (\boldsymbol{\tau} - \boldsymbol{\theta}) \tag{44}$$
$$\Rightarrow \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\le d} \left\| \frac{\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \le \left(2H^3 H_1(\mathcal{F}_{t-1})^2 + H^2 H_2(\mathcal{F}_{t-1})\right) \cdot d$$

By Assumption 2(D), $\mathbb{E}[V(\mathcal{F}_{t-1})V_1(\mathcal{F}_{t-1})H_1(\mathcal{F}_{t-1})]$, $\mathbb{E}[H_1(\mathcal{F}_{t-1})^2|Y_t|]$, $\mathbb{E}[H_2(\mathcal{F}_{t-1})|Y_t|] < \infty$. Therefore, $\mathbb{E}[\mu_t(\boldsymbol{\theta}, d)^{(5)}] \le cd$, where c is a strictly positive constant.

Sixth term:

$$\mu_t^{(6)}(\boldsymbol{\theta}, d) = \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\le d} \left\| \frac{\nabla' e_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' e_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| \tag{45}$$

By the mean-value theorem,

$$\frac{\nabla' e_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' e_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} = \left\| \frac{\nabla' e_t(\boldsymbol{\tau}^*)\nabla e_t(\boldsymbol{\tau}^*)}{e_t(\boldsymbol{\tau}^*)^2} + \frac{\nabla^2 e_t(\boldsymbol{\tau}^*)}{-e_t(\boldsymbol{\tau}^*)} \right\| \cdot (\boldsymbol{\tau} - \boldsymbol{\theta}) \tag{46}$$

$$\Rightarrow \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\le d} \left\| \frac{\nabla' e_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' e_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| \le \left(H^2 H_1(\mathcal{F}_{t-1})^2 + H \cdot H_2(\mathcal{F}_{t-1})\right) \cdot d. \tag{47}$$

By Assumption 2(D), $\mathbb{E}[H_1(\mathcal{F}_{t-1})^2]$, $\mathbb{E}[H_2(\mathcal{F}_{t-1})] < \infty$. Therefore, $\mathbb{E}[\mu_t(\boldsymbol{\theta}, d)^{(6)}] \le cd$, where c is a strictly positive constant.

Thus we have shown that $\mu_t(\boldsymbol{\theta}, d) \le \sum_{i=1}^{6} \mu_t(\boldsymbol{\theta}, d)^{(i)}$ with $\mathbb{E}[\mu_t(\boldsymbol{\theta}, d)^{(i)}] \le cd$, $\forall i = 1, 2, \cdots, 6$, where c is a strictly positive constant, proving the lemma. ■

Lemma 5 *Under Assumptions 1-2, Assumption N3(iii) of Weiss (1991) holds:*

$$\mathbb{E}[\mu_t(\boldsymbol{\theta}, d)^q] \leq cd, \text{ for } \|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| + d \leq d_0, \text{ and some } q > 2$$

for T sufficiently large, and where $c > 0$, $d \geq 0$ and $d_0 > 0$.

Proof of Lemma 5. In this proof, the strictly positive constant c and the mean-value expansion term, $\boldsymbol{\tau}^*$, can change from line to line. Pick d_0 such that for any $\boldsymbol{\theta}$ that satisfies $\|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| \leq d_0$, all the conditions in Assumption 2(C) and 2(D) hold as well as $e_t(\boldsymbol{\theta}) \leq v_t(\boldsymbol{\theta}) \leq 0$. Similar to Lemma 4, we will decompose $\mu_t(\boldsymbol{\theta}, d)$ into six terms, $\mu_t(\boldsymbol{\theta}, d)^{(i)}$, for $i = 1, 2, ..., 6$. By Jensen's inequality, $\mathbb{E}\left[\mu_t(\boldsymbol{\theta}, d)^q\right] \leq 6^{q-1} \sum_{i=1}^{6} \mathbb{E}[\left(\mu_t(\boldsymbol{\theta}, d)^{(i)}\right)^q]$, $q > 2$. We will show that for some $0 < \delta < 1$, $\mathbb{E}[\left(\mu_t(\boldsymbol{\theta}, d)^{(i)}\right)^{2+\delta}] \leq cd$, $\forall\ i = 1, 2, \cdots, 6$, where c is a strictly positive constant.

First term:

$$\mu_t(\boldsymbol{\theta}, d)^{(1)} = \frac{1}{\alpha} \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} \mathbf{1}\{Y_t \leq v_t(\boldsymbol{\tau})\} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\} \right\| \tag{48}$$

Set $\boldsymbol{\tau}_1 = \arg\min_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} v_t(\boldsymbol{\tau})$ and $\boldsymbol{\tau}_2 = \arg\max_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} v_t(\boldsymbol{\tau})$. Following the same argument as in the proof of Lemma 4, we obtain

$$\begin{aligned}[\alpha \cdot \mu_t(\boldsymbol{\theta}, d)^{(1)}]^{2+\delta} \leq\ & \left(\mathbf{1}\{v_t(\boldsymbol{\tau}_2) < Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\tau}_1) \leq Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\theta}) < Y_t \leq v_t(\boldsymbol{\tau}_2)\}\right) \\ & \cdot \left(\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\| \leq d_0} \left\| \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| \right)^{2+\delta} + \left(\sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| \right)^{2+\delta}\end{aligned} \tag{49}$$

where

$$\mathbb{E}_{t-1}\left[\mathbf{1}\{v_t(\boldsymbol{\tau}_2) < Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\tau}_1) \leq Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\theta}) < Y_t \leq v_t(\boldsymbol{\tau}_2)\}\right] \leq 3KV_1(\mathcal{F}_{t-1})d \tag{50}$$

and

$$\left(\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\| \leq d} \left\| \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| \right)^{2+\delta} \leq (HV_1(\mathcal{F}_{t-1}))^{2+\delta} \tag{51}$$

For $\left(\sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| \right)^{2+\delta}$, we need to combine the two following two results:

$$\sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| \leq \left(HV_2(\mathcal{F}_{t-1}) + H^2 V_1(\mathcal{F}_{t-1}) H_1(\mathcal{F}_{t-1})\right) d \tag{52}$$

$$\left(\sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\| \leq d} \left\| \frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| \right)^{1+\delta} \leq (2HV_1(\mathcal{F}_{t-1}))^{1+\delta}$$

Combining with Assumption 2(D), we thus have $\mathbb{E}[\big(\mu_t(\boldsymbol{\theta},d)^{(1)}\big)^{2+\delta}] \leq cd$.

Second term: $\mu_t(\boldsymbol{\theta},d)^{(2)} = \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{\nabla' v_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' v_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\|$. It was shown in the derivations for the first term that $\mathbb{E}[\big(\mu_t(\boldsymbol{\theta},d)^{(2)}\big)^{2+\delta}] \leq cd$.

Third term:

$$\mu_t(\boldsymbol{\theta},d)^{(3)} = \frac{1}{\alpha} \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{v_t(\boldsymbol{\tau})\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} \mathbf{1}\{Y_t \leq v_t(\boldsymbol{\tau})\} - \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\} \right\| \tag{53}$$

Similar to the first term, $\big(\alpha \cdot \mu_t(\boldsymbol{\theta},d)^{(3)}\big)^{2+\delta}$ can be bounded by

$$\begin{aligned} &\left(\mathbf{1}\{v_t(\boldsymbol{\tau}_2) < Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\tau}_1) \leq Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\theta}) < Y_t \leq v_t(\boldsymbol{\tau}_2)\}\right) \\ &\cdot \left(\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\|\leq d_0} \left\| \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \right)^{2+\delta} + \left(\sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{v_t(\boldsymbol{\tau})\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \right)^{2+\delta} \end{aligned} \tag{54}$$

where

$$\mathbb{E}_{t-1}\left(\mathbf{1}\{v_t(\boldsymbol{\tau}_2) < Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\tau}_1) \leq Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\theta}) < Y_t \leq v_t(\boldsymbol{\tau}_2)\}\right) \leq 3KV_1(\mathcal{F}_{t-1})d \tag{55}$$

and

$$\left(\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\|\leq d} \left\| \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \right)^{2+\delta} \leq (H \cdot H_1(\mathcal{F}_{t-1}))^{2+\delta} \tag{56}$$

For $\left(\sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{v_t(\boldsymbol{\tau})\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \right)^{2+\delta}$, we need to combine the following two results:

$$\sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{v_t(\boldsymbol{\tau})\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \leq \left(H^2 V_1(\mathcal{F}_{t-1}) H_1(\mathcal{F}_{t-1}) + 2H^2 H_1(\mathcal{F}_{t-1})^2 + H \cdot H_2(\mathcal{F}_{t-1})\right) d \tag{57}$$

$$\left(\sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{v_t(\boldsymbol{\tau})\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \right)^{1+\delta} \leq (2H \cdot H_1(\mathcal{F}_{t-1}))^{1+\delta} \tag{58}$$

Combining with Assumption 2(D), we thus have $\mathbb{E}[\big(\mu_t(\boldsymbol{\theta},d)^{(3)}\big)^{2+\delta}] \leq cd$.

Fourth term: $\mu_t(\boldsymbol{\theta},d)^{(4)} = \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{v_t(\boldsymbol{\tau})\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{v_t(\boldsymbol{\theta})\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\|$. It was shown in the derivations for the third term that $\mathbb{E}[\big(\mu_t(\boldsymbol{\theta},d)^{(4)}\big)^{2+\delta}] \leq cd$.

Fifth term:

$$\mu_t(\boldsymbol{\theta}, d)^{(5)} = \frac{1}{\alpha} \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} \mathbf{1}\{Y_t \leq v_t(\boldsymbol{\tau})\} Y_t - \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\} Y_t \right\| \tag{59}$$

Similar to the first and third terms, $\left(\alpha \cdot \mu_t(\boldsymbol{\theta}, d)^{(5)}\right)^{2+\delta}$ can be bounded by

$$\begin{aligned} &(\mathbf{1}\{v_t(\boldsymbol{\tau}_2) < Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\tau}_1) \leq Y_t \leq v_t(\boldsymbol{\theta})\} + \mathbf{1}\{v_t(\boldsymbol{\theta}) < Y_t \leq v_t(\boldsymbol{\tau}_2)\}) \\ &\cdot |Y_t|^{2+\delta} \left(\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\|\leq d_0} \left\| \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \right)^{2+\delta} + |Y_t|^{2+\delta} \left(\sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \right)^{2+\delta} \end{aligned} \tag{60}$$

where

$$\begin{aligned} \mathbb{E}_{t-1}[\mathbf{1}\{v_t(\boldsymbol{\tau}_2) < Y_t \leq v_t(\boldsymbol{\theta})\}|Y_t|^{2+\delta}] &= \int_{v_t(\boldsymbol{\tau}_2)}^{v_t(\boldsymbol{\theta})} |y|^{2+\delta} f_t(y) dy \leq K|v_t(\boldsymbol{\tau}_2)|^{2+\delta} \cdot |v_t(\boldsymbol{\tau}_2) - v_t(\boldsymbol{\theta})| \\ &\leq KV(\mathcal{F}_{t-1})^{2+\delta} V_1(\mathcal{F}_{t-1}) \|\boldsymbol{\tau}_2 - \boldsymbol{\theta}\| \leq KV(\mathcal{F}_{t-1})^{2+\delta} V_1(\mathcal{F}_{t-1}) d \end{aligned} \tag{61}$$

and similarly,

$$\begin{aligned} \mathbb{E}\left[\mathbf{1}\{v_t(\boldsymbol{\tau}_1) < Y_t \leq v_t(\boldsymbol{\theta})\}|Y_t|^{2+\delta} \ |\mathcal{F}_{t-1}\right] &\leq KV(\mathcal{F}_{t-1})^{2+\delta} V_1(\mathcal{F}_{t-1}) d \\ \text{and} \ \ \mathbb{E}\left[\mathbf{1}\{v_t(\boldsymbol{\theta}) < Y_t \leq v_t(\boldsymbol{\tau}_2)\}|Y_t|^{2+\delta} \ |\mathcal{F}_{t-1}\right] &\leq KV(\mathcal{F}_{t-1})^{2+\delta} V_1(\mathcal{F}_{t-1}) d \end{aligned} \tag{62}$$

Further

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\|\leq d} \left\| \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \leq H^2 H_1(\mathcal{F}_{t-1}) \tag{63}$$

For $\left(\sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \right)^{2+\delta}$, we need to combine the following two results:

$$\begin{aligned} \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{\nabla' e_t(\boldsymbol{\tau})}{e_t(\boldsymbol{\tau})^2} - \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| &\leq \left(2H^3 H_1(\mathcal{F}_{t-1})^2 + H^2 H_2(\mathcal{F}_{t-1})\right) d \\ \left(\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\|\leq d} \left\| \frac{\nabla' e_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})^2} \right\| \right)^{1+\delta} &\leq \left(2H^2 H_1(\mathcal{F}_{t-1})\right)^{1+\delta} \end{aligned} \tag{64}$$

Combining with Assumption 2(D), we thus have $\mathbb{E}[\left(\mu_t(\boldsymbol{\theta} d)^{(5)}\right)^{2+\delta}] \leq cd$.

Sixth term:

$$\mu_t^{(6)}(\boldsymbol{\theta}, d) = \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{\nabla' e_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' e_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| \tag{65}$$

We have

$$\begin{aligned} \sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{\nabla' e_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' e_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| &\leq \left(H^2 H_1(\mathcal{F}_{t-1})^2 + H H_2(\mathcal{F}_{t-1})\right) d \\ \left(\sup_{\|\boldsymbol{\tau}-\boldsymbol{\theta}\|\leq d} \left\| \frac{\nabla' e_t(\boldsymbol{\tau})}{-e_t(\boldsymbol{\tau})} - \frac{\nabla' e_t(\boldsymbol{\theta})}{-e_t(\boldsymbol{\theta})} \right\| \right)^{1+\delta} &\leq (2H H_1(\mathcal{F}_{t-1}))^{1+\delta} \end{aligned} \tag{66}$$

Combining with Assumption 2(D), we thus have $\mathbb{E}[\left(\mu_t(\boldsymbol{\theta}, d)^{(6)}\right)^{2+\delta}] \leq cd$. Thus $\mathbb{E}[\mu_t(\boldsymbol{\theta}, d)^{(i)}]^{2+\delta} \leq cd$, $\forall i = 1, 2, ..., 6$, proving the lemma. ■

Lemma 6 *Under Assumptions 1-2,* $E\|g_t(\boldsymbol{\theta}^0)\|^{2+\delta} \leq M$*, for all* t *and some* $M > 0$*.*

Proof of Lemma 6.

$$
\begin{aligned}
\mathbb{E}\|g_t(\boldsymbol{\theta}^0)\|^{2+\delta} \leq & 4^{1+\delta}\left\{\mathbb{E}\left\|\frac{\nabla' v_t(\boldsymbol{\theta}^0)}{-e_t(\boldsymbol{\theta}^0)}\left(\frac{1}{\alpha}\mathbf{1}\left\{Y_t \leq v_t(\boldsymbol{\theta}^0)\right\} - 1\right)\right\|^{2+\delta}\right. \\
& + \mathbb{E}\left\|\frac{v_t(\boldsymbol{\theta}^0)\nabla' e_t(\boldsymbol{\theta}^0)}{e_t(\boldsymbol{\theta}^0)^2}\left(\frac{1}{\alpha}\mathbf{1}\left\{Y_t \leq v_t(\boldsymbol{\theta}^0)\right\} - 1\right)\right\|^{2+\delta} \\
& + \mathbb{E}\left\|\frac{\nabla' e_t(\boldsymbol{\theta}^0)}{e_t(\boldsymbol{\theta}^0)^2}\frac{1}{\alpha}\mathbf{1}\left\{Y_t \leq v_t(\boldsymbol{\theta}^0)\right\} Y_t\right\|^{2+\delta} \\
& \left. + \mathbb{E}\left\|\frac{\nabla' e_t(\boldsymbol{\theta}^0)}{e_t(\boldsymbol{\theta}^0)}\right\|^{2+\delta}\right\} \\
\leq & 4^{1+\delta}\left\{(\frac{1}{\alpha} + 1)^{2+\delta} H^{2+\delta}\mathbb{E}\left[V_1(\mathcal{F}_{t-1})^{2+\delta}\right]\right. \\
& + (\frac{1}{\alpha} + 1)^{2+\delta} H^{2+\delta}\mathbb{E}\left[H_1(\mathcal{F}_{t-1})^{2+\delta}\right] \\
& + \frac{1}{\alpha^{2+\delta}} H^{4+2\delta}\mathbb{E}[H_1(\mathcal{F}_{t-1})^{2+\delta}\left|Y_t\right|^{2+\delta}] \\
& \left. + H^{2+\delta}\mathbb{E}\left[H_1(\mathcal{F}_{t-1})^{2+\delta}\right]\right\} \\
\leq & M
\end{aligned}
$$

since all the four expectations in the penultimate inequality are finite by assumption 2(D). Assumption N4 of Weiss (1991) only requires $E\|g_t(\boldsymbol{\theta}^0)\|^2 \leq M$, which is implied by the above. ■

Lemma 7 *Under Assumptions 1-2, we have* $T^{-1/2}\sum_{t=1}^{T} g_t(\boldsymbol{\theta}^{\mathbf{0}}) \xrightarrow{d} N\left(0, \mathbf{A}_0\right)$ *as* $T \to \infty$*, where* $\mathbf{A}_0 \equiv \mathbb{E}\left[g_t(\boldsymbol{\theta}^{\mathbf{0}})g_t(\boldsymbol{\theta}^{\mathbf{0}})'\right]$.

Proof of Lemma 7. First note that the sequence $\left\{g_t(\boldsymbol{\theta}^{\mathbf{0}})\right\}$ is stationary by Assumption 1(B)(ii), and has zero mean. Under Assumption 2(F) and Lemma 6, we can use Corollary 5.1 of Hall and Heyde (1980) and the Cramer-Wold device to obtain the result. ■

Appendix SA.2: Estimating a GARCH(1,1) model by FZ loss minimization

In this appendix we show that we can estimate the popular GARCH(1,1) model via FZ loss minimization. We then verify that the assumptions required to show this are implied by the Assumptions 1-2 in the main paper. Throughout, $||\mathbf{x}||$ refers to the Euclidean norm if $\mathbf{x}$ is a vector and to the Frobenius norm if $\mathbf{x}$ is a matrix.

Appendix SA.2.1: Model specification

Assume that the data generating process for Y_t is:

$$\begin{aligned} Y_t &= \sigma_t \eta_t, \quad \eta_t \perp\!\!\!\perp \sigma_t, \quad \eta_t \sim iid\ F_\eta(0,1) \\ \sigma_t^2 &= \omega_0 + \beta_0 \sigma_{t-1}^2 + \gamma_0 Y_{t-1}^2 \end{aligned} \tag{67}$$

Under this model, the conditional VaR and ES of Y_t at a probability level $\alpha \in (0,1)$, that is $VaR_\alpha(Y_t|\mathcal{F}_{t-1})$ and $ES_\alpha(Y_t|\mathcal{F}_{t-1})$, follow the dynamics:

$$\begin{bmatrix} VaR_\alpha(Y_t|\mathcal{F}_{t-1}) \\ ES_\alpha(Y_t|\mathcal{F}_{t-1}) \end{bmatrix} = \begin{bmatrix} c_0 \cdot ES_\alpha(Y_t|\mathcal{F}_{t-1}) \\ b_0 \cdot \sigma_t \end{bmatrix} \tag{68}$$

$$\begin{aligned} \text{where} \quad c_0 &\equiv F_\eta^{-1}(\alpha)/\mathbb{E}[\eta_t|\eta_t \leq F_\eta^{-1}(\alpha)] \\ b_0 &\equiv \mathbb{E}[\eta_t|\eta_t \leq F_\eta^{-1}(\alpha)] \end{aligned}$$

We fix the level $\alpha \in (0,1)$ throughout this appendix. Our goal is to estimate the parameter vector $\boldsymbol{\theta}^0 = [\beta_0, \gamma_0, b_0, c_0]$ by minimizing the FZ loss function. Note that the parameters do not include ω_0 because only two of the three parameters ω_0, b_0, γ_0 are identifiable under this model. A detailed discussion about the identification of the GARCH model via FZ loss minimization is provided in Section SA.2.3 of this appendix.

In the simulation study (Section 4 of the main paper), for estimating the GARCH model via FZ loss minimization, we fix ω at its true value ω_0. Put $\boldsymbol{\theta} = [\beta, \gamma, b, c]$ and its true value is

$\boldsymbol{\theta}^0 = [\beta_0, \gamma_0, b_0, c_0]$. We will estimate $\boldsymbol{\theta}^0$ by

$$\begin{aligned}
\boldsymbol{\theta}_T &\equiv \arg\min_{\boldsymbol{\theta}\in\Theta} \; L_T(\boldsymbol{\theta}) \\
\text{where} \;\; L_T(\boldsymbol{\theta}) &= \frac{1}{T}\sum_{t=1}^{T} L_{FZ0}(Y_t, v_t(\boldsymbol{\theta}), e_t(\boldsymbol{\theta}); \alpha) \\
\sigma_t^2(\boldsymbol{\theta}) &= \omega_0 + \beta\sigma_{t-1}^2(\boldsymbol{\theta}) + \gamma Y_{t-1}^2 \\
v_t(\boldsymbol{\theta}) &= c \cdot e_t(\boldsymbol{\theta}) \\
e_t(\boldsymbol{\theta}) &= b \cdot \sigma_t(\boldsymbol{\theta})
\end{aligned}$$

and the FZ loss function L_{FZ0} is defined in equation (6)

Appendix SA.2.2: Assumptions to estimate GARCH by FZ minimization

GARCH Assumption 1: F_η has zero mean, unit variance, finite fourth moment, and a unique α-quantile, which is non-positive. It has density $f_\eta(\cdot)$ that satisfies $f_\eta(\cdot) \le K$ and $|f_\eta(\lambda_1) - f_\eta(\lambda_2)| \le K|\lambda_1 - \lambda_2|$.

The distributions we often assume for the innovations of GARCH model, like the normal distribution or t-distribution with degrees of freedom greater than four, all satisfy this assumption.

GARCH Assumption 2: $0 < \omega_0 < \infty$. The true parameter vector $\boldsymbol{\theta}^0 = [\beta_0, \gamma_0, b_0, c_0] \in \Theta \in \mathbb{R}^4$ is in the interior of Θ, a compact and convex parameter space. Specifically, for any vector $[\beta, \gamma, b, c] \in \Theta$, assume that $\delta_1 \le \beta \le (1-\delta_1)$, $\delta_1 \le \gamma \le (1-\delta_1)$ for some constant $\delta_1 > 0$, $\delta_2 \le c \le (1-\delta_2)$, $-B_1 \le b \le -B_2$, for some constants $\delta_2, B_1, B_2 > 0$, and $(\beta+\gamma)^2 + \left(\mathbb{E}\left[\eta_t^4\right] - 1\right)\gamma^2 \le 1 - \delta_3$ for some constant $\delta_3 > 0$.

This assumption is similar to Assumption 1 of Lumsdaine (1996) with the exception of the third condition on the parameter vector, which is used to validate the mixing condition in Assumption 2(F), which we now discuss. It is not hard to show that

$$\sigma\left\{\left(Y_t, v_t(\theta), e_t(\theta), \nabla' v_t(\theta), \nabla' e_t(\theta)\right)\right\} \subset \sigma\left\{\left(Y_t, \sigma_t^2(\theta), \partial\sigma_t^2(\theta)/\beta\right)\right\}$$

and thus we need to consider the mixing properties of $\left(Y_t, \sigma_t^2(\theta), \partial\sigma_t^2(\theta)/\beta\right)$. Using Definition 3 of Carrasco and Chen (2002), $\left\{\left(Y_t, \sigma_t^2(\theta), \partial\sigma_t^2(\theta)/\beta\right), t \ge 0\right\}$ is a generalized hidden Markov model with a hidden chain $\left\{\left(\sigma_t^2(\theta), \partial\sigma_t^2(\theta)/\beta\right), t \ge 0\right\}$. By their Proposition 4, if $\left(\sigma_t^2(\theta), \partial\sigma_t^2(\theta)/\beta\right)$ is stationary and β-mixing then $\left(Y_t, \sigma_t^2(\theta), \partial\sigma_t^2(\theta)/\beta\right)$ is stationary and β-mixing with a decay rate at least as fast as that of $\left\{\left(\sigma_t^2(\theta), \partial\sigma_t^2(\theta)/\beta\right), t \ge 0\right\}$.

We use Proposition 3 of Carrasco and Chen (2002). First, we express $\left\{\left(\sigma_t^2(\theta), \partial\sigma_t^2(\theta)/\beta\right), t \geq 0\right\}$ in the polynomial random coefficient form:

$$\begin{pmatrix} \sigma_t^2(\theta) \\ \partial\sigma_t^2(\theta)/\beta \end{pmatrix} = \begin{pmatrix} \omega_0 \\ 0 \end{pmatrix} + \begin{pmatrix} \gamma\eta_{t-1}^2 + \beta & 0 \\ 1 & \beta \end{pmatrix} \begin{pmatrix} \sigma_{t-1}^2(\theta) \\ \partial\sigma_{t-1}^2(\theta)/\beta \end{pmatrix} \tag{69}$$

First, note that by GARCH Assumption 1 $\left\{\eta_t^2\right\}$ satisfies their condition (e) and that by GARCH Assumption 2, their Assumption A_0 is obviously satisfied. For Assumption A_1, the spectral radius of $\begin{pmatrix} \beta & 0 \\ 1 & \beta \end{pmatrix}$ is $\beta < 1$. For Assumption A_2', the spectral radius of $\begin{pmatrix} \gamma\eta_{t-1}^2 + \beta & 0 \\ 1 & \beta \end{pmatrix}$ is $\mathbb{E}\left[\left(\gamma\eta_{t-1}^2 + \beta\right)^2\right] = (\beta+\gamma)^2 + \left(\mathbb{E}\left[\eta_t^4\right] - 1\right)\gamma^2 < 1$.Then, if $\left(\sigma_t^2(\theta), \partial\sigma_t^2(\theta)/\beta\right)$ is initialized from the invariant distribution (which we did in our simulations) then $\left\{\left(\sigma_t^2(\theta), \partial\sigma_t^2(\theta)/\beta\right), t \geq 0\right\}$ is strictly stationary and β-mixing with exponential decay. It is well known that β-mixing implies α-mixing and so Assumption 2(F) of the paper is satisfied.

GARCH Assumptions 1–2 imply that the distribution and density of Y_t conditional on $\mathcal{F}_{t-1}$ satisfy Assumption 2(B)(i). Since $Y_t = \sigma_t\eta_t$ and $\sigma_t \in \mathcal{F}_{t-1}$,

$$\begin{aligned} F_t(x|\mathcal{F}_{t-1}) &= F_\eta\left(\frac{x}{\sigma_t}\right) \\ f_t(x|\mathcal{F}_{t-1}) &= \frac{1}{\sigma_t} f_\eta\left(\frac{x}{\sigma_t}\right) \end{aligned}$$

Thus,

$$|f_t(x|\mathcal{F}_{t-1})| \leq \frac{K}{\sqrt{\omega_0}}, \text{ since } \sigma_t^2 = \omega_0 + \beta\sigma_{t-1}^2 + \gamma y_{t-1}^2 \geq \omega_0 > 0$$

$$|f_t(\lambda_1|\mathcal{F}_{t-1}) - f_t(\lambda_2|\mathcal{F}_{t-1})| = \frac{1}{\sigma_t}|f_\eta(\frac{\lambda_1}{\sigma_t}) - f_\eta(\frac{\lambda_1}{\sigma_t})| \leq \frac{K}{\sigma_t^2}|\lambda_1 - \lambda_2| \leq \frac{K}{\omega_0}|\lambda_1 - \lambda_2|$$

GARCH Assumption 3: $\mathbb{E}|Y_t|^{5+\delta} < \infty$, for some $\delta > 0$.

GARCH Assumption 3 is needed to show the uniform LLN Assumption 1(A) of the paper and also to ensure the moment conditions in Assumptions 2 (C) and (D).

For the GARCH model it is possible to obtain the results of the paper under a weaker version of Assumption 2(D). An inspection of the proofs shows that it is sufficient to replace Assumption 2(D) by the following.

Assumption 2(D'): For some $0 < \delta < 1$ and $\forall\ t$:

(i) $\mathbb{E}\left[V_1(\mathcal{F}_{t-1})^{3+\delta}\right]$, $\mathbb{E}\left[H_1(\mathcal{F}_{t-1})^{3+\delta}\right]$, $\mathbb{E}\left[V_2(\mathcal{F}_{t-1})^{\frac{3+\delta}{2}}\right]$, $\mathbb{E}\left[H_2(\mathcal{F}_{t-1})^{\frac{3+\delta}{2}}\right] \leq K$,

(ii) $\mathbb{E}\left[V\left(\mathcal{F}_{t-1}\right)^{2+\delta}V_1\left(\mathcal{F}_{t-1}\right)^{1+\delta}\right] \leq K$,

(iii) $\mathbb{E}\left[H_1\left(\mathcal{F}_{t-1}\right)^{2+\delta}|Y_t|^{2+\delta}\right]$, $\mathbb{E}\left[H_2\left(\mathcal{F}_{t-1}\right)^{1+\delta}|Y_t|^{2+\delta}\right] \leq K$.

Assumption 2(D') is in turn fulfilled if $\mathbb{E}|Y_t|^{4+\delta} < \infty$, for some $\delta > 0$. For reasons of brevity, we omit the arguments and work instead with the stronger GARCH Assumption 3.

Appendix SA.2.3: Identification

In Theorem 1, we discussed the identification of a general dynamic model for ES and VaR model by minimizing the FZ loss, with the form of a general model given by equation (4). Under correct specification of the model, that is $(VaR_\alpha(Y_t|\mathcal{F}_{t-1}), ES_\alpha(Y_t|\mathcal{F}_{t-1})) = (v_t(\boldsymbol{\theta}^0), e_t(\boldsymbol{\theta}^0))\ \forall\, t$ a.s., the condition required for identification is given by Assumption 1(B) (iv): $\Pr\left[v_t(\boldsymbol{\theta}) = v_t(\boldsymbol{\theta}^0) \cap e_t(\boldsymbol{\theta}) = e_t(\boldsymbol{\theta}^0)\right] = 1$,$\forall\, t \Rightarrow \theta = \theta^0$. This assumption is equivalent to

$$\Pr\left[v_t(\boldsymbol{\theta}) = v_t(\boldsymbol{\theta}^0) \cap e_t(\boldsymbol{\theta}) = e_t(\boldsymbol{\theta}^0),\ \forall t\right] = 1$$

In the case of the GARCH model we have:

$$\begin{aligned}
&\Pr\left[\{v_t(\boldsymbol{\theta}) = v_t(\boldsymbol{\theta}^0)\} \cap \{e_t(\boldsymbol{\theta}) = e_t(\boldsymbol{\theta}^0)\},\quad \forall t\right] = 1\\
\Rightarrow &\Pr\left[\{c \cdot e_t(\boldsymbol{\theta}) = c_0 \cdot e_t(\boldsymbol{\theta}^0)\} \cap \{e_t(\boldsymbol{\theta}) = e_t(\boldsymbol{\theta}^0)\},\quad \forall t\right] = 1\\
\Rightarrow &\Pr\left[\{c = c_0\} \cap \{b \cdot \sigma_t(\boldsymbol{\theta}) = b_0 \cdot \sigma_t(\boldsymbol{\theta}^0)\},\quad \forall t\right] = 1\\
\Rightarrow &c = c_0,\quad \Pr\left[b^2 \cdot \sigma_t^2(\boldsymbol{\theta}) = b_0^2 \cdot \sigma_t^2(\boldsymbol{\theta}^0),\quad \forall t\right] = 1\\
\Rightarrow &c = c_0,\quad \Pr\left[b^2(\omega + \beta\sigma_{t-1}^2(\boldsymbol{\theta}) + \gamma Y_{t-1}^2) = b_0^2(\omega_0 + \beta_0\sigma_{t-1}^2(\boldsymbol{\theta}^0) + \gamma_0 Y_{t-1}^2),\quad \forall t\right] = 1\\
\Rightarrow &c = c_0,\quad \Pr\left[b^2\omega + \beta b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0) + b^2\gamma Y_{t-1}^2 = b_0^2\omega_0 + \beta_0 b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0) + b_0^2\gamma_0 Y_{t-1}^2,\quad \forall t\right] = 1
\end{aligned}$$

where the third line holds because $e_t(\boldsymbol{\theta}^0) = b_0\sigma_t(\boldsymbol{\theta}^0)$ and we assume that $b_0 < 0$, thus $e_t(\boldsymbol{\theta}^0) < 0$, and in the last line, we replaced $b^2\sigma_{t-1}^2(\boldsymbol{\theta})$ by $b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0)$ because we started with $b^2\sigma_t^2(\boldsymbol{\theta}) = b_0^2\sigma_t^2(\boldsymbol{\theta}^0), \forall t$ almost surely.

Since the GARCH model assumes that $Y_{t-1}|\sigma_{t-1}(\boldsymbol{\theta}^0) \sim F_\eta(0, \sigma_{t-1}^2(\boldsymbol{\theta}^0))$,

$$\begin{aligned}
&\Pr\left[b^2\omega + \beta b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0) + b^2\gamma Y_{t-1}^2 = b_0^2\omega_0 + \beta_0 b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0) + b_0^2\gamma_0 Y_{t-1}^2,\ \forall t\right] = 1\\
&\Rightarrow \Pr\left[\{b^2\gamma = b_0^2\gamma_0\} \cap \{b^2\omega + \beta b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0) = b_0^2\omega_0 + \beta_0 b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0)\},\ \forall t\right] = 1\\
&\Rightarrow b^2\gamma = b_0^2\gamma_0,\quad \Pr\left[b^2\omega + \beta b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0) = b_0^2\omega_0 + \beta_0 b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0),\ \forall t\right] = 1
\end{aligned}$$

If $\beta b_0^2 \neq \beta_0 b_0^2$ and $Pr\left[b^2\omega + \beta b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0) = b_0^2\omega_0 + \beta_0 b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0),\ \forall t\right] = 1$ hold at the same time, then we have

$$\Pr\left[\sigma_{t-1}^2(\boldsymbol{\theta}^0) = \frac{b_0^2\omega_0 - b^2\omega}{\beta b_0^2 - \beta_0 b_0^2},\ \forall t\right] = 1$$

$$\Rightarrow \Pr\left[\omega_0 + \beta_0\sigma_{t-2}^2(\boldsymbol{\theta}^0) + \gamma_0 Y_{t-2}^2 = \frac{b_0^2\omega_0 - b^2\omega}{\beta b_0^2 - \beta_0 b_0^2},\ \forall t\right] = 1$$

This contradicts the assumption of the GARCH model,that $Y_{t-2}|\sigma_{t-2}(\boldsymbol{\theta}^0) \sim F_\eta(0, \sigma_{t-2}^2(\boldsymbol{\theta}^0))$. Thus, $\beta b_0^2 \neq \beta_0 b_0^2$ and $\Pr\left[b^2\omega + \beta b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0) = b_0^2\omega_0 + \beta_0 b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0),\ \forall t\right] = 1$ cannot hold at the same time. This means that $\Pr\left[b^2\omega + \beta b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0) = b_0^2\omega_0 + \beta_0 b_0^2\sigma_{t-1}^2(\boldsymbol{\theta}^0),\ \forall t\right] = 1$ implies $\beta b_0^2 = \beta_0 b_0^2$, which further implies that $\beta = \beta_0$ and $b^2\omega = b_0^2\omega_0$. In summary, we have shown that

$$\Pr\left[v_t(\boldsymbol{\theta}) = v_t(\boldsymbol{\theta}^0) \cap e_t(\boldsymbol{\theta}) = e_t(\boldsymbol{\theta}^0)\right] = 1\ ,\forall\ t$$

$$\Rightarrow c = c_0,\ b^2\gamma = b_0^2\gamma_0,\ \beta b_0^2 = \beta_0 b_0^2,\ b^2\omega = b_0^2\omega_0$$

$$\Rightarrow c = c_0,\ \beta = \beta_0,\ b^2\gamma = b_0^2\gamma_0,\ b^2\omega = b_0^2\omega_0$$

Therefore, Assumption 1(B)(iv) holds if we normalize one of the three parameters b, γ, ω. We choose to normalize ω.

Appendix SA.2.4: Uniform LLN

In this section, we show that under the GARCH assumptions we have made in Section SA.2.2, Assumption 1(A) is satisfied: $L_{FZ0}\left(Y_t, v_t\left(\boldsymbol{\theta}\right), e_t\left(\boldsymbol{\theta}\right); \alpha\right)$ obeys the uniform law of large numbers.

Since the parameter space is assumed to be compact, we can establish the uniform LLN by combining the pointwise LLN with stochastic equicontinuity.

Appendix SA.2.4.1: LLN

The LLN is based on Davidson (1994, Corollary 19.3) which we restate here as Theorem 4 for convenience.

Theorem 4 (Davidson) *Suppose that $(X_t)_{t\in\mathbb{N}}$ satisfies:*

$$\sup_{t\in\mathbb{N}} \mathbb{E}|X_t|^{2+\delta} < \infty \text{ for some } \delta > 0$$

and $(X_t)_{t\in\mathbb{N}}$ is α mixing with $\sum_{m=1}^{\infty} m^{-1}\alpha\left(m\right)^{\delta/(2+\delta)} < \infty$, then

$$\frac{1}{n}\sum_{t=1}^{n}(X_t - \mathbb{E}[X_t]) \stackrel{L_2}{\to} 0.$$

Under Assumption 2(F), which we discussed in the context of the GARCH model in Section SA.2.2 above, implies that $L_{FZ0}\left(Y_t, v_t\left(\boldsymbol{\theta}\right), e_t\left(\boldsymbol{\theta}\right)\right)$ is α-mixing with a decay rate no slower than that required by Theorem 4. Also, $L_{FZ0}\left(Y_t, v_t\left(\boldsymbol{\theta}\right), e_t\left(\boldsymbol{\theta}\right)\right)$ is strictly stationary as we have shown that $\left(Y_t, \sigma_t^2\left(\boldsymbol{\theta}\right), \partial\sigma_t^2\left(\boldsymbol{\theta}\right)/\partial\beta\right)$ is strictly stationary. We then need only show that

$$\mathbb{E}|L_{FZ0}(Y_t, v_t(\boldsymbol{\theta}), e_t(\boldsymbol{\theta}); \alpha)|^{2+\delta} < \infty$$

$$\begin{aligned}
&|L_{FZ0}(Y_t, v_t(\boldsymbol{\theta}), e_t(\boldsymbol{\theta}); \alpha)| \\
&\quad =| -\frac{1}{\alpha e_t(\boldsymbol{\theta})}\mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\}(v_t(\boldsymbol{\theta}) - Y_t) + \frac{v_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})} + \log(-e_t(\boldsymbol{\theta})) - 1| \\
&\quad =|\frac{v_t(\boldsymbol{\theta})}{e_t(\boldsymbol{\theta})} \cdot (1 - \frac{1}{\alpha}\mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\}) + \frac{Y_t}{\alpha e_t(\boldsymbol{\theta})}\mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\} + \log(-e_t(\boldsymbol{\theta})) - 1| \\
&\quad =|c \cdot (1 - \frac{1}{\alpha}\mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\}) - \frac{\eta_t}{\alpha b}\mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\} + \log(-b\sigma_t(\boldsymbol{\theta})) - 1| \\
&\quad \leq c(1 + \frac{1}{\alpha}) + |\log(-b)| + 1 + \frac{|\eta_t|}{\alpha|b|} + |\log\sigma_t|
\end{aligned}$$

By Cr-inequality, it is sufficient to show that

$$\mathbb{E}|\eta_t|^{2+\delta} < \infty \;\text{ and }\; \mathbb{E}|\log\sigma_t|^{2+\delta} < \infty.$$

The moment condition on η_t is directly implied by the structure of the model and GARCH Assumption 3. Recall that $\sigma_t^2 = \omega_0 + \beta\sigma_{t-1}^2 + \gamma y_{t-1}^2 \geq \omega_0 > 0$. Therefore, if $\sigma_t^2 < 1$ then $|\log\sigma_t| \leq |\log\sqrt{\omega_0}|$, and if $\sigma_t^2 \geq 1$ then $|\log\sigma_t| \leq \sigma_t$. In summary, $|\log\sigma_t| \leq |\log\sqrt{\omega_0}| + \sigma_t$. Therefore, by Cr-inequality

$$\begin{aligned}
\mathbb{E}|\log\sigma_t|^{2+\delta} &\leq \mathbb{E}(|\log\sqrt{\omega_0}| + \sigma_t)^{2+\delta} \\
&\leq 2^{1+\delta}(|\log\sqrt{\omega_0}|^{2+\delta} + \mathbb{E}\sigma_t^{2+\delta}).
\end{aligned}$$

Thus, a sufficient condition for $\mathbb{E}|\log\sigma_t|^{2+\delta} < \infty$ is $\mathbb{E}\left[\sigma_t^{2+\delta}\right] < \infty$ which is implied by GARCH Assumption 3. Hence, $L_{FZ0}\left(Y_t, v_t\left(\boldsymbol{\theta}\right), e_t\left(\boldsymbol{\theta}\right)\right)$ obeys the law of large numbers for any fixed $\boldsymbol{\theta}$ by Theorem 4.

Appendix SA.2.4.2: Stochastic equicontinuity

The stochastic equicontinuity condition is derived using Davidson (1994, Theorem 21.10) which we restate here as Theorem 5 for convenience.

Theorem 5 (Davidson) *Let $Q_n(\cdot)$ be the objective function for an M-estimator. Suppose there exists $N \in \mathbb{N}$ such that*

$$|Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}')| \leq a_n h(\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|), \ a.s.$$

holds for all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ and $n \geq \mathbb{N}$, where h is a deterministic function with $h(x) \downarrow 0$ as $x \downarrow 0$, and $a_n = \mathcal{O}_p(1)$. Then $(Q_n)_{n\in\mathbb{N}}$ is stochastically equicontinuous.

Observe that

$$\mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta})\}(v_t(\boldsymbol{\theta}) - Y_t) = \frac{1}{2}(v_t(\boldsymbol{\theta}) - Y_t + |v_t(\boldsymbol{\theta}) - Y_t|). \tag{70}$$

Let $\boldsymbol{\theta}_1 = [\beta_1, \gamma_1, b_1, c_1], \boldsymbol{\theta}_2 = [\beta_2, \gamma_2, b_2, c_2] \in \Theta$. Then, using (70), we obtain

$$\begin{aligned}
&\left|\frac{\mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta}_1)\}(v_t(\boldsymbol{\theta}_1) - Y_t)}{e_t(\boldsymbol{\theta}_1)} - \frac{\mathbf{1}\{Y_t \leq v_t(\boldsymbol{\theta}_2)\}(v_t(\boldsymbol{\theta}_2) - Y_t)}{e_t(\boldsymbol{\theta}_2)}\right| \\
=&\frac{1}{2}\left|\frac{v_t(\boldsymbol{\theta}_1) - Y_t + |v_t(\boldsymbol{\theta}_1) - Y_t|}{e_t(\boldsymbol{\theta}_1)} - \frac{v_t(\boldsymbol{\theta}_2) - Y_t + |v_t(\boldsymbol{\theta}_2) - Y_t|}{e_t(\boldsymbol{\theta}_2)}\right| \\
\leq&\frac{1}{2}\left|\frac{v_t(\boldsymbol{\theta}_1) - Y_t}{e_t(\boldsymbol{\theta}_1)} - \frac{v_t(\boldsymbol{\theta}_2) - Y_t}{e_t(\boldsymbol{\theta}_2)}\right| + \frac{1}{2}\left|\frac{|v_t(\boldsymbol{\theta}_1) - Y_t|}{e_t(\boldsymbol{\theta}_1)} - \frac{|v_t(\boldsymbol{\theta}_2) - Y_t|}{e_t(\boldsymbol{\theta}_2)}\right| \qquad (71) \\
\leq&\left|\frac{v_t(\boldsymbol{\theta}_1) - Y_t}{e_t(\boldsymbol{\theta}_1)} - \frac{v_t(\boldsymbol{\theta}_2) - Y_t}{e_t(\boldsymbol{\theta}_2)}\right| \qquad (72) \\
=&\left|\frac{v_t(\boldsymbol{\theta}_1)}{e_t(\boldsymbol{\theta}_1)} - \frac{v_t(\boldsymbol{\theta}_2)}{e_t(\boldsymbol{\theta}_2)} - \left(\frac{Y_t}{e_t(\boldsymbol{\theta}_1)} - \frac{Y_t}{e_t(\boldsymbol{\theta}_2)}\right)\right| \\
=&\left|c_1 - c_2 - \left(\frac{\eta_t}{b_1} - \frac{\eta_t}{b_2}\right)\right| \\
\leq&|c_1 - c_2| + \frac{|b_1 - b_2|}{|b_1 b_2|}|\eta_t|.
\end{aligned}$$

The inequality between (71) and (72) holds because

$$\begin{aligned}
\left|\frac{|v_t(\boldsymbol{\theta}_1) - Y_t|}{e_t(\boldsymbol{\theta}_1)} - \frac{|v_t(\boldsymbol{\theta}_2) - Y_t|}{e_t(\boldsymbol{\theta}_2)}\right| &= \left|\frac{|v_t(\boldsymbol{\theta}_1) - Y_t|}{-|e_t(\boldsymbol{\theta}_1)|} - \frac{|v_t(\boldsymbol{\theta}_2) - Y_t|}{-|e_t(\boldsymbol{\theta}_2)|}\right| \\
&= \left|\frac{|v_t(\boldsymbol{\theta}_2) - Y_t|}{|e_t(\boldsymbol{\theta}_2)|} - \frac{|v_t(\boldsymbol{\theta}_1) - Y_t|}{|e_t(\boldsymbol{\theta}_1)|}\right| \\
&\leq \left|\frac{v_t(\boldsymbol{\theta}_2) - Y_t}{e_t(\boldsymbol{\theta}_2)} - \frac{v_t(\boldsymbol{\theta}_1) - Y_t}{e_t(\boldsymbol{\theta}_1)}\right|.
\end{aligned}$$

By Taylor's theorem,

$$|\log(-e_t(\boldsymbol{\theta}_1)) - \log(-e_t(\boldsymbol{\theta}_2))| = \left|\frac{1}{-e_t(\boldsymbol{\theta}_1^*)}\right| \cdot \|\nabla e_t(\boldsymbol{\theta}_1^*)\| \cdot \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$$

for some $\boldsymbol{\theta}_1^*$ between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Since,

$$\begin{aligned} \|\nabla e_t(\boldsymbol{\theta})\| =& \|b \cdot \nabla \sigma_t(\boldsymbol{\theta}) + \sigma_t(\boldsymbol{\theta}) \cdot [0,0,1,0]\| \\ \leq& |b| \cdot \|\nabla \sigma_t(\boldsymbol{\theta})\| + \sigma_t(\boldsymbol{\theta}) \\ \Rightarrow \frac{\|\nabla e_t(\boldsymbol{\theta})\|}{|e_t(\boldsymbol{\theta})|} \leq& \frac{|b| \cdot \|\nabla \sigma_t(\boldsymbol{\theta})\| + \sigma_t(\boldsymbol{\theta})}{|b \cdot \sigma_t(\boldsymbol{\theta})|} \leq \frac{\|\nabla \sigma_t(\boldsymbol{\theta})\|}{\sigma_t(\boldsymbol{\theta})} + \frac{1}{|b|}, \end{aligned}$$

we obtain

$$|\log(-e_t(\boldsymbol{\theta}_1)) - \log(-e_t(\boldsymbol{\theta}_2))| \leq \left(\frac{\|\nabla \sigma_t(\boldsymbol{\theta}_1^*)\|}{\sigma_t(\boldsymbol{\theta}_1^*)} + \frac{1}{|b_1^*|} \right) \cdot \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.$$

Therefore,

$$\begin{aligned} &|L_{FZ0}(Y_t, v_t(\boldsymbol{\theta}_1), e_t(\boldsymbol{\theta}_1); \alpha) - L_{FZ0}(Y_t, v_t(\boldsymbol{\theta}_2), e_t(\boldsymbol{\theta}_2); \alpha)| \\ &\leq \frac{1}{\alpha} \left(|c_1 - c_2| + \frac{|b_1 - b_2|}{|b_1 b_2|} |\eta_t| \right) + |c_1 - c_2| + \left(\frac{\|\nabla \sigma_t(\boldsymbol{\theta}_1^*)\|}{\sigma_t(\boldsymbol{\theta}_1^*)} + \frac{1}{|b_1^*|} \right) \cdot \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \\ &\leq \left(1 + \frac{1}{\alpha} + \frac{1}{B_2} + \frac{|\eta_t|}{\alpha B_2^2} + \frac{\|\nabla \sigma_t(\boldsymbol{\theta}_1^*)\|}{\sigma_t(\boldsymbol{\theta}_1^*)} \right) \cdot \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \end{aligned}$$

The last inequality holds because by GARCH Assumption 2, $|b_1|, |b_2|, |b_1^*| \geq B_2 > 0$.

Using Lemma 8 in Section SA.2.5, we obtain

$$\begin{aligned} \frac{\|\nabla \sigma_t(\boldsymbol{\theta})\|}{\sigma_t(\boldsymbol{\theta})} \leq& \frac{1}{2} \cdot \left[\gamma^{1/2} \beta^{-1/2} \sigma_t(\boldsymbol{\theta})^{-1} \sum_{i=2}^{\infty} (i-1) \beta^{(i-2)/2} |Y_{t-i}| + \gamma^{-1} \right] \\ \leq& \frac{1}{2} \cdot \left[(1-\delta_1)^{1/2} \delta_1^{-1/2} \omega_0^{-1/2} \sum_{i=2}^{\infty} (i-1)(1-\delta_1)^{(i-2)/2} |Y_{t-i}| + \delta_1^{-1} \right] \end{aligned}$$

using the bounds on γ and β in GARCH Assumption 2. Define

$$Z_t = \frac{1}{2} \cdot \left[(1-\delta_1)^{1/2} \delta_1^{-1/2} \omega_0^{-1/2} \sum_{i=2}^{\infty} (i-1)(1-\delta_1)^{(i-2)/2} |Y_{t-i}| + \delta_1^{-1} \right]$$

Then,

$$|L_T(\boldsymbol{\theta}_1) - L_T(\boldsymbol{\theta}_2)| \leq \left(1 + \frac{1}{\alpha} + \frac{1}{B_2} + \frac{1}{\alpha B_2^2} \cdot \frac{1}{T} \sum_{t=1}^{T} |\eta_t| + \frac{1}{T} \sum_{t=1}^{T} Z_t \right) \cdot \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.$$

$\mathbb{E}|\eta_t| = \text{const} < \infty$ and $\mathbb{E}[Z_t] = \text{const} < \infty$ (because $\mathbb{E}|Y_t| = \text{const} < \infty$ by GARCH Assumptions 2 and 3). Then, $\frac{1}{T}\sum_{t=1}^{T} |\eta_t| = \mathcal{O}_p(1)$ and $\frac{1}{T}\sum_{t=1}^{T} Z_t = \mathcal{O}_p(1)$.

Therefore, by Theorem 5, L_T is stochastically equicontinuous.

Appendix SA.2.5: Assumptions 2(C) and 2(D)

We define

$$X_1(t;\beta) = \sum_{i=2}^{\infty}(i-1)\beta^{i-2}Y_{t-i}^2$$
$$X_2(t;\beta) = \sum_{i=2}^{\infty}(i-1)\beta^{(i-2)/2}|Y_{t-i}|$$
$$X_3(t;\beta) = \sum_{i=3}^{\infty}\frac{(i-1)(i-2)}{2}\beta^{i-3}Y_{t-i}^2$$

Lemma 8 *Under GARCH Assumption 2, we have*

$$\begin{aligned}
\|\nabla\sigma_t^2(\boldsymbol{\theta})\| &\le \gamma X_1(t;\beta) + \gamma^{-1}\sigma_t^2(\boldsymbol{\theta}) \\
\|\nabla\sigma_t(\boldsymbol{\theta})\| &\le \frac{1}{2}\cdot[\gamma^{1/2}\beta^{-1/2}X_2(t;\beta) + \gamma^{-1}\sigma_t(\boldsymbol{\theta})] \\
\|\nabla^2\sigma_t^2(\boldsymbol{\theta})\| &\le 2[\frac{\omega_0}{(1-\beta)^3} + \gamma X_3(t;\beta) + X_1(t;\beta)] \\
\|\nabla^2\sigma_t(\boldsymbol{\theta})\| &\le \frac{\omega_0^{-1/2}}{4}\cdot\gamma\beta^{-1}X_2(t;\beta)^2 + \gamma^{-1/2}\beta^{-1/2}X_2(t;\beta) + \frac{1}{4}\gamma^{-2}\sigma_t(\boldsymbol{\theta}) \\
&\quad + \omega_0^{-1/2}\cdot[\frac{\omega_0}{(1-\beta)^3} + \gamma X_3(t;\beta) + X_1(t;\beta)]
\end{aligned}$$

Proof of Lemma 8.

$$\sigma_t^2(\boldsymbol{\theta}) = \omega_0 + \beta\sigma_{t-1}^2(\boldsymbol{\theta}) + \gamma Y_{t-1}^2 = \frac{\omega_0}{1-\beta} + \gamma\sum_{i=1}^{\infty}\beta^{i-1}Y_{t-i}^2. \tag{73}$$

Therefore,

$$\begin{aligned}
\nabla\sigma_t^2(\boldsymbol{\theta}) &= \left[\omega_0/(1-\beta)^2 + \gamma\sum_{i=2}^{\infty}(i-1)\beta^{i-2}Y_{t-i}^2,\ \sum_{i=1}^{\infty}\beta^{i-1}Y_{t-i}^2, 0, 0\right] \\
&= \left[\omega_0/(1-\beta)^2 + \gamma X_1(t;\beta),\ \gamma^{-1}(\sigma_t^2 - \omega_0/(1-\beta)), 0, 0\right] \qquad (74) \\
\|\nabla\sigma_t^2(\boldsymbol{\theta})\| &\le \omega_0/(1-\beta)^2 + \gamma X_1(t;\beta) + \gamma^{-1}(\sigma_t^2 - \omega_0/(1-\beta)) \\
&= \gamma X_1(t;\beta) + \gamma^{-1}\sigma_t^2 + \frac{\omega_0}{1-\beta}\left(\frac{1}{1-\beta} - \frac{1}{\gamma}\right) \\
&\le \gamma X_1(t;\beta) + \gamma^{-1}\sigma_t^2,
\end{aligned}$$

since $\beta+\gamma\le 1$ implies $1/(1-\beta) - 1/\gamma \le 0$. Furthermore,

$$\|\nabla\sigma_t(\boldsymbol{\theta})\| = \frac{\|\nabla\sigma_t^2(\boldsymbol{\theta})\|}{2\sigma_t(\boldsymbol{\theta})} \le \frac{1}{2}\left[\frac{\gamma\sum_{i=2}^{\infty}(i-1)\beta^{i-2}Y_{t-i}^2}{\sqrt{\frac{\omega_0}{1-\beta} + \gamma\sum_{i=1}^{\infty}\beta^{i-1}Y_{t-i}^2}} + \frac{\gamma^{-1}\sigma_t^2}{\sigma_t}\right].$$

For all $j \geq 2$, we have

$$\frac{\gamma(j-1)\beta^{j-2}Y_{t-j}^2}{\sqrt{\frac{\omega_0}{1-\beta}+\gamma\sum_{i=1}^{\infty}\beta^{i-1}Y_{t-i}^2}} \leq \frac{\gamma(j-1)\beta^{j-2}Y_{t-j}^2}{\sqrt{\gamma\beta^{j-1}Y_{t-j}^2}} = (j-1)\gamma^{1/2}\beta^{(j-3)/2}|Y_{t-j}|$$

as all summands in the denominator are positive. This implies

$$\begin{aligned}\|\nabla\sigma_t(\boldsymbol{\theta})\| &\leq \frac{1}{2}\cdot\left[\gamma^{1/2}\beta^{-1/2}\sum_{i=2}^{\infty}(i-1)\beta^{(i-2)/2}|Y_{t-i}| + \gamma^{-1}\sigma_t(\boldsymbol{\theta})\right] \\ &= \frac{1}{2}\cdot[\gamma^{1/2}\beta^{-1/2}X_2(t;\beta)+\gamma^{-1}\sigma_t(\boldsymbol{\theta})].\end{aligned}$$

Using equation (74), we obtain

$$\nabla^2\sigma_t^2(\boldsymbol{\theta}) = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

where

$$\begin{aligned} a_{11} &= \frac{2\omega_0}{(1-\beta)^3} + \gamma\sum_{i=3}^{\infty}(i-1)(i-2)\beta^{i-3}Y_{t-i}^2 \\ a_{12} = a_{21} &= \sum_{i=2}^{\infty}(i-1)\beta^{i-2}Y_{t-i}^2 \end{aligned}$$

Thus, since the Frobenius norm is always less than the sum of the absolute values of the matrix entries,

$$\|\nabla^2\sigma_t^2(\boldsymbol{\theta})\| \leq 2[\frac{\omega_0}{(1-\beta)^3} + \gamma X_3(t;\beta) + X_1(t;\beta)].$$

Using that $\nabla\sigma_t(\boldsymbol{\theta}) = \nabla\sigma_t^2(\boldsymbol{\theta})/(2\sigma_t(\boldsymbol{\theta}))$, we find that

$$\nabla^2\sigma_t(\boldsymbol{\theta}) = \frac{\nabla_t'^2(\boldsymbol{\theta})\nabla\sigma_t(\boldsymbol{\theta})}{-2\sigma_t^2(\boldsymbol{\theta})} + \frac{\nabla^2\sigma_t^2(\boldsymbol{\theta})}{2\sigma_t(\boldsymbol{\theta})} = \frac{\nabla'\sigma_t(\boldsymbol{\theta})\nabla\sigma_t(\boldsymbol{\theta})}{-\sigma_t(\boldsymbol{\theta})} + \frac{\nabla^2\sigma_t^2(\boldsymbol{\theta})}{2\sigma_t(\boldsymbol{\theta})},$$

therefore,

$$\|\nabla^2\sigma_t(\boldsymbol{\theta})\| \leq \frac{\|\nabla\sigma_t(\boldsymbol{\theta})\|^2}{\sigma_t(\boldsymbol{\theta})} + \frac{\|\nabla^2\sigma_t^2(\boldsymbol{\theta})\|}{2\sigma_t(\boldsymbol{\theta})}$$

Since $\sigma_t^2 \geq \omega_0 > 0$ and using our previous results, we obtain the claimed bound on $\|\nabla^2\sigma_t(\boldsymbol{\theta})\|$. ∎

Lemma 9 *Under GARCH Assumption 2, it holds that*

$$\begin{aligned}
|v_t(\boldsymbol{\theta})| \leq & V(\mathcal{F}_{t-1}) = B_1 \cdot S_1(\mathcal{F}_{t-1}) \\
\|\nabla e_t(\boldsymbol{\theta})\| \leq & H_1(\mathcal{F}_{t-1}) = B_1 \cdot S_2(\mathcal{F}_{t-1}) + S_1(\mathcal{F}_{t-1}) \\
\|\nabla v_t(\boldsymbol{\theta})\| \leq & V_1(\mathcal{F}_{t-1}) = H_1(\mathcal{F}_{t-1}) + V(\mathcal{F}_{t-1}) \\
\|\nabla^2 e_t(\boldsymbol{\theta})\| \leq & H_2(\mathcal{F}_{t-1}) = B_1 \cdot S_3(\mathcal{F}_{t-1}) + 2S_2(\mathcal{F}_{t-1}) \\
\|\nabla^2 v_t(\boldsymbol{\theta})\| \leq & V_2(\mathcal{F}_{t-1}) = H_2(\mathcal{F}_{t-1}) + 2H_1(\mathcal{F}_{t-1}),
\end{aligned}$$

where

$$\begin{aligned}
S_1(\mathcal{F}_{t-1}) = & \sqrt{\omega_0 \delta_1^{-1} + \gamma \sum_{i=1}^{\infty} (1-\delta_1)^{i-1} Y_{t-i}^2} \\
S_2(\mathcal{F}_{t-1}) = & \frac{1}{2} \cdot [(1-\delta_1)^{1/2} \delta_1^{-1/2} X_2(t; 1-\delta_1) + \delta_1^{-1} S_1(\mathcal{F}_{t-1})] \\
S_3(\mathcal{F}_{t-1}) = & \frac{\omega_0^{-1/2}}{4} \cdot (\delta_1^{-1} - 1) X_2(t; 1-\delta_1)^2 + \delta_1^{-1} X_2(t; 1-\delta_1) + \frac{1}{4} \delta_1^{-2} S_1(\mathcal{F}_{t-1}) \\
& + \omega_0^{-1/2} \cdot [\omega_0 \delta_1^{-3} + (1-\delta_1) X_3(t; 1-\delta_1) + X_1(t; 1-\delta_1)].
\end{aligned}$$

Proof of Lemma 9. As a function in β and γ, $\sigma_t(\boldsymbol{\theta})$ is increasing in both arguments, see equation (73), and, in fact, it does not depend on the parameters b and c. Therefore, $\sigma_t(\boldsymbol{\theta}) \leq S_1(\mathcal{F}_{t-1})$. The quantities $X_1(t,\beta)$, $X_2(t,\beta)$, $X_3(t,\beta)$ defined in the beginning of this section are all increasing in β, and thus, bounded by $X_1(t, 1-\delta_1)$, $X_2(t, 1-\delta_1)$, $X_3(t, 1-\delta_1)$, respectively. Recall that $|b| \leq B_1$ under GARCH Assumption 2.

The first inequality holds because $|v_t(\boldsymbol{\theta})| \leq |e_t(\boldsymbol{\theta})| = |b| \cdot \sigma_t(\boldsymbol{\theta})$. The remaining ones are implied by Lemma 8 and

$$\begin{aligned}
\|\nabla e_t(\boldsymbol{\theta})\| = & \|b \cdot \nabla \sigma_t(\boldsymbol{\theta}) + \sigma_t(\boldsymbol{\theta}) \cdot [0,0,1,0]\| \leq |b| \cdot \|\nabla \sigma_t(\boldsymbol{\theta})\| + \sigma_t(\boldsymbol{\theta}) \\
\|\nabla v_t(\boldsymbol{\theta})\| = & \|c \cdot \nabla e_t(\boldsymbol{\theta}) + e_t(\boldsymbol{\theta}) \cdot [0,0,0,1]\| \leq \|\nabla e_t(\boldsymbol{\theta})\| + |b| \cdot \sigma_t(\boldsymbol{\theta}) \\
\|\nabla^2 e_t(\boldsymbol{\theta})\| = & \|b \nabla^2 \sigma_t(\boldsymbol{\theta}) + [0,0,1,0]' \nabla \sigma_t(\boldsymbol{\theta}) + \nabla' \sigma_t(\boldsymbol{\theta})[0,0,1,0]\| \\
\leq & |b| \cdot \|\nabla^2 \sigma_t(\boldsymbol{\theta})\| + 2\|\nabla \sigma_t(\boldsymbol{\theta})\| \\
\|\nabla^2 v_t(\boldsymbol{\theta})\| = & \|c \nabla^2 e_t(\boldsymbol{\theta}) + [0,0,0,1]' \nabla e_t(\boldsymbol{\theta}) + \nabla' e_t(\boldsymbol{\theta})[0,0,0,1]\| \\
\leq & \|\nabla^2 e_t(\boldsymbol{\theta})\| + 2\|\nabla e_t(\boldsymbol{\theta})\|.
\end{aligned}$$

■

Lemma 10 *Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of random variables and define $X = \sum_{i=1}^{\infty} a_i|X_i|$, where $a_i > 0$ for all $i \in \mathbb{N}$ and $\sum_{i=1}^{\infty} a_i < \infty$. Let $p > 1$. If $\sup_{i\in\mathbb{N}} \mathbb{E}|X_i|^p \le K < \infty$ for some constant K, then $\mathbb{E}|X|^p \le (\sum_{i=1}^{\infty} a_i)^p K$.*

Proof of Lemma 10. By Jensen's inequality, $\mathbb{E}|Z|^p \ge |\mathbb{E}Z|^p$. We rewrite X as

$$X = \sum_{i=1}^{\infty} a_i|X_i| = \left(\sum_{i=1}^{\infty} a_i\right) \cdot \sum_{i=1}^{\infty} \left(\sum_{i=1}^{\infty} a_i\right)^{-1} a_i|X_i|.$$

Note that $\sum_{i=1}^{\infty}(\sum_{i=1}^{\infty} a_i)^{-1}a_i = 1$, namely $\{(\sum_{i=1}^{\infty} a_i)^{-1}a_i\}_{i=1}^{\infty}$ is a probability measure. Then, using Jensen's inequality,

$$X^p = \left(\sum_{j=1}^{\infty} a_j\right)^p \cdot \left(\sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} a_j\right)^{-1} a_i|X_i|\right)^p \le \left(\sum_{j=1}^{\infty} a_j\right)^p \cdot \sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} a_j\right)^{-1} a_i|X_i|^p.$$

Thus,

$$\mathbb{E}\left[X^p\right] \le \left(\sum_{j=1}^{\infty} a_j\right)^p \cdot \sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} a_j\right)^{-1} a_i\mathbb{E}|X_i|^p \le \left(\sum_{j=1}^{\infty} a_j\right)^p \cdot \sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} a_j\right)^{-1} a_i K = \left(\sum_{j=1}^{\infty} a_j\right)^p K$$

■

Lemma 11 *Under GARCH Assumption 2 and for any $p > 1$, $p_1, \dots, p_6 > 0$, the following statements hold for all t:*

(1) If $\mathbb{E}|Y_t|^p < \infty$, then the following quantities are all finite: $\mathbb{E}[V^p(\mathcal{F}_{t-1})]$, $\mathbb{E}[V_1^p(\mathcal{F}_{t-1})]$, $\mathbb{E}[H_1^p(\mathcal{F}_{t-1})]$, $\mathbb{E}[V_2^{p/2}(\mathcal{F}_{t-1})]$, $\mathbb{E}[H_2^{p/2}(\mathcal{F}_{t-1})]$.

(2) If $p = p_1 + p_2 + p_3 + 2p_4 + 2p_5 + p_6$ and $\mathbb{E}|Y_t|^p < \infty$, then

$$\mathbb{E}[V^{p_1}(\mathcal{F}_{t-1})V_1^{p_2}(\mathcal{F}_{t-1})H_1^{p_3}(\mathcal{F}_{t-1})V_2^{p_4}(\mathcal{F}_{t-1})H_2^{p_5}(\mathcal{F}_{t-1})|Y_t|^{p_6}] < \infty.$$

(3) If $\mathbb{E}|Y_t|^{4+\delta} < \infty$ for some $\delta > 0$, all the moment conditions in Assumption 2(D)' could be satisfied.

Proof of Lemma 11. Part (1) follows by combining Lemma 9 with Lemma 10, and part (2) is a consequence of part (1) and Hölder's inequality. ■

Lemma 11 implies that GARCH Assumption 3 implies Assumption 2(D) of the paper.

Appendix SA.2.6: Assumption 2(E)

$\mathbf{D}_0$ is the Hessian of the expected loss at $\boldsymbol{\theta}^0$, so it is positive semi-definite. Let $x = (x_1, \ldots, x_4)' \in \mathbb{R}^4$ such that $x'\mathbf{D}_0 x = 0$. This implies that $x'\nabla v_t(\boldsymbol{\theta}^0) = 0$, $x'\nabla e_t(\boldsymbol{\theta}^0) = 0$ almost surely. We have, $x'\nabla v_t(\boldsymbol{\theta}^0) = cx'\nabla e_t(\boldsymbol{\theta}^0) + x_4 e_t(\boldsymbol{\theta}^0)$. Therefore, $x_4 = 0$. Furthermore,

$$\begin{aligned} 2\sigma_t(\boldsymbol{\theta}^0)x'\nabla e_t(\boldsymbol{\theta}^0) &= 2\sigma_t(\boldsymbol{\theta}^0)bx'\nabla\sigma_t(\boldsymbol{\theta}^0) + 2x_3\sigma_t^2(\boldsymbol{\theta}^0) \\ &= bx'\nabla\sigma_t^2(\boldsymbol{\theta}^0) + 2x_3\sigma_t^2(\boldsymbol{\theta}^0) = 0, \text{ a.s.} \end{aligned} \tag{75}$$

The stationarity of $\{Y_t\}$ implies that $\sigma_t^2(\boldsymbol{\theta}^0)$ is stationary. Therefore it also holds that

$$bx'\nabla\sigma_{t-1}^2(\boldsymbol{\theta}^0) + 2x_3\sigma_{t-1}^2(\boldsymbol{\theta}^0) = 0, \text{ a.s.} \tag{76}$$

Computing (75) $-\beta\cdot$(76), we obtain that a.s.

$$0 = bx'[\sigma_{t-1}^2(\boldsymbol{\theta}^0), Y_{t-1}^2, 0, 0]' + 2x_3(\omega_0 + \gamma Y_{t-1}^2) = (bx_2 + 2\gamma x_3)Y_{t-1}^2 + (2\omega_0 x_3 + bx_1\sigma_{t-1}^2(\boldsymbol{\theta}^0)). \tag{77}$$

By the assumption that $Y_{t-1}|\sigma_{t-1}^2 \sim F_\eta(0, \sigma_{t-1}^2(\boldsymbol{\theta}^0))$ and that $\sigma_{t-1}^2(\boldsymbol{\theta}^0) = \omega_0 + \beta_0\sigma_{t-2}^2(\boldsymbol{\theta}^0) + \gamma_0 Y_{t-2}^2$, we can conclude from the above equation that $x_1 = x_2 = x_3 = 0$. Thus $\mathbf{D}_0$ is positive definite.

Appendix SA.2.7: Assumption 2(G)

We now verify this assumption for the GARCH(1,1) model. Set $a = bc$, so that $v_t = a\sigma_t$. Then for $T \geq 5$, a necessary condition for $Y_t = v_t(\boldsymbol{\theta})$, $t = 1, \ldots, T$ is given by the set of equations

$$Y_t^2 = a^2\beta^t\sigma_0^2 + a^2\beta^{t-1}(\omega_0 + \gamma Y_0^2) + a^2\sum_{k=1}^{t-1}\beta^{t-1-k}(\omega_0 + \gamma Y_k^2), \quad t = 1, \ldots, 4$$

or, equivalently,

$$Y_1^2 = a^2\beta\sigma_0^2 + a^2(\omega_0 + \gamma Y_0^2) \tag{78}$$

$$Y_2^2 = \beta Y_1^2 + a^2(\omega_0 + \gamma Y_1^2) \tag{79}$$

$$Y_3^2 = \beta Y_2^2 + a^2(\omega_0 + \gamma Y_2^2) \tag{80}$$

$$Y_4^2 = \beta Y_3^2 + a^2(\omega_0 + \gamma Y_3^2). \tag{81}$$

Solving equations (79)-(81) for β and equating the results, we obtain

$$\frac{a^2}{\omega_0} = \frac{Y_2^4 - Y_1^2Y_3^2}{Y_2^2 - Y_1^2} = \frac{Y_3^4 - Y_2^2Y_4^2}{Y_3^2 - Y_2^2}. \tag{82}$$

Therefore, a necessary condition such that $Y_t = v_t(\theta)$, $t = 1, \ldots, T$ for some parameter $\boldsymbol{\theta} \in \Theta$ is that $(Y_1, \ldots, Y_T)$ lies in the set $p^{-1}(0) = \{(Y_1, ..., Y_T) \in \mathbb{R}^T | p(Y_1, ..., Y_T) = 0\}$, where p is the polynomial function

$$p(Y_1, \ldots, Y_T) = (Y_2^4 - Y_1^2 Y_3^2)(Y_3^2 - Y_2^2) - (Y_3^4 - Y_2^2 Y_4^2)(Y_2^2 - Y_1^2).$$

The set $p^{-1}(0)$ has Hausdorff dimension less than T. Therefore, as the distribution of $(Y_1, \ldots, Y_T)$ is assumed to be absolutely continuous from GARCH Assumption 1, we obtain the claim with $K = 4$.

Appendix SA.2.8: Summary

We summarize the arguments showing that Assumption 1 and 2 of the paper are satisfied under GARCH Assumptions 1–3.

Assumption 1: Part (A) holds as it has been shown in Section SA.2.4.1 that the uniform law of large number holds under our GARCH Assumptions. Part (B)(i)-(ii) are satisfied under GARCH Assumptions 1-2. Part (B)(iii) is easy to check. Concerning Part (B)(iv), we have shown in Section SA.2.3 that the GARCH model is identifiable when ω is normalized.

Assumption 2: Part (A)(i) is easy to check, (ii) is satisfied by GARCH Assumption 1. Part (B)(i) is satisfied by GARCH Assumption 1, (ii) is clearly weaker than GARCH Assumption 3. Part (C)(i) follows easily from $\sigma_t(\boldsymbol{\theta})^2 \geq \omega_0 > 0$ and the bounds on the parameter $|b|$. Part (C)(ii) has been shown in Lemma 9. Part (D) is implied by Lemma 11. Part (E) is discussed in Section SA.2.6. Part (F) is satisfied under GARCH Assumptions 2–3 as discussed in Section SA.2.2, and Part (G) is satisfied by GARCH Assumption 1 as discussed in Section SA.2.7.

Appendix SA.3: Additional tables

Table S1: Finite-sample performance of (Q)MLE

	$T = 2500$			$T = 5000$		
	ω	β	γ	ω	β	γ
Panel A: N(0,1) innovations						
True	0.050	0.950	0.050	0.050	0.950	0.050
Median	0.053	0.897	0.050	0.051	0.899	0.050
Avg bias	0.011	(0.011)	0.000	0.005	(0.005)	0.000
St dev	0.056	0.064	0.013	0.023	0.029	0.009
Coverage	0.936	0.930	0.928	0.936	0.933	0.937
Panel B: Skew t (5,-0.5) innovations						
True	0.050	0.950	0.050	0.050	0.950	0.050
Median	0.052	0.895	0.049	0.052	0.897	0.050
Avg bias	0.017	(0.023)	0.005	0.006	(0.008)	0.002
St dev	0.077	0.095	0.028	0.026	0.037	0.017
Coverage	0.899	0.907	0.897	0.913	0.907	0.903

Notes: This table presents results from 1000 replications of the estimation of the parameters of a GARCH(1,1) model, using the Normal likelihood. In Panel A the innovations are standard Normal, and so estimation is then ML. In Panel B the innovations are standardized skew t, and so estimation is QML. Details are described in Section 4 of the main paper. The top row of each panel presents the true values of the parameters. The second, third, and fourth rows present the median estimated parameters, the average bias, and the standard deviation (across simulations) of the estimated parameters. The last row of each panel presents the coverage rates for 95% confidence intervals constructed using estimated standard errors.

Table S2: Simulation results for Normal innovations, estimation by CAViaR

	$T = 2500$			$T = 5000$		
	β	γ	a_α	β	γ	a_α
			$\alpha = 0.01$			
True	0.900	0.050	-2.326	0.900	0.050	-2.326
Median	0.901	0.048	-2.275	0.899	0.048	-2.347
Avg bias	-0.017	0.012	-0.120	-0.011	0.006	-0.095
St dev	0.079	0.066	0.957	0.051	0.034	0.718
Coverage	0.881	0.874	0.907	0.892	0.886	0.905
			$\alpha = 0.025$			
True	0.900	0.050	-1.960	0.900	0.050	-1.960
Median	0.898	0.047	-1.953	0.896	0.047	-2.009
Avg bias	-0.018	0.005	-0.136	-0.012	0.002	-0.110
St dev	0.068	0.038	0.728	0.052	0.023	0.566
Coverage	0.906	0.879	0.934	0.913	0.892	0.918
			$\alpha = 0.05$			
True	0.900	0.050	-1.645	0.900	0.050	-1.645
Median	0.901	0.047	-1.639	0.899	0.049	-1.667
Avg bias	-0.014	0.005	-0.085	-0.009	0.002	-0.070
St dev	0.068	0.037	0.597	0.045	0.023	0.436
Coverage	0.909	0.884	0.930	0.918	0.900	0.935
			$\alpha = 0.10$			
True	0.900	0.050	-1.282	0.900	0.050	-1.282
Median	0.898	0.047	-1.291	0.898	0.048	-1.289
Avg bias	-0.016	0.006	-0.076	-0.010	0.003	-0.055
St dev	0.069	0.041	0.482	0.047	0.025	0.364
Coverage	0.916	0.883	0.933	0.921	0.896	0.937
			$\alpha = 0.20$			
True	0.900	0.050	-0.842	0.900	0.050	-0.842
Median	0.898	0.048	-0.848	0.899	0.048	-0.840
Avg bias	-0.023	0.022	-0.058	-0.016	0.007	-0.049
St dev	0.091	0.107	0.391	0.063	0.044	0.304
Coverage	0.914	0.876	0.931	0.929	0.901	0.940

Notes: This table presents results from 1000 replications of the estimation of VaR from a GARCH(1,1) DGP with standard Normal innovations. Details are described in Section 4 of the main paper. The top row of each panel presents the true values of the parameters. The second, third, and fourth rows present the median estimated parameters, the average bias, and the standard deviation (across simulations) of the estimated parameters. The last row of each panel presents the coverage rates for 95% confidence intervals constructed using estimated standard errors.

Table S3: Simulation results for skew t innovations, estimation by CAViaR

	$T = 2500$			$T = 5000$		
	β	γ	a_α	β	γ	a_α
			$\alpha = 0.01$			
True	0.900	0.050	-3.290	0.900	0.050	-3.290
Median	0.898	0.045	-3.272	0.899	0.045	-3.306
Avg bias	-0.041	0.022	-0.355	-0.027	0.008	-0.306
St dev	0.142	0.097	1.928	0.103	0.044	1.546
Coverage	0.771	0.805	0.827	0.785	0.808	0.823
			$\alpha = 0.025$			
True	0.900	0.050	-2.408	0.900	0.050	-2.408
Median	0.899	0.047	-2.371	0.898	0.049	-2.414
Avg bias	-0.026	0.012	-0.190	-0.016	0.004	-0.144
St dev	0.103	0.067	1.135	0.070	0.033	0.862
Coverage	0.832	0.841	0.877	0.830	0.862	0.859
			$\alpha = 0.05$			
True	0.900	0.050	-1.800	0.900	0.050	-1.800
Median	0.899	0.047	-1.780	0.899	0.049	-1.792
Avg bias	-0.023	0.008	-0.146	-0.013	0.004	-0.087
St dev	0.092	0.060	0.782	0.057	0.028	0.563
Coverage	0.863	0.861	0.892	0.883	0.871	0.890
			$\alpha = 0.10$			
True	0.900	0.050	-1.223	0.900	0.050	-1.223
Median	0.900	0.049	-1.205	0.900	0.049	-1.217
Avg bias	-0.019	0.008	-0.074	-0.010	0.004	-0.043
St dev	0.080	0.050	0.495	0.050	0.027	0.356
Coverage	0.895	0.892	0.919	0.892	0.905	0.910
			$\alpha = 0.20$			
True	0.900	0.050	-0.652	0.900	0.050	-0.652
Median	0.903	0.051	-0.619	0.902	0.051	-0.636
Avg bias	-0.027	0.026	-0.035	-0.016	0.009	-0.028
St dev	0.122	0.109	0.353	0.084	0.042	0.271
Coverage	0.867	0.887	0.897	0.890	0.889	0.916

Notes: This table presents results from 1000 replications of the estimation of VaR from a GARCH(1,1) DGP with skew t innovations. Details are described in Section 4 of the main paper. The top row of each panel presents the true values of the parameters. The second, third, and fourth rows present the median estimated parameters, the average bias, and the standard deviation (across simulations) of the estimated parameters. The last row of each panel presents the coverage rates for 95% confidence intervals constructed using estimated standard errors.

Table S4: Diebold-Mariano t-statistics on average out-of-sample loss differences for the DJIA, NIKKEI and FTSE100 (alpha=0.05)

	RW125	RW250	RW500	G-N	G-Skt	G-EDF	FZ-2F	FZ-1F	G-FZ	Hybrid
					Panel A: DJIA					
RW125		-2.200	-3.536	2.324	2.860	2.935	3.006	3.821	3.244	3.494
RW250	2.200		-3.349	2.983	3.411	3.502	3.989	4.522	3.926	3.957
RW500	3.536	3.349		3.979	4.336	4.417	4.805	5.321	4.829	4.860
G-N	-2.324	-2.983	-3.979		3.573	2.787	0.791	1.419	1.472	1.670
G-Skt	-2.860	-3.411	-4.336	-3.573		1.385	-0.034	0.625	0.195	0.302
G-EDF	-2.935	-3.502	-4.417	-2.787	-1.385		-0.266	0.432	-0.119	-0.031
FZ-2F	-3.006	-3.989	-4.805	-0.791	0.034	0.266		1.085	0.192	0.247
FZ-1F	-3.821	-4.522	-5.321	-1.419	-0.625	-0.432	-1.085		-0.796	-0.613
G-FZ	-3.244	-3.926	-4.829	-1.472	-0.195	0.119	-0.192	0.796		0.126
Hybrid	-3.494	-3.957	-4.86	-1.670	-0.302	0.031	-0.247	0.613	-0.126	
					Panel B: NIKKEI					
RW125		-0.225	-1.047	3.703	3.687	3.719	3.733	3.219	3.692	3.868
RW250	0.225		-1.162	4.048	4.058	4.098	3.897	3.582	4.076	4.249
RW500	1.047	1.162		3.733	3.748	3.785	3.768	3.387	3.773	3.847
G-N	-3.703	-4.048	-3.733		1.165	2.110	-1.841	-1.261	1.861	0.457
G-Skt	-3.687	-4.058	-3.748	-1.165		1.797	-1.888	-1.378	1.468	0.295
G-EDF	-3.719	-4.098	-3.785	-2.110	-1.797		-1.984	-1.522	-0.797	0.100
FZ-2F	-3.733	-3.897	-3.768	1.841	1.888	1.984		1.209	1.958	2.489
FZ-1F	-3.219	-3.582	-3.387	1.261	1.378	1.522	-1.209		1.487	2.624
G-FZ	-3.692	-4.076	-3.773	-1.861	-1.468	0.797	-1.958	-1.487		0.134
Hybrid	-3.868	-4.249	-3.847	-0.457	-0.295	-0.100	-2.489	-2.624	-0.134	

Table continued on next page.

Table S4 (cont'd): Diebold-Mariano t-statistics on average out-of-sample loss differences for the DJIA, NIKKEI and FTSE100 (alpha=0.05)

	RW125	RW250	RW500	G-N	G-Skt	G-EDF	FZ-2F	FZ-1F	G-FZ	Hybrid
					Panel C: FTSE					
RW125		-2.329	-3.439	3.275	3.485	3.450	2.732	3.279	3.300	3.141
RW250	2.329		-2.751	4.146	4.337	4.305	3.663	4.264	4.160	4.025
RW500	3.439	2.751		4.682	4.845	4.817	4.232	4.848	4.696	4.661
G-N	-3.275	-4.146	-4.682		4.327	4.446	-0.210	-0.070	0.581	1.048
G-Skt	-3.485	-4.337	-4.845	-4.327		-3.853	-0.746	-0.877	-4.066	0.428
G-EDF	-3.450	-4.305	-4.817	-4.446	3.853		-0.648	-0.731	-3.949	0.545
FZ-2F	-2.732	-3.663	-4.232	0.210	0.746	0.648		0.213	0.249	1.401
FZ-1F	-3.279	-4.264	-4.848	0.070	0.877	0.731	-0.213		0.128	1.321
G-FZ	-3.300	-4.160	-4.696	-0.581	4.066	3.949	-0.249	-0.128		1.006
Hybrid	-3.141	-4.025	-4.661	-1.048	-0.428	-0.545	-1.401	-1.321	-1.006	

Notes: This table presents t-statistics from Diebold-Mariano tests comparing the average losses, using the FZ0 loss function, over the out-of-sample period from January 2000 to December 2016, for ten different forecasting models. A positive value indicates that the row model has higher average loss than the column model. Values greater than 1.96 in absolute value indicate that the average loss difference is significantly different from zero at the 95% confidence level. Values along the main diagonal are all identically zero and are omitted for interpretability. The first three rows correspond to rolling window forecasts, the next three rows correspond to GARCH forecasts based on different models for the standardized residuals, and the last four rows correspond to models introduced in Section 2 of the main paper.

Table S5: Out-of-sample average losses and goodness-of-fit tests (alpha=0.025)

	Average loss				*GoF p-values: VaR*				*GoF p-values: ES*			
	S&P	**DJIA**	**NIK**	**FTSE**	**S&P**	**DJIA**	**NIK**	**FTSE**	**S&P**	**DJIA**	**NIK**	**FTSE**
RW-125	1.119	1.088	1.525	1.166	0.036	0.004	0.001	0.000	0.017	0.006	0.001	0.001
RW-250	1.164	1.117	1.525	1.209	0.009	0.009	0.006	0.000	0.037	*0.056*	0.015	0.006
RW-500	1.245	1.187	1.561	1.294	0.003	0.001	0.011	0.000	0.032	0.025	0.014	0.000
GCH-N	1.089	1.021	1.341	1.052	0.000	0.001	**0.177**	0.000	0.000	0.000	*0.053*	0.000
GCH-Skt	1.044	0.978	**1.328**	**1.026**	0.008	0.009	**0.796**	0.001	0.011	0.006	**0.725**	0.001
GCH-EDF	*1.028*	*0.969*	1.329	*1.042*	**0.188**	0.031	**0.796**	0.000	**0.258**	0.017	**0.593**	0.000
FZ-2F	1.039	0.998	1.421	1.242	0.000	0.002	**0.341**	0.000	0.001	0.001	**0.158**	0.000
FZ-1F	1.030	0.985	1.390	1.056	*0.057*	0.007	**0.773**	0.000	**0.130**	*0.058*	**0.415**	0.000
GCH-FZ	**1.020**	**0.951**	**1.328**	1.055	**0.125**	**0.364**	**0.688**	0.000	*0.222*	*0.403*	**0.521**	0.000
Hybrid	1.053	1.030	1.345	1.079	0.001	**0.114**	**0.558**	0.000	0.002	*0.075*	**0.464**	0.000

Notes: The left panel of this table presents the average losses, using the FZ0 loss function, for four daily equity return series, over the out-of-sample period from January 2000 to December 2016, for ten different forecasting models. The lowest average loss in each column is highlighted in bold, the second-lowest is highlighted in italics. The first three rows correspond to rolling window forecasts, the next three rows correspond to GARCH forecasts based on different models for the standardized residuals, and the last four rows correspond to models introduced in Section 2 of the main paper. The middle and right panels of this table present p-values from goodness-of-fit tests of the VaR and ES forecasts respectively. Values that are greater than 0.10 (indicating no evidence against optimality at the 0.10 level) are in bold, and values between 0.05 and 0.10 are in italics.

Table S6: Diebold-Mariano t-statistics on average out-of-sample loss differences for the S&P 500, DJIA, NIKKEI and FTSE100 (alpha=0.025)

	RW125	RW250	RW500	G-N	G-Skt	G-EDF	FZ-2F	FZ-1F	G-FZ	Hybrid
					Panel A: S&P 500					
RW125		-1.836	-2.988	1.025	2.479	2.788	2.146	3.371	2.891	2.419
RW250	1.836		-2.815	1.725	2.747	3.004	2.602	3.712	3.135	2.992
RW500	2.988	2.815		2.823	3.673	3.893	3.630	4.624	4.023	4.045
G-N	-1.025	-1.725	-2.823		4.019	3.368	2.083	2.429	3.698	1.928
G-Skt	-2.479	-2.747	-3.673	-4.019		2.275	0.270	0.815	2.742	-0.594
G-EDF	-2.788	-3.004	-3.893	-3.368	-2.275		-0.592	-0.074	1.393	-1.483
FZ-2F	-2.146	-2.602	-3.630	-2.083	-0.270	0.592		0.487	1.227	-0.729
FZ-1F	-3.371	-3.712	-4.624	-2.429	-0.815	0.074	-0.487		0.579	-1.605
G-FZ	-2.891	-3.135	-4.023	-3.698	-2.742	-1.393	-1.227	-0.579		-2.172
Hybrid	-2.419	-2.992	-4.045	-1.928	0.594	1.483	0.729	1.605	2.172	
					Panel B: DJIA					
RW125		-0.971	-2.294	1.892	2.981	3.051	3.132	3.590	3.332	1.840
RW250	0.971		-2.527	1.954	2.844	2.968	3.640	3.732	3.311	2.043
RW500	2.294	2.527		2.891	3.717	3.852	4.680	4.679	4.195	3.093
G-N	-1.892	-1.954	-2.891		3.705	2.900	0.765	1.305	3.236	-0.459
G-Skt	-2.981	-2.844	-3.717	-3.705		1.421	-0.706	-0.291	2.335	-2.666
G-EDF	-3.051	-2.968	-3.852	-2.900	-1.421		-1.022	-0.705	2.213	-2.693
FZ-2F	-3.132	-3.640	-4.680	-0.765	0.706	1.022		1.344	1.740	-1.229
FZ-1F	-3.590	-3.732	-4.679	-1.305	0.291	0.705	-1.344		1.539	-1.943
G-FZ	-3.332	-3.311	-4.195	-3.236	-2.335	-2.213	-1.740	-1.539		-3.127
Hybrid	-1.840	-2.043	-3.093	0.459	2.666	2.693	1.229	1.943	3.127	

Table continued on next page.

Table S6 (cont'd): Diebold-Mariano t-statistics on average out-of-sample loss differences for the S&P 500, DJIA, NIKKEI and FTSE100 (alpha=0.025)

	RW125	RW250	RW500	G-N	G-Skt	G-EDF	FZ-2F	FZ-1F	G-FZ	Hybrid
					Panel C: NIKKEI					
RW125		0.010	-0.901	3.956	3.896	3.944	3.703	3.093	3.895	3.829
RW250	-0.010		-1.486	4.105	4.149	4.177	3.544	3.340	4.136	4.102
RW500	0.901	1.486		3.935	3.999	4.012	3.886	3.441	3.980	3.996
G-N	-3.956	-4.105	-3.935		1.799	2.032	-2.541	-2.010	2.052	-0.226
G-Skt	-3.896	-4.149	-3.999	-1.799		-0.785	-2.726	-2.532	-0.310	-0.977
G-EDF	-3.944	-4.177	-4.012	-2.032	0.785		-2.741	-2.499	0.459	-0.903
FZ-2F	-3.703	-3.544	-3.886	2.541	2.726	2.741		1.481	2.687	2.739
FZ-1F	-3.093	-3.34	-3.441	2.010	2.532	2.499	-1.481		2.454	2.971
G-FZ	-3.895	-4.136	-3.98	-2.052	0.310	-0.459	-2.687	-2.454		-0.919
Hybrid	-3.829	-4.102	-3.996	0.226	0.977	0.903	-2.739	-2.971	0.919	
					Panel D: FTSE					
RW125		-1.557	-3.197	2.938	3.467	3.157	-1.683	2.978	2.570	2.173
RW250	1.557		-2.864	3.646	4.172	3.863	-0.758	3.788	3.355	2.985
RW500	3.197	2.864		4.350	4.789	4.532	1.179	4.688	4.173	3.972
G-N	-2.938	-3.646	-4.350		4.520	3.634	-3.549	-0.239	-0.340	-2.352
G-Skt	-3.467	-4.172	-4.789	-4.520		-4.471	-3.863	-1.996	-3.05	-3.991
G-EDF	-3.157	-3.863	-4.532	-3.634	4.471		-3.686	-0.949	-1.612	-3.218
FZ-2F	1.683	0.758	-1.179	3.549	3.863	3.686		3.924	3.468	3.271
FZ-1F	-2.978	-3.788	-4.688	0.239	1.996	0.949	-3.924		0.046	-1.602
G-FZ	-2.570	-3.355	-4.173	0.340	3.050	1.612	-3.468	-0.046		-2.354
Hybrid	-2.173	-2.985	-3.972	2.352	3.991	3.218	-3.271	1.602	2.354	

Notes: This table presents t-statistics from Diebold-Mariano tests comparing the average losses, using the FZ0 loss function, over the out-of-sample period from January 2000 to December 2016, for ten different forecasting models. A positive value indicates that the row model has higher average loss than the column model. Values greater than 1.96 in absolute value indicate that the average loss difference is significantly different from zero at the 95% confidence level. Values along the main diagonal are all identically zero and are omitted for interpretability. The first three rows correspond to rolling window forecasts, the next three rows correspond to GARCH forecasts based on different models for the standardized residuals, and the last four rows correspond to models introduced in Section 2 of the main paper.

References

[1] Carrasco, M. and X. Chen, 2002, Mixing and moment properties of various GARCH and stochastic volatility models, *Econometric Theory*, 18(1), 17-39.

[2] Davidson, J. 1994, *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press.

[3] Lumsdaine, R., 1996, Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models, *Econometrica*, 64, 575–596.