Technical Appendix to

Maximum Likelihood Estimation for Score-Driven Time Series Models

Contents

B Proofs of Remaining Results in the Main Paper

B.1 Two propositions on SE and the existence of moments

We start with the two propositions used in the proof of Propositions 3.1 and 3.3. These propositions encompass the score-driven model and are written for the case of general random sequences $\{x_t(\boldsymbol{\theta},\bar{x})\}_{t\in\mathbb{N}}$ taking values in $\mathcal{X}\subseteq\mathbb{R}$, where $x_t(\boldsymbol{\theta},\bar{x})$ is generated by a stochastic recurrence equation of the form

$$x_{t+1}(\boldsymbol{\theta},\bar{x})=\phi\big(x_t(\boldsymbol{\theta},\bar{x}),v_t,\boldsymbol{\theta}\big), \tag{B.1}$$

where $\bar{x}\in\mathcal{X}$ is a fixed initialization value at $t=1$, $\phi:\mathcal{X}\times\mathcal{V}\times\Theta\to\mathcal{X}$ is a continuous map, $\mathcal{X}$ is a convex set $\mathcal{X}\subseteq\mathcal{X}^*\subseteq\mathbb{R}$, and $\boldsymbol{\theta}\in\Theta$ is a static parameter vector. For the results that follow we define the supremum

$$r_t^k(\boldsymbol{\theta}):=\sup_{(x,x')\in\mathcal{X}^*\times\mathcal{X}^*:x\neq x'}\frac{|\phi(x,v_t,\boldsymbol{\theta})-\phi(x',v_t,\boldsymbol{\theta})|^k}{|x-x'|^k},\quad k\geq 0.$$

Moreover, for random sequences $\{x_{1,t}\}_{t\in\mathbb{Z}}$ and $\{x_{2,t}\}_{t\in\mathbb{Z}}$, we say that $x_{1,t}$ converges exponentially fast almost surely (e.a.s.) to $x_{2,t}$ if there exists a constant $c>1$ such that $c^t\,\|x_{1,t}-x_{2,t}\|\xrightarrow{a.s.}0$; see also Straumann and Mikosch (2006) (hereafter referred to as SM06).

Proposition TA.1. *For every $\boldsymbol{\theta}\in\Theta$, let $\{v_t\}_{t\in\mathbb{Z}}$ be an i.i.d. sequence and assume $\exists\ \bar{x}\in\mathcal{X}$ such that*

(i) $\mathbb{E}\log^+|\phi(\bar{x},v_1,\boldsymbol{\theta})-\bar{x}|<\infty$;

(ii) $\mathbb{E}\ \log r_1^1(\boldsymbol{\theta})\ <\ 0$.

Then $\{x_t(\boldsymbol{\theta},\bar{x})\}_{t\in\mathbb{N}}$ converges e.a.s. to a unique SE solution $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ for every $\boldsymbol{\theta}\in\Theta$ as $t\to\infty$. If furthermore, for every $\boldsymbol{\theta}\in\Theta\ \exists\ n>0$ such that

(iii) $\|\phi(\bar{x},v_1,\boldsymbol{\theta})\|_n<\infty$;

(iv) $\mathbb{E}r_1^n(\boldsymbol{\theta})<1$;

then $\mathbb{E}|x_t(\boldsymbol{\theta})|^n<\infty\ \forall\ \boldsymbol{\theta}\in\Theta$.

Proof of Proposition TA.1. Step 1, SE: The assumption that $\{v_t\}_{t\in\mathbb{Z}}$ is i.i.d. and therefore SE $\forall\ \boldsymbol{\theta}\in\Theta$ together with the continuity of ϕ on $\mathcal{X}\times\mathcal{V}\times\Theta$ (and resulting measurability w.r.t. the Borel σ-algebra) implies that $\{\phi_t:=\phi(\cdot,v_t,\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is SE for every $\boldsymbol{\theta}\in\Theta$ by Krengel (1985, Proposition 4.3). Condition C1 in Bougerol (1993, Theorem 3.1) is immediately implied by assumption *(i)* for every $\boldsymbol{\theta}\in\Theta$. Condition C2 in Bougerol (1993, Theorem 3.1) is implied, for every $\boldsymbol{\theta}\in\Theta$, by condition *(ii)* since for every $\boldsymbol{\theta}\in\Theta$,

$$\mathbb{E}\log\sup_{(x,x')\in\mathcal{X}\times\mathcal{X}:x\neq x'}\frac{|\phi(x,v_t,\boldsymbol{\theta})-\phi(x',v_t,\boldsymbol{\theta})|}{|x-x'|}=\mathbb{E}\log r_t^1(\boldsymbol{\theta})<0.$$

Also due to the stationarity of $\{v_t\}$ we have $\mathbb{E}\log r_t^1(\boldsymbol{\theta})=\mathbb{E}\log r_1^1(\boldsymbol{\theta})$. As a result, for every $\boldsymbol{\theta}\in\Theta$, $\{x_t(\boldsymbol{\theta},\bar{x})\}_{t\in\mathbb{N}}$ converges to an SE solution $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$. Uniqueness and e.a.s. convergence are obtained by Straumann and Mikosch (2006, Theorem 2.8).

Step 2, moment bounds: For $n \geq 1$ the moment bounds are obtained by first noting that $|x_t(\boldsymbol{\theta})|$ can be bounded as follows:

$$\begin{aligned}
|x_t(\boldsymbol{\theta})| &= |\phi(x_{t-1}(\boldsymbol{\theta}), v_{t-1}, \boldsymbol{\theta})| \\
&\leq |\phi(x_{t-1}(\boldsymbol{\theta}, \bar{x}), v_{t-1}, \boldsymbol{\theta}) - \phi(\bar{x}, v_{t-1}, \boldsymbol{\theta})| + |\phi(\bar{x}, v_{t-1}, \boldsymbol{\theta})| \\
&\leq |x_{t-1}(\boldsymbol{\theta}) - \bar{x}| \times \frac{|\phi(x_{t-1}(\boldsymbol{\theta}), v_{t-1}, \boldsymbol{\theta}) - \phi(\bar{x}, v_{t-1}, \boldsymbol{\theta})|}{|x_{t-1}(\boldsymbol{\theta}) - \bar{x}|} + |\phi(\bar{x}, v_{t-1}, \boldsymbol{\theta})| \\
&\leq r^1_{t-1}(\boldsymbol{\theta})\ |x_{t-1}(\boldsymbol{\theta}) - \bar{x}| + |\phi(\bar{x}, v_{t-1}, \boldsymbol{\theta})| \\
&\leq r^1_{t-1}(\boldsymbol{\theta})\ |x_{t-1}(\boldsymbol{\theta})| + r^1_{t-1}(\boldsymbol{\theta})\ |\bar{x}| + |\phi(\bar{x}, v_{t-1}, \boldsymbol{\theta})|\,,
\end{aligned}$$

so if we keep unfolding this recursion backwards k times, we obtain

$$\begin{aligned}
|x_t(\boldsymbol{\theta})| \ &\leq \ \left(\prod_{j=1}^{k} r^1_{t-j}(\boldsymbol{\theta})\right)\ |x_{t-k}(\boldsymbol{\theta})| \\
&\quad + \sum_{i=1}^{k} \left(\prod_{j=1}^{i-1} r^1_{t-j}(\boldsymbol{\theta})\right) \left(|\phi(\bar{x}, v_{t-i}, \boldsymbol{\theta})| + r^1_{t-i}(\boldsymbol{\theta})\ |\bar{x}|\right)\,.
\end{aligned} \tag{B.2}$$

Because for every $\boldsymbol{\theta} \in \Theta$, $\{r^1_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is an i.i.d. (and therefore SE) sequence of nonnegative random variables with $\mathbb{E}\log r^1_1(\boldsymbol{\theta}) < 0$ (by condition *(ii)*), it follows from Lemma 2.4 of Straumann and Mikosch (2006) that for every t:

$$\prod_{j=1}^{k} r^1_{t-j}(\boldsymbol{\theta}) \ \stackrel{e.a.s.}{\to}\ 0\,, \qquad \text{as } k \to \infty\,.$$

Using this result and the fact that $\{|x_{t-k}(\boldsymbol{\theta})|\}_{k\in\mathbb{Z}}$ is SE by the first part of this proposition and Krengel (1985, Proposition 4.3), it now follows that for large enough k, we have that the first term of (B.2) is smaller than 1 almost surely. So there exists a large k such that

$$\left(\prod_{j=1}^{k} r^1_{t-j}(\boldsymbol{\theta})\right)\ |x_{t-k}(\boldsymbol{\theta})| < 1\,, \quad \text{a.s.} \tag{B.3}$$

Now use that for every $\boldsymbol{\theta} \in \Theta$ we have $\mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty$ if and only if $\|x_t(\boldsymbol{\theta})\|_n < \infty$. For now we consider the case where $n \geq 1$, meaning that we can use the sub-additivity of $\|\cdot\|_n$, because it is a norm. By combining (B.2) and (B.3) for some large enough k and by taking the norm $\|\cdot\|_n$, we obtain

$$\begin{aligned}
\|x_t(\boldsymbol{\theta})\|_n \ &\leq\ 1 + \sum_{i=1}^{k} \left\| \left(\prod_{j=1}^{i-1} r^1_{t-j}(\boldsymbol{\theta})\right) |\phi(\bar{x}, v_{t-i}, \boldsymbol{\theta})| \right\|_n + \sum_{i=1}^{k} \left\| \left(\prod_{j=1}^{i-1} r^1_{t-j}(\boldsymbol{\theta})\right) r^1_{t-i}(\boldsymbol{\theta})\ |\bar{x}| \right\|_n \\
&=\ 1 + \sum_{i=1}^{k} \left(\prod_{j=1}^{i-1} (\mathbb{E}r^n_{t-j}(\boldsymbol{\theta}))^{1/n}\right) \|\phi(\bar{x}, v_{t-i}, \boldsymbol{\theta})\|_n \\
&\quad + \sum_{i=1}^{k} \left(\prod_{j=1}^{i-1} (\mathbb{E}r^n_{t-j}(\boldsymbol{\theta}))^{1/n}\right) (\mathbb{E}r^n_{t-i}(\boldsymbol{\theta}))^{1/n}\ |\bar{x}| \\
&=\ 1 + (\|\phi(\bar{x}, v_t, \boldsymbol{\theta})\|_n + \bar{c}_n\ |\bar{x}|) \sum_{i=1}^{k} (\bar{c}_n)^{i-1} \\
&\leq\ 1 + \frac{\|\phi(\bar{x}, v_t, \boldsymbol{\theta})\|_n + \bar{c}_n\ |\bar{x}|}{1 - \bar{c}_n} \ <\ \infty\,, \quad \forall\, \boldsymbol{\theta} \in \Theta\,,
\end{aligned}$$

where $\bar{c}_n := (\mathbb{E}r_t^n(\boldsymbol{\theta}))^{1/n} = (\mathbb{E}r_1^n(\boldsymbol{\theta}))^{1/n} < 1$ by assumption *(iv)* and the stationarity of $\{v_t\}$. The first inequality follows from the sub-additivity of the norm and the first equality holds by the serial independence of the elements of the sequence $\{v_t\}$, which also implies the serial independence of $\{r_t^1\}$. The final quantity is finite because $\|\phi(\bar{x}, v_t, \boldsymbol{\theta})\|_n < \infty$ by condition *(iii)*. This finishes the proof of $\mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty$ for $n \geq 1$

For $0 < n < 1$, the function $\| \cdot \|_n$ is only a pseudo-norm as it is not sub-additive. However, the proof still follows by instead using the metric $\| \cdot \|_n^* := (\| \cdot \|_n)^n$ which is sub-additive; see the C_n inequality in Loève (1977). ■

Proposition TA.1 not only establishes the convergence to a unique SE solution, but also establishes the existence of unconditional moments. The latter property is key to proving the consistency and asymptotic normality of the MLE in Section 4 of the paper. To establish convergence to an SE solution, condition *(ii)* requires the stochastic recurrence equation to be contracting on average. For the subsequent existence of moments, the contraction condition *(iv)*, together with the moment bound in *(iii)*, are sufficient. Note that conditions *(i)*–*(ii)* are implied by *(iii)*–*(iv)*.

Following SM06, we also note that conditions *(i)* and *(ii)* in Proposition TA.1 provide us with an almost sure representation of $x_t(\boldsymbol{\theta}, \bar{x})$ as a measurable function of $\{v_s\}_{s \leq t-1}$. Let $\circ$ denote the composition of maps, e.g.,

$$\phi(\cdot, v_{t-1}, \boldsymbol{\theta}) \circ \phi(\cdot, v_{t-2}, \boldsymbol{\theta}) = \phi\Big(\phi(\cdot, v_{t-2}, \boldsymbol{\theta}) \ , \ v_{t-1} \ , \ \boldsymbol{\theta} \Big).$$

Then we have the following result.

Remark TA.2. Let conditions *(i)* and *(ii)* of Proposition TA.1 hold. Then $x_t(\boldsymbol{\theta})$ admits the following a.s. representation for every $\boldsymbol{\theta} \in \Theta$

$$x_t(\boldsymbol{\theta}) = \lim_{r \to \infty} \phi(\cdot, v_{t-1}, \boldsymbol{\theta}) \circ \phi(\cdot, v_{t-2}, \boldsymbol{\theta}) \circ ... \circ \phi(\cdot, v_{t-r}, \boldsymbol{\theta}),$$

and $x_t(\boldsymbol{\theta})$ is measurable with respect to the σ-algebra generated by $\{v_s\}_{s \leq t-1}$.

Proposition TA.3 deals with sequences $\{x_t(\boldsymbol{\theta}, \bar{x})\}_{t \in \mathbb{N}}$ that, for a given initialization $\bar{x} \in \mathcal{X}$, are generated by

$$x_{t+1}(\boldsymbol{\theta}, \bar{x}) = \phi\big(x_t(\boldsymbol{\theta}, \bar{x}), v_t, \boldsymbol{\theta}\big) \ \forall \ (\boldsymbol{\theta}, t) \in \Theta \times \mathbb{N},$$

where $\phi : \mathcal{X} \times \mathcal{V} \times \Theta \to \mathcal{X}$ is continuous. We have the following proposition.

Proposition TA.3. *Let Θ be compact, $\{v_t\}_{t \in \mathbb{Z}}$ be stationary and ergodic (SE) and assume there exists an $\bar{x} \in \mathcal{X}$, such that*

(i) $\mathbb{E} \log^+ \sup_{\boldsymbol{\theta} \in \Theta} |\phi(\bar{x}, v_t, \boldsymbol{\theta}) - \bar{x}| < \infty$;

(ii) $\mathbb{E} \log \sup_{\boldsymbol{\theta} \in \Theta} r_1^1(\boldsymbol{\theta}) < 0$.

Then $\{x_t(\boldsymbol{\theta}, \bar{x})\}_{t \in \mathbb{N}}$ converges e.a.s. to a unique SE solution $\{x_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ uniformly on Θ as $t \to \infty$.

If furthermore $\exists \ n > 0$ such that

(iii) $\|\phi(\bar{x}, v_t, \cdot)\|_n^\Theta < \infty$;

(iv) $\sup_{\boldsymbol{\theta} \in \Theta} |\phi(x, v, \boldsymbol{\theta}) - \phi(x', v, \boldsymbol{\theta})| \ < \ |x - x'| \ \ \forall \ (x, x', v) \in \mathcal{X} \times \mathcal{X} \times \mathcal{V}$ *with* $x \neq x'$.

then $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |x_t(\boldsymbol{\theta})|^n < \infty$.

The contraction condition *(iv)* in Proposition TA.3 is stricter than the similar condition *(iv)* in Proposition TA.1. Rather than only requiring the contraction property to hold in expectation, condition *(iv)* in Proposition TA.3 holds for all $v \in \mathcal{V}$.

Again, we note that conditions *(i)* and *(ii)* in Proposition TA.3 provide us with an almost sure representation of $x_t(\boldsymbol{\theta})$.

Remark TA.4. Let conditions *(i)* and *(ii)* of Proposition TA.3 hold. Then $x_t(\boldsymbol{\theta})$ admits the following a.s. representation for every $\boldsymbol{\theta} \in \Theta$

$$x_t(\boldsymbol{\theta}) = \lim_{r\to\infty} \phi\big(\cdot, v_{t-1}, \boldsymbol{\theta}\big) \circ \phi\big(\cdot, v_{t-2}, \boldsymbol{\theta}\big) \circ \ldots \circ \phi\big(\cdot, v_{t-r}, \boldsymbol{\theta}\big)$$

and $x_t(\boldsymbol{\theta})$ is measurable with respect to the σ-algebra generated by $\{v_s\}_{s\le t-1}$.

Proof of Proposition TA.3. Step 0, additional notation: Following Straumann and Mikosch (2006, Proposition 3.12), the uniform convergence of the process $\sup_{\boldsymbol{\theta}\in\Theta} |x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta})| \overset{e.a.s.}{\to} 0$ is obtained by appealing to Bougerol (1993, Theorem 3.1) using sequences of random functions $\{x_t(\cdot, \bar{x})\}_{t\in\mathbb{N}}$ rather than sequences of real numbers. This change is subtle in the notation, but important. We refer to SM06 for details.

The elements $x_t(\cdot, \bar{x})$ are random functions that take values in the separable Banach space $\mathcal{X}_\Theta \subseteq (\mathbb{C}(\Theta, \mathcal{X}), \|\cdot\|^\Theta)$, where $\|x_t(\cdot)\|_n^\Theta \equiv (\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x_t(\boldsymbol{\theta})|^n)^{1/n}$ and $\|x_t(\cdot)\|^\Theta \equiv \sup_{\boldsymbol{\theta}\in\Theta} |x_t(\boldsymbol{\theta})|$. The functions $x_t(\cdot, \bar{x})$ are generated by

$$x_t(\cdot, \bar{x}) = \phi^*(x_{t-1}(\cdot, \bar{x}), v_{t-1}, \cdot) \;\forall\; t \in \{2, 3, \ldots\},$$

with starting function $x_1(\emptyset, \boldsymbol{\theta}, \bar{x}) = \bar{x} \;\forall\; \boldsymbol{\theta} \in \Theta$, and where $\{\phi^*(\cdot, v_t, \cdot)\}_{t\in\mathbb{Z}}$ is a sequence of stochastic recurrence equations $\phi^* : \mathbb{C}(\Theta) \times \Theta \to \mathbb{C}(\Theta) \;\forall\; t$ as in Straumann and Mikosch (2006, Proposition 3.12). Note the subtle but important difference between $\phi^*(\cdot, v_t, \cdot) : \mathbb{C}(\Theta) \times \Theta \to \mathbb{C}(\Theta)$ and $\phi(\cdot, v_t, \cdot) : \mathcal{X} \times \Theta \to \mathcal{X}$ as alluded to earlier.

Step 1, SE: With the above notation in place, we now first prove the SE part of the proposition. The assumption that $\{v_t\}_{t\in\mathbb{Z}}$ is SE together with the continuity of ϕ on $\mathcal{X} \times \mathcal{V} \times \Theta$ implies that $\{\phi^*(\cdot, v_t, \cdot)\}_{t\in\mathbb{Z}}$ is SE. Condition C1 in Bougerol (1993, Theorem 3.1) is now implied directly by condition *(i)*, since there exists a function $\bar{x}_\Theta \in \mathbb{C}(\Theta)$ with $\bar{x}_\Theta(\boldsymbol{\theta}) = \bar{x} \;\forall\; \boldsymbol{\theta} \in \Theta$ that satisfies $\mathbb{E}\log^+ \|\phi^*(\bar{x}_\Theta(\cdot), v_t, \cdot) - \bar{x}_\Theta(\cdot)\|^\Theta = \mathbb{E}\log^+ \sup_{\boldsymbol{\theta}\in\Theta} |\phi(\bar{x}, v_t, \boldsymbol{\theta}) - \bar{x}| < \infty$.

Condition C2 in Bougerol (1993, Theorem 3.1) is directly implied by condition *(ii)*, since

$$\mathbb{E}\log \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \frac{\|\phi^*(\bar{x}_\Theta(\cdot), v_t, \cdot) - \phi^*(\bar{x}'_\Theta(\cdot), v_t, \cdot)\|^\Theta}{\|\bar{x}_\Theta(\cdot) - \bar{x}'_\Theta(\cdot)\|^\Theta} =$$

$$\mathbb{E}\log \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \frac{\sup_{\boldsymbol{\theta}\in\Theta} |\phi(\bar{x}_\Theta(\boldsymbol{\theta}), v_t, \boldsymbol{\theta}) - \phi(\bar{x}'_\Theta(\boldsymbol{\theta}), v_t, \boldsymbol{\theta})|}{\|\bar{x}_\Theta(\cdot) - \bar{x}'_\Theta(\cdot)\|^\Theta} =$$

$$\mathbb{E}\log \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \sup_{\boldsymbol{\theta}\in\Theta} \frac{|\phi(\bar{x}_\Theta(\boldsymbol{\theta}), v_t, \boldsymbol{\theta}) - \phi(\bar{x}'_\Theta(\boldsymbol{\theta}), v_t, \boldsymbol{\theta})|}{\|\bar{x}_\Theta(\cdot) - \bar{x}'_\Theta(\cdot)\|^\Theta} =$$

$$\mathbb{E}\log \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \sup_{\boldsymbol{\theta}\in\Theta | \bar{x}_\Theta(\boldsymbol{\theta}) \neq \bar{x}'_\Theta(\boldsymbol{\theta})} \frac{|\phi(\bar{x}_\Theta(\boldsymbol{\theta}), v_t, \boldsymbol{\theta}) - \phi(\bar{x}'_\Theta(\boldsymbol{\theta}), v_t, \boldsymbol{\theta})|}{\|\bar{x}_\Theta(\cdot) - \bar{x}'_\Theta(\cdot)\|^\Theta} \leq$$

$$\mathbb{E}\log \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \sup_{\boldsymbol{\theta}\in\Theta | \bar{x}_\Theta(\boldsymbol{\theta}) \neq \bar{x}'_\Theta(\boldsymbol{\theta})} \frac{|\phi(\bar{x}_\Theta(\boldsymbol{\theta}), v_t, \boldsymbol{\theta}) - \phi(\bar{x}'_\Theta(\boldsymbol{\theta}), v_t, \boldsymbol{\theta})|}{|\bar{x}_\Theta(\boldsymbol{\theta}) - \bar{x}'_\Theta(\boldsymbol{\theta})|} \times \frac{|\bar{x}_\Theta(\boldsymbol{\theta}) - \bar{x}'_\Theta(\boldsymbol{\theta})|}{\|\bar{x}_\Theta(\cdot) - \bar{x}'_\Theta(\cdot)\|^\Theta} \leq$$

$$\mathbb{E}\log \Bigg(\sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \sup_{\boldsymbol{\theta}\in\Theta | \bar{x}_\Theta(\boldsymbol{\theta}) \neq \bar{x}'_\Theta(\boldsymbol{\theta})} \frac{|\phi(\bar{x}_\Theta(\boldsymbol{\theta}), v_t, \boldsymbol{\theta}) - \phi(\bar{x}'_\Theta(\boldsymbol{\theta}), v_t, \boldsymbol{\theta})|}{|\bar{x}_\Theta(\boldsymbol{\theta}) - \bar{x}'_\Theta(\boldsymbol{\theta})|} \Bigg) \times \Bigg(\sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \sup_{\boldsymbol{\theta}\in\Theta} \frac{|\bar{x}_\Theta(\boldsymbol{\theta}) - \bar{x}'_\Theta(\boldsymbol{\theta})|}{\|\bar{x}_\Theta(\cdot) - \bar{x}'_\Theta(\cdot)\|^\Theta} \Bigg) \leq$$

$$\mathbb{E}\log \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \sup_{\boldsymbol{\theta}\in\Theta} \sup_{x \neq x'} \frac{|\phi(x, v_t, \boldsymbol{\theta}) - \phi(x', v_t, \boldsymbol{\theta})|}{|x - x'|} =$$

$$\mathbb{E}\log \sup_{\boldsymbol{\theta}\in\Theta} \sup_{x \neq x'} \frac{|\phi(x, v_t, \boldsymbol{\theta}) - \phi(x', v_t, \boldsymbol{\theta})|}{|x - x'|} = \mathbb{E}\log \sup_{\boldsymbol{\theta}\in\Theta} r_t^1(\boldsymbol{\theta}) =$$

$$\mathbb{E}\log \sup_{\boldsymbol{\theta}\in\Theta} r_1^1(\boldsymbol{\theta}) < 0.$$

As a result, $\{x_t(\cdot, \bar{x})\}_{t\in\mathbb{N}}$ converges to an SE solution $\{x_t(\cdot)\}_{t\in\mathbb{Z}}$ in $\|\cdot\|^\Theta$-norm. Uniqueness and e.a.s. convergence is obtained in Straumann and Mikosch (2006, Theorem 2.8), such that $\sup_{\boldsymbol{\theta}\in\Theta} |x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$.

Step 2, moment bounds: We use a similar argument as in the proof of Proposition TA.1. First consider $n \geq 1$. Using condition *(iv.a)*, define $\bar{c} < 1$ such that $\sup_{\boldsymbol{\theta}\in\Theta} |\phi(x, v, \boldsymbol{\theta}) - \phi(x', v, \boldsymbol{\theta})| \leq \bar{c}\,|x - x'|$ for all (x, x', v). and note that we can bound $\|x_t(\cdot)\|^\Theta$ as follows:

$$\begin{aligned}
\|x_t(\cdot)\|^\Theta &= \|\phi^*(x_{t-1}(\cdot), v_{t-1}, \cdot)\|^\Theta \\
&\leq \|\phi^*(x_{t-1}(\cdot), v_{t-1}, \cdot) - \phi(\bar{x}, v_{t-1}, \cdot)\|^\Theta + \|\phi(\bar{x}, v_{t-1}, \cdot)\|^\Theta \\
&\leq \sup_{\boldsymbol{\theta}\in\Theta} |\phi(x_{t-1}(\boldsymbol{\theta}), v_{t-1}, \boldsymbol{\theta}) - \phi(\bar{x}, v_{t-1}, \boldsymbol{\theta})| + \|\phi(\bar{x}, v_{t-1}, \cdot)\|^\Theta \\
&\leq \bar{c} \cdot \sup_{\boldsymbol{\theta}\in\Theta} |x_{t-1}(\boldsymbol{\theta}) - \bar{x}| + \|\phi(\bar{x}, v_{t-1}, \cdot)\|^\Theta \\
&\leq \bar{c} \cdot \|x_{t-1}(\cdot)\|^\Theta + \bar{c}\,|\bar{x}| + \|\phi(\bar{x}, v_{t-1}, \cdot)\|^\Theta .
\end{aligned}$$

Unfolding this recursion k steps backwards, leads to

$$\begin{aligned}
\|x_t(\cdot)\|^\Theta &\leq (\bar{c})^k \cdot \|x_{t-k}(\cdot)\|^\Theta + \sum_{i=1}^{k} (\bar{c})^{i-1} \left(\|\phi(\bar{x}, v_{t-i}, \cdot)\|^\Theta + \bar{c}\,|\bar{x}| \right) \\
&\leq 1 + \sum_{i=1}^{k} (\bar{c})^{i-1} \left(\|\phi(\bar{x}, v_{t-i}, \cdot)\|^\Theta + \bar{c}\,|\bar{x}| \right), \qquad \text{(B.4)}
\end{aligned}$$

where the final inequality holds for some large enough k, because $(\bar{c})^k$ goes to zero at an exponential rate as $k \to \infty$ and because $\{\|x_{t-k}(\cdot)\|^{\Theta}\}_{k\in\mathbb{Z}}$ is SE by the first part of the proposition and Krengel (1985, Proposition 4.3). Now use that $\sup_t \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x_t(\boldsymbol{\theta}, \bar{x})|^n < \infty$ if and only if $\sup_t \|x_t(\boldsymbol{\theta}, \bar{x})\|_n^{\Theta} < \infty$. Consider taking the norm $\|\cdot\|_n$ of both sides of the inequality (B.4) for some large enough k. Then

$$\begin{aligned} \|x_t(\cdot)\|_n^{\Theta} &\leq 1 + \sum_{i=1}^{k} (\bar{c})^{i-1} \left(\|\phi(\bar{x}, v_{t-i}, \cdot)\|_n^{\Theta} + \bar{c}\,|\bar{x}|\right) \\ &\leq 1 + \frac{\|\phi(\bar{x}, v_t, \cdot)\|_n^{\Theta} + \bar{c}\,|\bar{x}|}{1-\bar{c}} < \infty\,, \end{aligned}$$

where the first inequality holds by the subadditivity of $\|\cdot\|_n$ for $n \geq 1$, the second inequality holds because $\bar{c} < 1$ and $\|\phi(\bar{x}, v_{t-i}, \cdot)\|_n^{\Theta} = \|\phi(\bar{x}, v_t, \cdot)\|_n^{\Theta}$ for every i because $\{v_t\}$ is SE and where the final expression is finite because $\|\phi(\bar{x}, v_t, \cdot)\|_n^{\Theta} < \infty$ by condition *(iii)*. This establishes the result for $n \geq 1$.

When $0 < n < 1$, then $\|\cdot\|_n$ is not sub-additive. Just as in the proof of Proposition TA.1, in this case the metric $\|\cdot\|_n^* := (\|\cdot\|_n)^n$ can be used. This works because $\|\cdot\|_n^*$ is sub-additive (see the C_n inequality in Loève (1977)). ∎

B.2 Proofs of remaining results in the main paper

Proof of Theorem 4.3. Recall that $\hat{f}_t$ denotes the initialized $\hat{f}_t(\boldsymbol{\theta}, \hat{f}_1)$. Assumption 4.2 implies that $\ell_T(\boldsymbol{\theta}, \hat{f}_1)$ is a.s. continuous (a.s.c.) in $\boldsymbol{\theta} \in \Theta$ through continuity of each $\tilde{\ell}_t(\boldsymbol{\theta}, \hat{f}_1) = \ell(f_t, y, \boldsymbol{\theta})$, ensured in turn by the continuous differentiability of $\bar{p}$, $\bar{g}$, $\bar{g}'$ and the continuity of S_t, the implied a.s.c. of $s(f_t, y; \lambda) = S_t \cdot (\partial \bar{p}_t/\partial f + \partial \log \bar{g}'/\partial f)$ in $(f_t; \lambda)$ and the resulting continuity of $\hat{f}_t$ in $\boldsymbol{\theta}$ as a composition of t continuous maps. The compactness of Θ implies by Weierstrass' theorem that the arg max set is non-empty a.s. and hence that $\hat{\boldsymbol{\theta}}_T$ exists a.s. $\forall\, T \in \mathbb{N}$. Similarly, Assumption 4.2 implies that $\ell_T(\boldsymbol{\theta}, \hat{f}_1) = \ell\big(\{y_t\}_{t=1}^T, \{\hat{f}_t\}_{t=1}^T, \boldsymbol{\theta}\big)$ is continuous in y_t $\forall\, \boldsymbol{\theta} \in \Theta$ and hence measurable w.r.t. a Borel σ-algebra. The measurability of $\hat{\boldsymbol{\theta}}_T$ follows from White (1994, Theorem 2.11) or Gallant and White (1988, Lemma 2.1, Theorem 2.2). ∎

The following two lemmas support the proof of Theorem 4.6.

Lemma TA.5. *Under the conditions of Theorem 4.6,* $\sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}, \hat{f}_1) - \ell_T(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$.

Proof. Note that instead of considering the average log likelihood $\ell_T(\boldsymbol{\theta}, \hat{f}_1)$ it is sufficient to show that $\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{\ell}_t(\boldsymbol{\theta}, \hat{f}_1) - \tilde{\ell}_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$, where $\tilde{\ell}_t(\cdot, \hat{f}_1) = \ell(f_t(\cdot, \hat{f}_1), y_t, \cdot)$ is the individual log likelihood. The expression for the likelihood in (2.5) and the differentiability conditions in Assumption 4.2 ensure that $\tilde{\ell}_t(\cdot, \hat{f}_1)$ is continuous in $(\hat{f}_t, y_t)$. All the assumptions of Proposition 3.3 relevant for the process $\{f_t\}$ hold as well. To see this, note that

- the compactness of Θ is imposed in Assumption 4.1;
- the moment bound $\mathbb{E}|y_t|^{n_y} < \infty$ is ensured in the statement of Theorem 4.6;
- the differentiability $s \in \mathbb{C}^{(2,0,2)}(\mathcal{F} \times \mathcal{Y} \times \Lambda)$ is implied by $\bar{g} \in \mathbb{C}^{(2,0)}(\mathcal{F} \times \mathcal{Y})$, $\bar{p} \in \mathbb{C}^{(2,2)}(\tilde{\mathcal{U}} \times \Lambda)$, and $S \in \mathbb{C}^{(2,2)}(\mathcal{F} \times \Lambda)$);
- and finally, conditions *(i)*-*(v)* in Proposition 3.3 are ensured by Assumption 4.4.

As a result, there exists a unique SE sequence $\{f_t\}_{t\in\mathbb{Z}}$ such that $\sup_{\boldsymbol{\theta}\in\Theta}|\hat{f}_t - f_t| \overset{e.a.s.}{\to} 0 \;\forall \hat{f}_1 \in \mathcal{F}$. Because $\tilde{\ell}_t$ is differentiable in f_t by assumption, the mean value theorem implies that

$$\sup_{\boldsymbol{\theta}\in\Theta}|\tilde{\ell}_t(\boldsymbol{\theta},\hat{f}_1) - \tilde{\ell}_t(\boldsymbol{\theta})| \leq \sup_{\boldsymbol{\theta}\in\Theta}\sup_{f}|\partial\ell(f,y_t,\lambda)/\partial f| \cdot \sup_{\boldsymbol{\theta}\in\Theta}|\hat{f}_t - f_t|$$

and therefore we obtain the required result $\sup_{\boldsymbol{\theta}\in\Theta}|\tilde{\ell}_t(\boldsymbol{\theta},\hat{f}_1) - \tilde{\ell}_t(\boldsymbol{\theta})| \overset{e.a.s.}{\to} 0$ by Lemma 2.1 in SM06 since $\sup_{\boldsymbol{\theta}\in\Theta}|\hat{f}_t - f_t| \overset{e.a.s.}{\to} 0$ by Proposition 3.3 and $\sup_{\boldsymbol{\theta}\in\Theta}\sup_f|\nabla_t|$ is SE by the continuity of the score which follows from Assumption 4.2, and has a logarithmic moment because $\bar{n}_\nabla > 0$ by Assumption 4.5.

∎

Lemma TA.6. *Under the conditions of Theorem 4.6,* $\sup_{\boldsymbol{\theta}\in\Theta}|\ell_T(\boldsymbol{\theta}) - \ell_\infty(\boldsymbol{\theta})| \overset{a.s.}{\to} 0$.

Proof. We apply the ergodic theorem for separable Banach spaces of Rao (1962) (see also Straumann and Mikosch (2006, Theorem 2.7)) to the sequence $\{\ell_T(\cdot)\}$ with elements taking values in $\mathbb{C}(\Theta)$, so that $\sup_{\boldsymbol{\theta}\in\Theta}|\ell_T(\boldsymbol{\theta}) - \ell_\infty(\boldsymbol{\theta})| \overset{a.s.}{\to} 0$, where $\ell_\infty(\boldsymbol{\theta}) = \mathbb{E}\tilde{\ell}_t(\boldsymbol{\theta}) \;\forall\; \boldsymbol{\theta} \in \Theta$. The ULLN $\sup_{\boldsymbol{\theta}\in\Theta}|\ell_T(\boldsymbol{\theta}) - \mathbb{E}\tilde{\ell}_t(\boldsymbol{\theta})| \overset{a.s.}{\to} 0$ as $T\to\infty$ follows, under a moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\tilde{\ell}_t(\boldsymbol{\theta})| < \infty$, by the SE nature of $\{\ell_T\}_{t\in\mathbb{Z}}$, which is implied by continuity of ℓ on the SE sequence $\{(f_t, y_t)\}_{t\in\mathbb{Z}}$ and Proposition 4.3 in Krengel (1985). Moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\tilde{\ell}_t(\boldsymbol{\theta})| < \infty$ is ensured by $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|f_t|^{n_f} < \infty$, $\mathbb{E}|y_t|^{n_y} < \infty$, and the fact that Assumption 4.5 implies $n_\ell \geq 1$. We stress that Assumption 4.5 can be checked via low-level conditions on n_y and n_f via the moment preserving maps as laid out in Technical Appendix G. ∎

The following lemmas support the proof of Theorem 4.10.

Lemma TA.7. *Under the conditions of Theorem 4.10,*

$$Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0) = \int\!\!\int\left[\int p_y(y|f,\lambda_0)\log\frac{p_y(y|\tilde{f};\lambda)}{p_y(y|f;\lambda_0)}\,\mathrm{d}y\right]\mathrm{d}P_{f_t,\tilde{f}_t}(f,\tilde{f};\boldsymbol{\theta}_0,\boldsymbol{\theta}),$$

for all $(\boldsymbol{\theta}_0,\boldsymbol{\theta}) \in \Theta\times\Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

Proof. Using the observation-driven dynamic structure of the score-driven model, we can substitute the conditioning on $\{y_s\}_{s\leq t-1}$ by the conditioning on f_t, where f_t is generated through the generalized autoregressive score recursion. Under the present conditions, the (non-initialized) limit process $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is a measurable function of $\{y_s\}_{s\leq t-1}$, and hence SE by Krengel's theorem for any $\boldsymbol{\theta}\in\Theta$; see also SM06. By substituting the conditioning, we obtain

$$\begin{aligned} Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0) &= \mathbb{E}\log p_y\Big(y_t\Big|f_t(\boldsymbol{\theta});\lambda\Big) \\ &\qquad\qquad - \mathbb{E}\log p_y\Big(y_t\Big|f_t(\boldsymbol{\theta}_0);\lambda_0\Big) \\ &= \int\!\!\int\!\!\int \log\frac{p_y(y|\tilde{f};\lambda)}{p_y(y|f;\lambda_0)}\,\mathrm{d}P_{y_t,f_t,\tilde{f}_t}(y,f,\tilde{f};\boldsymbol{\theta}_0,\boldsymbol{\theta}), \end{aligned} \tag{B.5}$$

$\forall\;(\boldsymbol{\theta}_0,\boldsymbol{\theta}) \in \Theta\times\Theta : \boldsymbol{\theta}\neq\boldsymbol{\theta}_0$, with $P_{y_t,f_t,\tilde{f}_t}(y,f,\tilde{f};\boldsymbol{\theta}_0,\boldsymbol{\theta})$ denoting the cdf of $(y_t, f_t(\boldsymbol{\theta}_0), \tilde{f}_t(\boldsymbol{\theta}))$. Define the bivariate cdf $P_{f_t,\tilde{f}_t}(f,\tilde{f};\boldsymbol{\theta}_0,\boldsymbol{\theta})$ for the pair $(f_t(\boldsymbol{\theta}_0), \tilde{f}_t(\boldsymbol{\theta}))$. Note that this bivariate cdf depends on $\boldsymbol{\theta}$ through the recursion defining $\tilde{f}_t(\boldsymbol{\theta})$, and on $\boldsymbol{\theta}_0$ through y_{t-1} and $f_t(\boldsymbol{\theta}_0)$. Also note that for any $(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \in \Theta \times \Theta$ this cdf does not depend on the initialization $\hat{f}_1$ because, under the present conditions, the limit criterion is a function of the unique limit SE process $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, and not of the initialized process $\{\hat{f}_t(\boldsymbol{\theta}, \hat{f}_1)\}_{t\in\mathbb{N}}$; see the proof of Theorem 4.6.

We re-write the normalized limit criterion function $Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0)$ by factorizing the joint distribution $P_{y_t,f_t,\tilde{f}_t}(y, f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta})$ as

$$\begin{aligned} P_{y_t,f_t,\tilde{f}_t}(y, f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}) &= P_{y_t|f_t,\tilde{f}_t}(y|f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}) \cdot P_{f_t,\tilde{f}_t}(f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}) \\ &= P_{y_t|f_t}(y|f, \lambda_0) \cdot P_{f_t,\tilde{f}_t}(f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}), \end{aligned}$$

where the second equality holds because under the axiom of correct specification, and conditional on $f_t(\boldsymbol{\theta}_0)$, observed data y_t does not depend on $\tilde{f}_t(\boldsymbol{\theta})$ $\forall$ $(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \in \Theta \times \Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. We also note that the conditional distribution $P_{y_t|f_t}(y|f, \lambda_0)$ has a density $p_y(y|f, \lambda_0)$ defined in equation (2.1). The existence of this density follows because $g(f, \cdot)$ is a diffeomorphism $g(f, \cdot) \in \mathbb{D}(\mathcal{U})$ for every $f \in \mathcal{F}$, i.e., it is continuously differentiable and uniformly invertible with differentiable inverse.

We can now re-write $Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0)$ as

$$\begin{aligned} &Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0) = \\ &\int\int\int \log \frac{p_y(y|\tilde{f}; \lambda)}{p_y(y|f; \lambda_0)} \, \mathrm{d}P_{y_t|f_t}(y|f, \lambda_0) \cdot \mathrm{d}P_{f_t,\tilde{f}_t}(f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}) = \\ &\int\int\left[\int \log \frac{p_y(y|\tilde{f}; \lambda)}{p_y(y|f; \lambda_0)} \, \mathrm{d}P_{y_t|f_t}(y|f, \lambda_0)\right] \mathrm{d}P_{f_t,\tilde{f}_t}(f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}) = \\ &\int\int\left[\int p_y(y|f, \lambda_0) \log \frac{p_y(y|\tilde{f}; \lambda)}{p_y(y|f; \lambda_0)} \, \mathrm{d}y\right] \mathrm{d}P_{f_t,\tilde{f}_t}(f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}), \end{aligned}$$

for all $(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \in \Theta \times \Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. ∎

Lemma TA.8. *Under the conditions of Theorem 4.10, for every $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ there exists a set $YF\tilde{F} \subseteq \mathcal{Y} \times \mathcal{F} \times \tilde{\mathcal{F}}$ with positive probability mass and with orthogonal projections $Y\tilde{F} \subseteq \mathcal{Y} \times \mathcal{F}$, $F\tilde{F} \subseteq \mathcal{F} \times \tilde{\mathcal{F}}$, etc., for which (i)–(ii) hold if $\lambda \neq \lambda_0$, and for which (i)–(iii) hold if $\lambda = \lambda_0$, where*

(i) $p_y(y|f, \lambda_0) > 0$ $\forall$ $(y, f) \in YF$;

(ii) *if* $(\tilde{f}, \lambda) \neq (f, \lambda_0)$, *then* $p_y(y|\tilde{f}; \lambda) \neq p_y(y|f; \lambda_0)$ $\forall$ $(y, f, \tilde{f}) \in YF\tilde{F}$;

(iii) *if* $\lambda = \lambda_0$ *and* $(\omega, \alpha, \beta) \neq (\omega_0, \alpha_0, \beta_0)$, *then* $f \neq \tilde{f}$ *for every* $(f, \tilde{f}) \in F\tilde{F}$.

Proof.
Part (i): The first result follows by noting that under the correct specification axiom, the conditional density $p_y(y|f, \lambda_0)$ is implicitly defined by $y_t(\boldsymbol{\theta}_0) = g(f, u_t)$, $u_t \sim p_u(u_t; \lambda_0)$. Note that $g(f, \cdot)$ is a diffeomorphism $g(f, \cdot) \in \mathbb{D}(\mathcal{U})$ for every $f \in \mathcal{F}_g$ and hence an open map, i.e., $g^{-1}(f, Y) \in \mathcal{T}(\mathcal{U}_g)$ for every $Y \in \mathcal{T}(\mathcal{Y}_g)$ where $\mathcal{T}(\mathbb{A})$ denotes a topology on the set $\mathbb{A}$. Therefore, since $p_u(u; \lambda) > 0$ $\forall$ $(u, \lambda) \in U \times \Lambda$ for some open set $U \subset \mathcal{U}$, which exists by the assumption that u_t has a density with respect to Lebesgue measure. As a result, we obtain that there exists an open set $Y \in \mathcal{T}(\mathcal{Y}_g)$ such that $p_y(y|f, \lambda_0) > 0$ $\forall$ $(y, f) \in Y \times \mathcal{F}_g$, namely the image of any open set $U \subseteq \mathcal{U}$ under $g(f, \cdot)$. Next, $YF\tilde{F}$ can be constructed by taking the union of Y over $F\tilde{F}$ for any $F\tilde{F}$ of positive measure for $\lambda \neq \lambda_0$, and for a set $F\tilde{F}$ satisfying (iii) below if $\lambda = \lambda_0$.

Part (ii): The second result is implied directly by the assumption that $p_y(y|f,\lambda) = p_y(y|f',\lambda')$ almost everywhere in Y for some open set $Y \subset \mathcal{Y}$ if and only if $f = f'$ and $\lambda = \lambda'$. The existence of an open set Y was already argued under (i) above.

Part (iii): The assumptions that $\alpha \neq 0 \; \forall \boldsymbol{\theta} \in \Theta$ (including $\alpha_0 \neq 0$); and that $\partial s(f,y;\lambda)/\partial y \neq 0$ almost everywhere in Y_s for every $(f,\lambda) \in \mathcal{F} \times \Lambda$; together with the fact that u_t has a density, together ensure that both F and $\tilde{F}$ can be chosen as open subsets, i.e., to have multiple different values.

The result is now obtained by a proof by contradiction: if $\lambda = \lambda_0 \wedge (\omega,\alpha,\beta) \neq (\omega_0,\alpha_0,\beta_0)$, but there is no set $YF\tilde{F}$ with positive probability mass satisfying $f \neq \tilde{f} \; \forall \; (f,\tilde{f}) \in F\tilde{F}$, then it must be that $(\omega,\alpha,\beta) = (\omega_0,\alpha_0,\beta_0)$, which is a contradiction.

The proof goes as follows. Let $(\boldsymbol{\theta}_0,\boldsymbol{\theta}) \in \Theta \times \Theta$ be a pair satisfying $\lambda = \lambda_0 \wedge (\omega,\alpha,\beta) \neq (\omega_0,\alpha_0,\beta_0)$. If there is no $YF\tilde{F}$ of positive probability mass with $f \neq \tilde{f}$ for all $(f,\tilde{f}) \in F\tilde{F}$, then it must be that $f = \tilde{f}$ except for a set of zero probability. This implies that $\tilde{f}_t(\boldsymbol{\theta}) \overset{a.s.}{=} f_t(\boldsymbol{\theta}_0)$ for arbitrary t. Putting this into the recurrence equation for both $f_t(\boldsymbol{\theta}_0)$ and $\tilde{f}_t(\boldsymbol{\theta})$ and subtracting the two, we obtain

$$\begin{aligned} 0 &= \phi(f_t(\boldsymbol{\theta}), y_e, \boldsymbol{\theta}) - \phi(f_t(\boldsymbol{\theta}), y_e, \boldsymbol{\theta}_0) \\ &= (\omega - \omega_0) + (\beta - \beta_0) f_t(\boldsymbol{\theta}_0) + (\alpha - \alpha_0) s(f_t(\boldsymbol{\theta}_0), y_t(\boldsymbol{\theta}_0), \lambda_0). \end{aligned} \tag{B.6}$$

Note that $s(f_t(\boldsymbol{\theta}_0), y_t, \lambda_0)$ is not constant in $y_t \in Y$ where Y is an open set, because $\alpha \neq 0 \; \forall \; \boldsymbol{\theta} \in \Theta$ and $\partial s(f,y,\lambda)/\partial y \neq 0$ for every $\lambda \in \Lambda$ and almost every $(y,f) \in \mathcal{Y}_s \times \mathcal{F}_s$. As a result, we must have $\alpha = \alpha_0$ for (B.6) to hold.

Given $\alpha = \alpha_0 \wedge \lambda = \lambda_0$, and given F can be chosen as an open set due to the fact that u_t has a density and $\alpha_0 > 0$, it follows that $\beta = \beta_0$. Given the result for α and β, the result $\omega = \omega_0$ follows directly from (B.6), which establishes the contradiction and the result. ■

C Derivative Expressions for the Main Example

In this part, we provide some of the technical details of the main example of the paper, including the detailed expressions for of the required derivatives.

Let $\{u_t\}_{t\in\mathbb{N}}$ be i.i.d. Student's t distributed noise with λ degrees of freedom. Consider the model $y_t = f_t^{1/2} u_t$ as in Creal et al. (2011, 2013). Following Creal et al. (2011, 2013), we scale the score by (a time-invariant multiple of) the conditional Fisher information, which in this case amounts to setting $S(f_t;\lambda) = 2f_t^2$.

The following set of derivatives is straightforward (though tedious) to compute, either by hand or by a symbolic computation package such as Maple or Mathematica.

$$
\begin{aligned}
\bar{p}_t &= \log \frac{\Gamma\left(\frac{\lambda+1}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)\sqrt{\pi\lambda}} - \tfrac{1}{2}(\lambda+1)\log\left(1+\frac{y_t^2}{\lambda f_t}\right),\\
\log \bar{g}_t' &= -\tfrac{1}{2}\log f_t\\
\nabla_t &= \frac{(1+\lambda^{-1})y_t^2/(2f_t^2)}{1+y_t^2/(\lambda f_t)} - \tfrac{1}{2}f_t^{-1},\\
s_t &= \frac{(1+\lambda^{-1})y_t^2}{1+y_t^2/(\lambda f_t)} - f_t,\\
\partial s_t/\partial f_t &= \frac{(1+\lambda^{-1})y_t^4/(\lambda f_t^2)}{\left(1+y_t^2/(\lambda f_t)\right)^2} - 1,\\
s_{u,t} &= \left(\frac{(1+\lambda^{-1})u_t^2}{1+\lambda^{-1}u_t^2} - 1\right) f_t,\\
\partial s_{u,t}/\partial f_t &= \frac{(1+\lambda^{-1})u_t^2}{1+\lambda^{-1}u_t^2} - 1.\\
\partial s_t/\partial \lambda &= \frac{y_t^2}{(\lambda + y_t^2/f_t)} - \frac{(1+\lambda)y_t^2}{\left(\lambda+y_t^2/f_t\right)^2} = \frac{\left((y_t^2/f_t)^2 - (y_t^2/f_t)\right)}{\left(\lambda+y_t^2/f_t\right)^2}\cdot f_t,\\
\partial^2 s_t/\partial\lambda^2 &= \frac{-2\left((y_t^2/f_t)^2-(y_t^2/f_t)\right)}{\left(\lambda+y_t^2/f_t\right)^3}\cdot f_t,\\
\partial^2 s_t/\partial\lambda\partial f_t &= \frac{(1+\lambda)(y_t^4/f_t^2)}{\left(\lambda+y_t^2/f_t\right)^2} = \frac{(y_t^4/f_t^2)}{\left(\lambda+y_t^2/f_t\right)^2} - \frac{2(1+\lambda)(y_t^4/f_t^2)}{\left(\lambda+y_t^2/f_t\right)^3},\\
\partial^2 s_t/\partial f_t^2 &= \frac{-2(1+\lambda^{-1})(y_t^4/f_t^2)}{\left(1+y_t^2/(\lambda f_t)\right)^3}\cdot\frac{1}{\lambda f_t},\\
\partial^3 s_t/\partial\lambda^3 &= \frac{6\left((y_t^2/f_t)^2-(y_t^2/f_t)\right)}{\left(\lambda+y_t^2/f_t\right)^4}\cdot f_t,\\
\partial^3 s_t/\partial\lambda^2\partial f_t &= \frac{2(y_t^2/f_t)^2\left(\lambda+3-2(y_t^2/f_t)\right)}{\left(\lambda+y_t^2/f_t\right)^4},\\
\partial^3 s_t/\partial\lambda\partial f_t^2 &= \frac{2(y_t^2/f_t)^2\left(\lambda^2+2\lambda-(1+2\lambda)(y_t^2/f_t)\right)}{\left(\lambda+y_t^2/f_t\right)^4}\cdot\frac{1}{f_t},\\
\partial^3 s_t/\partial f_t^3 &= \frac{6(1+\lambda^{-1})y_t^4/(\lambda f_t^2)}{\left(1+y_t^2/(\lambda f_t)\right)^4}\cdot\frac{1}{f_t^2}.
\end{aligned}
$$

We obtain directly that

- $|s_t| \leq \sup_{y_t} |s_t| < c_1 \cdot |f_t|$ for some constant c_1, and thus $n_s \leq n_f$.
- $\sup_{y_t} |\partial s_t/\partial\lambda| \leq c_1 \cdot |f_t|$ and thus $n_s^\lambda \leq n_f$.
- $\sup_{y_t} |\partial s_t/\partial f_t| \leq c_1$ and thus $\bar{n}_s^f \to \infty$.
- $\sup_{y_t} |\partial^2 s_t/\partial f_t \partial\lambda| \leq c_1$ and thus $\bar{n}_s^{f\lambda} \to \infty$.
- $\sup_{y_t} |\partial^2 s_t/\partial f_t^2| \leq c_1 f_t^{-1} \leq c_1/\omega$ and thus $\bar{n}_s^{ff} \to \infty$.
- $\sup_{y_t} |\partial^2 s_t/\partial\lambda^2| \leq c_1 f_t$ and thus $n_s^{\lambda\lambda} \leq n_f$.
- $\sup_{y_t} |\partial^3 s_t/\partial f_t^3| \leq c_1 f_t^{-2} \leq c_1/\omega^2$ and thus $\bar{n}_s^{fff} \to \infty$.
- $\sup_{y_t} |\partial^3 s_t/\partial\lambda^2 \partial f_t| \leq c_1$ and thus $\bar{n}_s^{\lambda\lambda f} \to \infty$.
- $\sup_{y_t} |\partial^3 s_t/\partial\lambda \partial f_t^2| \leq c_1 f_t^{-1} \leq c_1/\omega$ and thus $\bar{n}_s^{\lambda ff} \to \infty$.
- $|\log \bar{g}_t'| \leq c_1 + c_2|f_t|^\delta$ for arbitrarily small positive δ given $f_t > \omega$, and thus $n_{\log \bar{g}'} \leq n_f/\delta$.
- $|\bar{p}_t| < c_1 + c_2 \log|1 + y_t^2/(\lambda\omega)| \leq c_3 + c_4|y_t|^\delta$ for arbitrarily small positive δ, and thus $n_{\bar{p}} \leq n_y/\delta$.
- $|\nabla_t| < \sup_{y_t} |\nabla_t| \leq c_1 f_t^{-1} + \frac{1}{2} f_t^{-1} \leq c_2/\omega$, and thus $\bar{n}_\nabla \to \infty$.

For asymptotic normality, a further sets of moments and derivatives need to be established. We have

$$\begin{aligned}
\bar{p}_t^\lambda &= \left(\frac{\partial}{\partial\lambda} \log \frac{\Gamma\big(\frac{1}{2}(\lambda+1)\big)}{\Gamma(\lambda/2)\sqrt{\pi\lambda}} \right) + \tfrac{1}{2} \frac{(1+\lambda^{-1}) y_t^2/(\lambda f_t)}{1 + y_t^2/(\lambda f_t)} - \tfrac{1}{2} \log\big(1 + y_t^2/(\lambda f_t)\big), \\
\bar{p}_t^{\lambda\lambda} &= -\tfrac{1}{2} \frac{\big(y_t^2/(\lambda f_t)\big)\lambda^{-2}\Big(2 + (1-\lambda)\big(y_t^2/(\lambda f_t)\big)\Big)}{\Big(1 + y_t^2/(\lambda f_t)\Big)^2}, \\
\bar{p}_t^{\lambda f} &= -\tfrac{1}{2} \frac{f_t^{-1}\big(y_t^2/f_t\big)\big(1 - y_t^2/f_t\big)}{\Big(\lambda + y_t^2/f_t\Big)^2}, \\
\bar{p}_t^{\lambda\lambda f} &= \frac{f_t^{-1}\big(y_t^2/f_t\big)\big(1 - y_t^2/f_t\big)}{\Big(\lambda + y_t^2/f_t\Big)^3}, \\
\bar{p}_t^{\lambda ff} &= \tfrac{1}{2} \frac{f_t^{-2}\big(y_t^2/f_t\big)\big(2\lambda - 3\lambda y_t^2/f_t - y_t^4/f_t^2\big)}{\Big(\lambda + y_t^2/f_t\Big)^3},
\end{aligned}$$

$$\partial \nabla_t / \partial f = \frac{\frac{1}{2}\lambda f_t^{-2}\Big(\lambda - 2\lambda y_t^2/f_t - y_t^4/f_t^2\Big)}{\Big(\lambda + y_t^2/f_t\Big)^2},$$
$$\partial \nabla_t / \partial \lambda = \frac{-\frac{1}{2} f_t^{-1}\big(y_t^2/f_t\big)\big(1 - y_t^2/f_t\big)}{\Big(\lambda + y_t^2/f_t\Big)^2},$$
$$\partial^2 \nabla_t / \partial f^2 = \frac{-\lambda f_t^{-3}\Big(\lambda^2 - 3\lambda y_t^4/f_t^2 - 3\lambda^2 y_t^2/f_t - y_t^6/f_t^3\Big)}{\Big(\lambda + y_t^2/f_t\Big)^3},$$
$$\partial^2 \nabla_t / \partial f \partial \lambda = \frac{\frac{1}{2} f_t^{-2}\big(y_t^2/f_t\big)\Big(2\lambda - 3\lambda y_t^2/f_t - y_t^4/f_t^2\Big)}{\Big(\lambda + y_t^2/f_t\Big)^3},$$
$$\partial^2 \nabla_t / \partial \lambda^2 = \frac{f_t^{-1}\big(y_t^2/f_t\big)\big(1 - y_t^2/f_t\big)}{\Big(\lambda + y_t^2/f_t\Big)^3}.$$

From these moments and derivatives, we obtain

- $|\bar{p}_t^{\lambda}| \leq c_1 + c_2 \log(1 + y_t/(\lambda\omega))$, such that $n_{\bar{p}}^{\lambda} \leq n_y/\delta$ for arbitrarily small positive δ.
- $|\bar{p}_t^{\lambda\lambda}| \leq c_1$, such that $n_{\bar{p}}^{\lambda\lambda} \to \infty$.
- $|\bar{p}_t^{\lambda f}| \leq c_1 f_t^{-1} \leq c_1/\omega$, such that $\bar{n}_{\bar{p}}^{\lambda f} \to \infty$.
- $|\bar{p}_t^{\lambda\lambda f}| \leq c_1 f_t^{-1} \leq c_1/\omega$, such that $\bar{n}_{\bar{p}}^{\lambda\lambda f} \to \infty$.
- $|\bar{p}_t^{\lambda f f}| \leq c_1 f_t^{-2} \leq c_1/\omega^{-2}$, such that $\bar{n}_{\bar{p}}^{\lambda f f} \to \infty$.
- $|\partial \nabla / \partial f| \leq c_1 f_t^{-2} \leq c_1/\omega^2$, such that $\bar{n}_{\nabla}^{f} \to \infty$.
- $|\partial \nabla / \partial \lambda| \leq c_1 f_t^{-1} \leq c_1/\omega$, such that $\bar{n}_{\nabla}^{\lambda} \to \infty$.
- $|\partial^2 \nabla / \partial f^2| \leq c_1 f_t^{-3} \leq c_1/\omega^3$, such that $\bar{n}_{\nabla}^{ff} \to \infty$.
- $|\partial^2 \nabla / \partial f \partial \lambda| \leq c_1 f_t^{-2} \leq c_1/\omega^2$, such that $\bar{n}_{\nabla}^{\lambda f} \to \infty$.
- $|\partial^2 \nabla / \partial \lambda^2| \leq c_1 f_t^{-1} \leq c_1/\omega$, such that $\bar{n}_{\nabla}^{\lambda\lambda} \to \infty$.

D Likelihood Derivatives of Time-Varying Parameter

D.1 Explicit expressions for the likelihood and its derivatives

We assume that $\lambda \in \mathbb{R}$. Similar derivations hold for vector valued $\lambda \in \mathbb{R}^{d_\lambda}$. The likelihood function of the score-driven model is given by

$$\begin{aligned}
\ell_T(\boldsymbol{\theta}, \hat{f}_1) &= \frac{1}{T}\sum_{t=1}^{T} \tilde{\ell}_t(\boldsymbol{\theta}, \hat{f}_1) = \frac{1}{T}\sum_{t=1}^{T} \ell(\hat{f}_t, y_t; \lambda) \\
&= \frac{1}{T}\sum_{t=1}^{T} \log p_u\Big(g^{-1}\big(\hat{f}_t, y_t\big)\ ;\ \lambda\Big) + \log \frac{\partial g^{-1}\big(f_t, y_t\big)}{\partial y} \\
&= \frac{1}{T}\sum_{t=1}^{T} \log p_u(\bar{g}_t; \lambda) + \log \frac{\partial \bar{g}_t}{\partial y} \\
&= \frac{1}{T}\sum_{t=1}^{T} \bar{p}_t + \log \bar{g}_t'.
\end{aligned} \tag{D.1}$$

Note that we have defined the score ∇_t as $\partial \ell(\hat{f}_t, y_t; \lambda)/\partial f_t = \partial(\bar{p}_t + \log \bar{g}_t')/\partial f_t$. The derivative of the likelihood is given by

$$\begin{aligned}
\ell_T'(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:1)}) &= \frac{\partial \ell_T(\boldsymbol{\theta}, \hat{f}_1)}{\partial \boldsymbol{\theta}} = \frac{1}{T}\sum_{t=1}^{T} \tilde{\ell}_t'(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:1)}) \\
&= \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}} \cdot A_t^* + \frac{\partial \bar{p}_t}{\partial \boldsymbol{\theta}} = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}} \cdot \nabla_t + \frac{\partial \bar{p}_t}{\partial \boldsymbol{\theta}},
\end{aligned} \tag{D.2}$$

with

$$A_t^* := \frac{\partial \bar{p}_t}{\partial f_t} + \frac{\partial \log \bar{g}_t'}{\partial f_t} = \nabla_t, \qquad \hat{\boldsymbol{f}}_1^{(0:1)} = \Big(\hat{f}_1\,,\, \partial \hat{f}_1/\partial \boldsymbol{\theta}\Big),$$

and

$$\frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \hat{f}_t}{\partial \omega} & \frac{\partial \hat{f}_t}{\partial \alpha} & \frac{\partial \hat{f}_t}{\partial \beta} & \frac{\partial \hat{f}_t}{\partial \lambda} \end{bmatrix}^\top, \qquad \frac{\partial \bar{p}_t}{\partial \boldsymbol{\theta}} := \begin{bmatrix} 0 & 0 & 0 & \frac{\partial \bar{p}_t}{\partial \lambda} \end{bmatrix}^\top.$$

Note that $\partial \hat{f}_t/\partial \boldsymbol{\theta} = \hat{\boldsymbol{f}}_t^{(1)}$. The second derivative of the log-likelihood function is given by

$$\begin{aligned}
&\ell_T''(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:2)}) = \frac{\partial^2 \ell_T(\boldsymbol{\theta}, \hat{f}_1)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(\frac{\partial^2 \hat{f}_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} \cdot A_t^* + \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}} \cdot \frac{\partial A_t^*}{\partial f_t} \cdot \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}} \cdot \frac{\partial A_t^*}{\partial \boldsymbol{\theta}^\top} + \frac{\partial^2 \bar{p}_t}{\partial \boldsymbol{\theta} \partial f_t} \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial^2 \bar{p}_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} \right) \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(\frac{\partial^2 \hat{f}_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} \cdot A_t^* + \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}} \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}^\top} \cdot B_t^* + \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}} (C_t^*)^\top + C_t^* \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial^2 \bar{p}_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} \right), \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(\frac{\partial^2 \hat{f}_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} \cdot \nabla_t + \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}} \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}^\top} \frac{\partial \nabla_t}{\partial f_t} + \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}} \frac{\partial \nabla_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \nabla_t}{\partial \boldsymbol{\theta}} \frac{\partial \hat{f}_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial^2 \bar{p}_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} \right),
\end{aligned} \tag{D.3}$$

where

$$
\begin{aligned}
\hat{\boldsymbol{f}}_1^{(0:2)} &= \left(\hat{f}_1\,,\, \partial\hat{f}_1/\partial\boldsymbol{\theta}\,,\, \partial^2\hat{f}_1/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\right),\\
B_t^* &= \frac{\partial^2 \bar{p}_t}{\partial f_t^2} + \frac{\partial^2 \log \bar{g}_t'}{\partial f_t^2} = \frac{\partial \nabla_t}{\partial f_t},\\
C_t^* &= \begin{bmatrix} 0 & 0 & 0 & \frac{\partial^2 \bar{p}_t}{\partial f_t \partial\lambda} \end{bmatrix}^\top = \begin{bmatrix} 0 & 0 & 0 & \frac{\partial \nabla_t}{\partial\lambda} \end{bmatrix}^\top = \frac{\partial A_t^*}{\partial\boldsymbol{\theta}},\\
\frac{\partial^2 \hat{f}_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} &= \begin{bmatrix} \frac{\partial^2\hat{f}_t}{\partial\omega^2} & \frac{\partial^2\hat{f}_t}{\partial\omega\partial\alpha} & \frac{\partial^2\hat{f}_t}{\partial\omega\partial\beta} & \frac{\partial^2\hat{f}_t}{\partial\omega\partial\lambda} \\ \frac{\partial^2\hat{f}_t}{\partial\alpha\partial\omega} & \frac{\partial^2\hat{f}_t}{\partial\alpha^2} & \frac{\partial^2\hat{f}_t}{\partial\alpha\partial\beta} & \frac{\partial^2\hat{f}_t}{\partial\alpha\partial\lambda} \\ \frac{\partial^2\hat{f}_t}{\partial\beta\partial\omega} & \frac{\partial^2\hat{f}_t}{\partial\beta\partial\alpha} & \frac{\partial^2\hat{f}_t}{\partial\beta^2} & \frac{\partial^2\hat{f}_t}{\partial\beta\partial\lambda} \\ \frac{\partial^2\hat{f}_t}{\partial\lambda\partial\omega} & \frac{\partial^2\hat{f}_t}{\partial\lambda\partial\alpha} & \frac{\partial^2\hat{f}_t}{\partial\lambda\partial\beta} & \frac{\partial^2\hat{f}_t}{\partial\lambda^2} \end{bmatrix},\\
\frac{\partial^2 \bar{p}_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\partial^2 \bar{p}_t}{\partial\lambda^2} \end{bmatrix},
\end{aligned}
$$

where $\partial^2\hat{f}_t/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top = \hat{\boldsymbol{f}}_t^{(2)}$.

D.2 Expressions for the derivative processes of f_t

We have $\boldsymbol{\theta} = (\omega, \alpha, \beta, \lambda) \in \Theta$ and write $\partial s(f_t, v_t; \lambda)/\partial\boldsymbol{\theta}_i$ as the derivative of the scaled score w.r.t. λ only, not accounting for the dependence of f_t on $\boldsymbol{\theta}$. Differentiating the transition equation of the score-driven model, we obtain

$$
\begin{aligned}
\frac{\partial f_{t+1}}{\partial\boldsymbol{\theta}_i} &= \frac{\partial\omega}{\partial\boldsymbol{\theta}_i} + \frac{\partial\alpha}{\partial\boldsymbol{\theta}_i} s_t + \alpha\frac{\partial s_t}{\partial f_t}\frac{\partial f_t}{\partial\boldsymbol{\theta}_i} + \alpha\frac{\partial s_t}{\partial\boldsymbol{\theta}_i} + \frac{\partial\beta}{\partial\boldsymbol{\theta}_i} f_t + \beta\frac{\partial f_t}{\partial\boldsymbol{\theta}_i}, \qquad \text{(D.4)}\\
&= \boldsymbol{A}_{j,t}^{(1)} + \frac{\partial f_t}{\partial\boldsymbol{\theta}_i}\boldsymbol{B}_t,
\end{aligned}
$$

with

$$
\begin{aligned}
\boldsymbol{A}_t^{(1)} &= \boldsymbol{A}_t^{(1)}(f_t,\boldsymbol{\theta}) = \left(\boldsymbol{A}_{1,t}^{(1)}(f_t,\boldsymbol{\theta}), \ldots, \boldsymbol{A}_{4,t}^{(1)}(f_t,\boldsymbol{\theta})\right)^\top = \frac{\partial\omega}{\partial\boldsymbol{\theta}} + \frac{\partial\alpha}{\partial\boldsymbol{\theta}} s_t + \alpha\frac{\partial s_t}{\partial\boldsymbol{\theta}} + \frac{\partial\beta}{\partial\boldsymbol{\theta}} f_t,\\
\boldsymbol{B}_t &= \boldsymbol{B}_t(f_t,\boldsymbol{\theta}) = \alpha\frac{\partial s_t}{\partial f_t} + \beta.
\end{aligned}
$$

For the second derivative process, we obtain a recursion

$$
\begin{aligned}
\frac{\partial^2 f_{t+1}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} &= \frac{\partial\boldsymbol{A}_t^{(1)}}{\partial\boldsymbol{\theta}^\top} + \frac{\partial\boldsymbol{A}_t^{(1)}}{\partial f_t}\frac{\partial f_t}{\partial\boldsymbol{\theta}^\top} + \frac{\partial f_t}{\partial\boldsymbol{\theta}}\frac{\partial\boldsymbol{B}_t}{\partial\boldsymbol{\theta}^\top} + \frac{\partial f_t}{\partial\boldsymbol{\theta}}\frac{\partial\boldsymbol{B}_t}{\partial f_t}\frac{\partial f_t}{\partial\boldsymbol{\theta}^\top} + \frac{\partial^2 f_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\boldsymbol{B}_t \qquad \text{(D.5)}\\
&= \boldsymbol{A}_t^{(2)} + \frac{\partial^2 f_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\boldsymbol{B}_t,
\end{aligned}
$$

with

$$
\begin{aligned}
\boldsymbol{A}_t^{(2)} &= \frac{\partial\boldsymbol{A}_t^{(1)}}{\partial\boldsymbol{\theta}^\top} + \frac{\partial\boldsymbol{A}_t^{(1)}}{\partial f_t}\frac{\partial f_t}{\partial\boldsymbol{\theta}^\top} + \frac{\partial f_t}{\partial\boldsymbol{\theta}}\frac{\partial\boldsymbol{B}_t}{\partial\boldsymbol{\theta}^\top} + \frac{\partial\boldsymbol{B}_t}{\partial f_t}\frac{\partial f_t}{\partial\boldsymbol{\theta}}\frac{\partial f_t}{\partial\boldsymbol{\theta}^\top} \qquad \text{(D.6)}\\
&= \left(\frac{\partial\alpha}{\partial\boldsymbol{\theta}}\frac{\partial s_t}{\partial\boldsymbol{\theta}^\top} + \frac{\partial s_t}{\partial\boldsymbol{\theta}}\frac{\partial\alpha}{\partial\boldsymbol{\theta}^\top} + \alpha\frac{\partial^2 s_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right) + \left(\frac{\partial\alpha}{\partial\boldsymbol{\theta}}\frac{\partial s_t}{\partial f_t} + \alpha\frac{\partial^2 s_t}{\partial\boldsymbol{\theta}\partial f_t} + \frac{\partial\beta}{\partial\boldsymbol{\theta}}\right)\frac{\partial f_t}{\partial\boldsymbol{\theta}^\top}\\
&\quad + \frac{\partial f_t}{\partial\boldsymbol{\theta}}\left(\frac{\partial s_t}{\partial f_t}\frac{\partial\alpha}{\partial\boldsymbol{\theta}^\top} + \alpha\frac{\partial^2 s_t}{\partial f_t\partial\boldsymbol{\theta}^\top} + \frac{\partial\beta}{\partial\boldsymbol{\theta}^\top}\right) + \alpha\frac{\partial^2 s_t}{\partial f_t^2}\frac{\partial f_t}{\partial\boldsymbol{\theta}}\frac{\partial f_t}{\partial\boldsymbol{\theta}^\top}.
\end{aligned}
$$

E Further Derivations for Example of Section 5

In this appendix we show that Assumptions 4.8 and 4.9 which are needed for consistency of the MLE under correct specification hold for the example discussed in Section 5 (a score-driven location model with Student's t-distributed innovations).

It is straightforward to check that condition *(i)*, *(ii)* and *(iii)* of Assumption 4.8 are satisfied in this setting. Also, by the C_r-inequality of Loève (1977, p.157) we have that $n_g = \min\{n_{f^u}, n_u\}$. Thus, $n_g > 0$, because f_t^u can be set arbitrarily high by Proposition 3.1 under the conditions of Assumption 4.9 and $0 < n_u < \inf_{\Theta_*} \nu$.

Finally, consider the conditions of Assumption 4.9. Defining the updating recursion of the time-varying parameter in terms of u_t and $\hat{f}_t^u$ as in equation (3.1), we obtain

$$\hat{f}_{t+1}^u = \omega + \alpha(1 + \lambda^{-1} e^{-2\kappa} u_t^2)^{-1} u_t + \beta \hat{f}_t^u \,,$$

using that $s_{u,t} = (1 + \lambda^{-1} e^{-2\kappa} u_t^2) u_t$. That condition *(i)* of Assumption 4.9 holds now follows immediately because $s_{u,t}$ is uniformly bounded. Also, because $\partial s_{u,t}/\partial f^u = 0$, condition *(ii)* simplifies to $|\beta| < 1$. The shaded area in Figure 1 represents (a part of) the (α, β)-pairs which meet this restriction. Clearly, the parameter restriction $|\beta| < 1$ holds for all $\boldsymbol{\theta} \in \Theta^*$. In other words, the parameter restrictions we needed for filter invertibility are sufficient for the true time-varying parameter to be SE. Assumption 4.9 also requires that for every $(f, \boldsymbol{\theta}) \in \mathcal{F}_s \times \Theta_*$, the derivative $\partial s(f, y, \lambda)/\partial y \neq 0$ for almost every $y \in \mathcal{Y}_g$. For this model, it is not hard to see that this holds for every $\nu < \infty$. Lastly, we must have $\alpha \neq 0$ for all $\boldsymbol{\theta} \in \Theta_*$. In other words, we can take some compact $\Theta_* \subseteq \{\boldsymbol{\theta} \in \mathbb{R}^5 : |\beta| < 1 \,, \alpha \neq 0, \nu > 0\}$. Notice that essentially Assumption 4.9 does not impose any further restrictions on the parameter region, except for the restriction $\alpha \neq 0$, which is required for identification of the model.

Under correct specification, it now follows from Corollary 4.11 that for any compact parameter set $\Theta \subseteq \Theta^* \cap \Theta_*$, where Θ^* and Θ_* meet the requirements described above, the MLE $\hat{\boldsymbol{\theta}}_T$ is consistent.

F Further Technical Lemmas and Proofs

This appendix contains a number of more technical results.

The following set of lemmas derives the bounds on the moments of the likelihood function based on moments of the inputs. The results follow from the properties of moment preserving maps as laid out in Technical Appendix G, but can also be proved directly.

Lemma TA.9. $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\ell_T'(\boldsymbol{\theta},\hat{f}_1)|^m<\infty$ *where*

$$m=\min\left\{n_{\bar{p}}^{\lambda}\ ,\ \frac{n_{\nabla}n_{f_{\boldsymbol{\theta}}}}{n_{\nabla}+n_{f_{\boldsymbol{\theta}}}}\right\}. \tag{F.1}$$

Proof. Using the explicit form of the first derivative of the likelihood in (D.2) in Technical Appendix D, the number of moments for the likelihood score is at least the minimum of the number of moments for each of the terms making up the score, namely

$$\frac{\partial\bar{p}}{\partial\boldsymbol{\theta}},\qquad \frac{\partial f_t}{\partial\boldsymbol{\theta}}\nabla_t.$$

The number of moments for the first term is $n_{\bar{p}}^{\lambda}$. Using a generalized Hölder inequality, the second term has moments $n_{\nabla}n_{f_{\boldsymbol{\theta}}}/(n_{\nabla}+n_{f_{\boldsymbol{\theta}}})$. This yields the expression for m in equation (F.1). ∎

Lemma TA.10. $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\ell_T''(\boldsymbol{\theta},f)|^m<\infty$ *where*

$$m\ =\ \min\left\{n_{\bar{p}}^{\lambda\lambda}\ ,\ \frac{n_{\nabla}n_{f_{\boldsymbol{\theta\theta}}}}{n_{\nabla}+n_{f_{\boldsymbol{\theta\theta}}}}\ ,\ \frac{n_{\nabla}^{\lambda}n_{f_{\boldsymbol{\theta}}}}{n_{\nabla}^{\lambda}+n_{f_{\boldsymbol{\theta}}}}\ ,\ \frac{n_{\nabla}^{f}n_{f_{\boldsymbol{\theta}}}}{2n_{\nabla}^{f}+n_{f_{\boldsymbol{\theta}}}}\right\}. \tag{F.2}$$

Proof. The statement follows by Hölder's generalized inequality and from the explicit expression for the second derivative of the likelihood in equation (D.3) in Technical Appendix D, we obtain that the number of moments m is at least that of the minimum number of moments of the following terms

$$\frac{\partial^2 f_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\nabla_t,\qquad \frac{\partial f_t}{\partial\boldsymbol{\theta}}\frac{\partial f_t}{\partial\boldsymbol{\theta}^\top}\frac{\partial\nabla_t}{\partial f_t},\qquad \frac{\partial f_t}{\partial\boldsymbol{\theta}}\frac{\partial\nabla_t}{\partial\lambda},\qquad \frac{\partial^2\bar{p}_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}.$$

Using generalized Hölder inequalities, the number of moments for each of these terms are, respectively,

$$\frac{n_{\nabla}n_{f_{\boldsymbol{\theta\theta}}}}{n_{\nabla}+n_{f_{\boldsymbol{\theta\theta}}}},\qquad \frac{n_{\nabla}^{f}n_{f_{\boldsymbol{\theta}}}}{2n_{\nabla}^{f}+n_{f_{\boldsymbol{\theta}}}},\qquad \frac{n_{\nabla}^{\lambda}n_{f_{\boldsymbol{\theta}}}}{n_{\nabla}^{\lambda}+n_{f_{\boldsymbol{\theta}}}},\qquad n_{\bar{p}}^{\lambda\lambda}.$$

This makes up the expression for m in equation (F.2). ∎

The following lemmas support the proof of Theorem 4.15.

Lemma TA.11. *Let the conditions of Theorem 4.15 hold. Then* $\ell_T'(\boldsymbol{\theta}_0)$ *is a sample average of a sequence that is SE and NED of size* -1 *on a strongly mixing sequence of size* $-\delta/(1-\delta)$ *for some* $\delta>2$.

Proof. By assumption, $\{y_t\}_{t\in\mathbb{Z}}$ satisfies $\mathbb{E}|y_t|^{n_y} < \infty$ for some $n_y \geq 0$ and is SE and NED of size -1 on a strongly mixing process of size $-\delta/(1-\delta)$ for some $\delta > 2$. Assumption 4.4 and the moment conditions of Assumption 4.12 ensure that the limit process $\{\boldsymbol{f}_t^{(0:1)}(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is both SE (Propositions 3.3 and 3.5) and NED (Pötscher and Prucha (1997, Theorem 6.10)) of size -1 on the strongly mixing process. The SE nature of the terms $\tilde{\ell}_t'\big(y_t, \boldsymbol{f}_t^{(0:1)}(\boldsymbol{\theta}_0); \lambda\big)$ that compose the score

$$\ell_T'(\boldsymbol{\theta}_0) \;=\; \frac{1}{T}\sum_{t=1}^{T}\tilde{\ell}_t'(\boldsymbol{\theta}_0) \;=\; \frac{1}{T}\sum_{t=1}^{T}\frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\cdot A_t^* + \frac{\partial \bar{p}_t}{\partial\boldsymbol{\theta}}$$

follows immediately by Krengel's theorem (Krengel (1985)) and the continuity of the score on the SE processes $\{y_t\}_{t\in\mathbb{Z}}$ and $\{\boldsymbol{f}_t^{(0:1)}(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$. Finally, the NED nature of the terms in $\ell_T'(\boldsymbol{\theta}_0)$ is ensured by noting that Assumption 4.14 ensures that A_t^* is uniformly bounded and A_t^* and $\frac{\partial \bar{p}_t}{\partial\boldsymbol{\theta}}$ are a.s. Lipschitz continuous, and hence that $\tilde{\ell}_t'$ is Lipschitz continuous on $(y_t, \boldsymbol{f}_t^{(0:1)}(\boldsymbol{\theta}_0))$, which implies by Theorem 17.12 of Davidson (1994) or Theorem 6.15 of Pötscher and Prucha (1997) that $\{\tilde{\ell}_t'(\boldsymbol{\theta}_0)\}$ is NED of size -1 on the mixing sequence. ∎

Lemma TA.12. *Under the conditions of Theorem 4.15,*

$$\sqrt{T}\|\ell_T'(\boldsymbol{\theta}_0, \hat{\boldsymbol{f}}_1^{(0:1)}) - \ell_T'(\boldsymbol{\theta}_0)\| \overset{a.s.}{\to} 0 \quad \text{as} \quad T\to\infty. \tag{F.3}$$

Proof. We establish the a.s. convergence in (F.3) by showing the e.a.s. convergence of the individual contributions of the score of the log likelihood

$$\|\tilde{\ell}_t'(\boldsymbol{\theta}_0, \hat{\boldsymbol{f}}_1^{(0:1)}) - \tilde{\ell}_t'(\boldsymbol{\theta}_0)\| \overset{e.a.s.}{\to} 0 \quad \text{as} \quad T\to\infty.$$

This e.a.s. convergence follows from $|\hat{f}_t - f_t| \overset{e.a.s.}{\to} 0$ and

$$\|\hat{\boldsymbol{f}}_t^{(1)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{f}}_1^{(0:1)}) - \boldsymbol{f}_t^{(1)}(\boldsymbol{\theta}_0)\| \overset{e.a.s.}{\to} 0,$$

as implied by Propositions 3.3 and 3.5 respectively, which hold because of Assumptions 4.4 and 4.12. Now consider the expression of $\tilde{\ell}_t'(\boldsymbol{\theta}_0, \hat{\boldsymbol{f}}_1^{(0:1)})$ given in (D.2) to rewrite the difference under investigation:

$$\begin{aligned}\|\tilde{\ell}_t'(\boldsymbol{\theta}_0, \boldsymbol{f}_1^{(0:1)}) - \tilde{\ell}_t'(\boldsymbol{\theta}_0)\| &= \left\|\hat{\boldsymbol{f}}_t^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{f}_1^{(0:1)})\cdot\hat{\nabla}_t + \frac{\partial\hat{\bar{p}}_t}{\partial\boldsymbol{\theta}} - \boldsymbol{f}_t^{(1)}(\boldsymbol{\theta}_0)\cdot\nabla_t - \frac{\partial\bar{p}_t}{\partial\boldsymbol{\theta}}\right\| \\ &\leq \left\|\hat{\boldsymbol{f}}_t^{(1)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{f}}_1^{(0:1)})\cdot\hat{\nabla}_t - \boldsymbol{f}_t^{(1)}(\boldsymbol{\theta}_0)\cdot\nabla_t\right\| + \left\|\frac{\partial\hat{\bar{p}}_t}{\partial\boldsymbol{\theta}} - \frac{\partial\bar{p}_t}{\partial\boldsymbol{\theta}}\right\|,\end{aligned} \tag{F.4}$$

where the quantities $\hat{\nabla}_t$ and $\frac{\partial\hat{\bar{p}}_t}{\partial\boldsymbol{\theta}}$ are based on $\hat{f}_t(\boldsymbol{\theta}_0, \hat{f}_1)$ and their analogues without hats are based on $f_t(\boldsymbol{\theta}_0)$. Both terms can be shown to converge to zero e.a.s. by application of Lemma 2.1 in SM06. The first term also requires the application of Lemma TA.14. We just argued that $\|\hat{\boldsymbol{f}}_t^{(0:1)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{f}}_1^{(0:1)}) - \boldsymbol{f}_t^{(0:1)}(\boldsymbol{\theta}_0)\| \overset{e.a.s.}{\to} 0$, meaning that each of the elements converge, and it follows from Proposition 3.5 that the limit sequence is SE and has a bounded moment, as $n_{f_{\boldsymbol{\theta}}} > 0$. Also, as ∇_t is continuously differentiable in f_t, it follows from the mean value theorem that

$$\left|\hat{\nabla}_t - \nabla_t\right| \leq \sup_f\left|\frac{\partial\nabla_t}{\partial f}\right|\cdot|\hat{f}_t(\boldsymbol{\theta}_0, \hat{f}_1) - f_t(\boldsymbol{\theta}_0)| \overset{e.a.s.}{\to} 0 \quad \text{as } t\to\infty, \tag{F.5}$$

where it must be noted that ∇_t only depends on f_t and not on $\boldsymbol{f}_t^{(1)}$. The convergence follows from Lemma 2.1 in SM06 since $\{\sup_f |\partial\nabla_t/\partial f|\}_{t\in\mathbb{N}}$ is SE with a logarithmic moment, because $\bar{n}_{\nabla}^{f} > 0$, and because the second factor vanishes e.a.s. Note that $\{\nabla_t\}_{t\in\mathbb{Z}}$ is SE by the continuity of ∇_t on the SE sequence $\{y_t, f_t(\cdot)\}_{t\in\mathbb{Z}}$ and Proposition 4.2 in Krengel (1985), and has a bounded moment because $\bar{n}_{\nabla} > 0$. So it follows from Lemma TA.14 that the first term of F.4 converges to zero e.a.s.

Next, we show that the second term of (F.4) vanishes e.a.s. by again invoking the mean value theorem, which can be done because $\partial\bar{p}_t/\partial\lambda$ is continuously differentiable in f:

$$\left\| \frac{\partial\hat{\bar{p}}_t}{\partial\lambda} - \frac{\partial\bar{p}_t}{\partial\lambda} \right\| \leq \sup_f \left| \frac{\partial^2\bar{p}_t}{\partial\lambda\partial f} \right| \cdot |\hat{f}_t(\boldsymbol{\theta}_0, \hat{f}_1) - f_t(\boldsymbol{\theta}_0)| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t\to\infty\,,$$

where the convergence again follows from Lemma 2.1 in SM06 because $\{\sup_f \left|\partial^2\bar{p}/\partial\lambda\partial f\right|\}_{t\in\mathbb{N}}$ is SE with a logarithmic moment, as $\bar{n}_{\bar{p}}^{\lambda f} > 0$. This finishes the proof. ∎

Lemma TA.13. *Under the conditions of Theorem 4.15,* $\sup_{\boldsymbol{\theta}\in\Theta} \|\ell_T''(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:2)}) - \ell_T''(\boldsymbol{\theta})\| \stackrel{a.s.}{\to} 0$ *as* $t\to\infty$.

Proof. The proof takes on a similar approach as the proof of Lemma TA.12. Again, instead of considering the average log likelihood, we prove that the individual contributions of the Hessian of the log likelihood vanish e.a.s.

$$\sup_{\boldsymbol{\theta}\in\Theta} \|\tilde{\ell}_t''(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:2)}) - \tilde{\ell}_t''(\boldsymbol{\theta})\|\,.$$

Recall that

$$\sup_{\boldsymbol{\theta}\in\Theta} \|\hat{\hat{\boldsymbol{f}}}_t^{(0:2)} - \boldsymbol{f}_t^{(0:2)}\| \stackrel{e.a.s.}{\to} 0,$$

by Proposition 3.3 and 3.5 under the maintained assumptions, where the limit sequences are SE and have a bounded moment. - Now consider the expression of the second derivative of the log likelihood given in (D.3) to rewrite this difference:

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\Theta} \|\tilde{\ell}_t''(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:2)}) - \tilde{\ell}_t''(\boldsymbol{\theta})\| \leq{}& \sup_{\boldsymbol{\theta}\in\Theta} \left\| \hat{\hat{\boldsymbol{f}}}_t^{(2)} \cdot \hat{\nabla}_t - \boldsymbol{f}_t^{(2)} \cdot \nabla_t \right\| \\
&+ \sup_{\boldsymbol{\theta}\in\Theta} \left\| \hat{\hat{\boldsymbol{f}}}_t^{(1)} (\hat{\hat{\boldsymbol{f}}}_t^{(1)})^\top \cdot \frac{\partial\hat{\nabla}_t}{\partial f_t} - \boldsymbol{f}_t^{(1)} (\boldsymbol{f}_t^{(1)})^\top \cdot \frac{\partial\nabla_t}{\partial f_t} \right\| \\
&+ \sup_{\boldsymbol{\theta}\in\Theta} \left\| \hat{\hat{\boldsymbol{f}}}_t^{(1)} \cdot \frac{\partial\hat{\nabla}_t}{\partial\boldsymbol{\theta}^\top} - \boldsymbol{f}_t^{(1)} \cdot \frac{\partial\nabla_t}{\partial\boldsymbol{\theta}^\top} \right\| \\
&+ \sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{\partial\hat{\nabla}_t}{\partial\boldsymbol{\theta}} \cdot (\hat{\hat{\boldsymbol{f}}}_t^{(1)})^\top - \frac{\partial\nabla_t}{\partial\boldsymbol{\theta}} \cdot (\boldsymbol{f}_t^{(1)})^\top \right\| \\
&+ \sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{\partial^2\hat{\bar{p}}_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \frac{\partial^2\bar{p}_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \right\|,
\end{aligned} \tag{F.6}$$

where all terms with hats are evaluated at elements of the initialized process $\{\hat{\hat{\boldsymbol{f}}}_t^{(0:2)}\}$ and all terms without hats are evaluated in the SE limit process $\{\boldsymbol{f}_t^{(0:2)}\}$. Every term of (F.6) vanishes e.a.s. We start by applying Corollary TA.16 to the first term, because we argued that $\sup_{\boldsymbol{\theta}\in\Theta} \|\hat{\hat{\boldsymbol{f}}}_t^{(2)} - \boldsymbol{f}_t^{(2)}\| \stackrel{e.a.s.}{\to}$

0, meaning that this convergence occurs for each element of $\boldsymbol{f}_t^{(2)}$, where the limit process is SE and has a bounded log moment, as $n_{f\boldsymbol{\theta\theta}} > 0$. Also, that

$$\sup_{\boldsymbol{\theta}\in\Theta} \left|\hat{\nabla}_t - \nabla_t\right| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t \to \infty\,,$$

can be shown using similar steps as in the proof of Lemma TA.12 in (F.5). Because the sequence $\{\sup_{\boldsymbol{\theta}\in\Theta} \sup_f |\partial\nabla_t/\partial f|\}_{t\in\mathbb{N}}$ is SE and has a log moment, because $\bar{n}_\nabla^f > 0$. It was also argued that $\{\nabla_t\}$ is SE and has a bounded log moment uniformly over Θ. Thus, Corollary TA.16 can be applied.

For the second term of (F.6), we note that

$$\sup_{\boldsymbol{\theta}\in\Theta} \left|\frac{\partial\hat{f}_t}{\partial\boldsymbol{\theta}_i}\frac{\partial\hat{f}_t}{\partial\boldsymbol{\theta}_j} - \frac{\partial f_t}{\partial\boldsymbol{\theta}_i}\frac{\partial f_t}{\partial\boldsymbol{\theta}_j}\right| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t \to \infty\,,$$

by Corollary TA.16, because we argued above that $\sup_{\boldsymbol{\theta}\in\Theta} \|\hat{\hat{\boldsymbol{f}}}_t^{(1)} - \boldsymbol{f}_t^{(1)}\| \stackrel{e.a.s.}{\to} 0$, where the limit process is SE and has some bounded moment $n_{f_{\boldsymbol{\theta}}} > 0$. The convergence to zero of the second term of of (F.6) now follows from a second application of Corollary TA.16, because by the mean value theorem we have

$$\sup_{\boldsymbol{\theta}\in\Theta} \left\|\frac{\partial\hat{\nabla}_t}{\partial f} - \frac{\partial\nabla_t}{\partial f}\right\| \leq \sup_{\boldsymbol{\theta}\in\Theta}\sup_f \left|\frac{\partial^2\nabla_t}{\partial f^2}\right| \cdot \sup_{\boldsymbol{\theta}\in\Theta} |\hat{f}_t - f_t| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t \to \infty\,,$$

where the convergence to zero follows from Lemma 2.1 in SM06 because $\{\sup_{\boldsymbol{\theta}\in\Theta} \sup_f |\partial^2\nabla_t/\partial f^2|\}_{t\in\mathbb{N}}$ is SE and has a bounded log moment because $\bar{n}_\nabla^{ff} > 0$ by assumption. Thus, the result follows from Corollary TA.16, because $\{\partial\nabla_t/\partial f\}_{t\in\mathbb{N}}$ is SE and has a bounded log moment as $\bar{n}_\nabla^f > 0$.

For the remaining three terms of (F.6) the convergence result follows by taking exactly the same steps, so we omit a detailed derivation. Note that the necessary moment conditions that correspond to the third and fourth term are $n_f > 0$, $n_{f_{\boldsymbol{\theta}}} > 0$, $\bar{n}_\nabla^{\lambda f} > 0$ and $n_\nabla^\lambda > 0$ and for the fifth term it is $\bar{n}_{\bar{p}}^{\lambda\lambda f} > 0$. By assumption all these moment conditions hold.

∎

Lemma TA.14. *Let $\{\hat{x}_t(\boldsymbol{\theta},\bar{x})\}_{t\in\mathbb{N}}$ and $\{\hat{x}_t(\boldsymbol{\theta},\bar{x})\}_{t\in\mathbb{N}}$ be sequences that converge e.a.s. to their SE limits $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ and $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, respectively, i.e.,*

$$|\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0, \qquad |\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t \to \infty.$$

Let $\mathbb{E}\log|x_t(\boldsymbol{\theta})| < \infty$ and $\mathbb{E}\log|x_t(\boldsymbol{\theta})| < \infty$. Then

$$|\hat{x}_t(\boldsymbol{\theta},\bar{x})\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})x_t(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t \to \infty.$$

Proof. We have

$$\begin{aligned}
&|\hat{x}_t(\boldsymbol{\theta},\bar{x})\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})x_t(\boldsymbol{\theta})| \\
&\qquad = |\hat{x}_t(\boldsymbol{\theta},\bar{x})\hat{x}_t(\boldsymbol{\theta},\bar{x}) - \hat{x}_t(\boldsymbol{\theta},\bar{x})x_t(\boldsymbol{\theta}) + \hat{x}_t(\boldsymbol{\theta},\bar{x})x_t(\boldsymbol{\theta}) - x_t(\boldsymbol{\theta})x_t(\boldsymbol{\theta})| \\
&\qquad \leq |\hat{x}_t(\boldsymbol{\theta},\bar{x})| \cdot |\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})| + |\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})| \cdot |x_t(\boldsymbol{\theta})| \\
&\qquad \leq |\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta}) + x_t(\boldsymbol{\theta})| \cdot |\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})| + \\
&\qquad\quad |\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})| \cdot |x_t(\boldsymbol{\theta})| \\
&\qquad \leq |\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})| \cdot |\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})| + \\
&\qquad\quad |x_t(\boldsymbol{\theta})| \cdot |\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})| + \\
&\qquad\quad |\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})| \cdot |x_t(\boldsymbol{\theta})|
\end{aligned}$$

The first term goes to zero e.a.s. due to the e.a.s. convergence of $\{\hat{x}_t(\boldsymbol{\theta},\bar{x})\}_{t\in\mathbb{N}}$ and $\{\hat{x}_t(\boldsymbol{\theta},\bar{x})\}_{t\in\mathbb{N}}$ to $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ and $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, respectively. The second and third term go to zero due to Lemma 2.1 in SM06 the e.a.s. convergence of $\{\hat{x}_t(\boldsymbol{\theta},\bar{x})\}_{t\in\mathbb{N}}$ and $\{\hat{x}_t(\boldsymbol{\theta},\bar{x})\}_{t\in\mathbb{N}}$, and the SE nature and existence of a log moment for both $x_t(\boldsymbol{\theta})$ and $x_t(\boldsymbol{\theta})$. ∎

Corollary TA.15. *Let $\{\hat{x}_t(\boldsymbol{\theta},\bar{x})\}_{t\in\mathbb{N}}$ be a sequence initialized at $\bar{x}$ that converges e.a.s. to an SE limit sequence $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, i.e.,*

$$|\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t\to\infty.$$

Let $\mathbb{E}\log|x_t(\boldsymbol{\theta})|<\infty$. Then

$$|\hat{x}_t(\boldsymbol{\theta},\bar{x})^2 - x_t(\boldsymbol{\theta})^2| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t\to\infty.$$

Proof. The result in this corollary follows immediately from Lemma TA.14. ∎

Corollary TA.16. *Let $\{\hat{x}_t(\boldsymbol{\theta},\bar{x})\}_{t\in\mathbb{N}}$ and $\{\hat{x}_t(\boldsymbol{\theta},\bar{x})\}_{t\in\mathbb{N}}$ be sequences that converges e.a.s. to their SE limits $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ and $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, respectively, uniformly over some set Θ i.e.,*

$$\sup_{\boldsymbol{\theta}\in\Theta}|\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0, \qquad \sup_{\boldsymbol{\theta}\in\Theta}|\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t\to\infty.$$

Let $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\log|x_t(\boldsymbol{\theta})|<\infty$ and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\log|x_t(\boldsymbol{\theta})|<\infty$. Then

$$\sup_{\boldsymbol{\theta}\in\Theta}|\hat{x}_t(\boldsymbol{\theta},\bar{x})\hat{x}_t(\boldsymbol{\theta},\bar{x}) - x_t(\boldsymbol{\theta})x_t(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t\to\infty.$$

Proof. This corollary can be proved using exactly the same steps as the proof of Lemma TA.14. The proof makes use of the subadditivity of the supremum and the fact that $\{\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|\}_{t\in\mathbb{N}}$ and $\{\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|\}_{t\in\mathbb{N}}$ are SE sequences. ∎

Lemma TA.17. *Let $\boldsymbol{A}_t^{(2)}(\boldsymbol{\theta},\hat{\boldsymbol{f}}_1^{(0:1)})$ be as defined in (D.6) and evaluated at the initialized series for $\hat{f}_t(\boldsymbol{\theta},\hat{f}_1)$ and $\hat{\boldsymbol{f}}_t^{(1)}(\boldsymbol{\theta},\hat{\boldsymbol{f}}_1^{(0:1)})$. Similarly, let $\boldsymbol{A}_t^{(2)}(\boldsymbol{\theta})$ denote the same quantity evaluated at the SE limits $f_t(\boldsymbol{\theta})$ and $\boldsymbol{f}_t^{(1)}(\boldsymbol{\theta})$. Then under the conditions of Proposition 3.5, we have*

$$\sup_{\boldsymbol{\theta}\in\Theta}|\boldsymbol{A}_t^{(2)}(\boldsymbol{\theta},\hat{\boldsymbol{f}}_1^{(0:1)}) - \boldsymbol{A}_t^{(2)}(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0.$$

Proof. Under the conditions of Proposition 3.5 it was already shown that $\sup_{\boldsymbol{\theta}\in\Theta}|\hat{f}_t - f_t| \stackrel{e.a.s.}{\to} 0$ and $\sup_{\boldsymbol{\theta}\in\Theta}\|\hat{\boldsymbol{f}}_t^{(1)} - \boldsymbol{f}_t^{(1)}\| \stackrel{e.a.s.}{\to} 0$. The expression for $\boldsymbol{A}_t^{(2)}$ in (D.6) has three different types of terms.
Type I: The terms

$$\frac{\partial\alpha}{\partial\boldsymbol{\theta}}\frac{\partial s_t}{\partial\boldsymbol{\theta}^\top}, \qquad \alpha\frac{\partial^2 s_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}, \qquad \frac{\partial\beta}{\partial\boldsymbol{\theta}}\frac{\partial f_t}{\partial\boldsymbol{\theta}^\top}.$$

These terms for $\sup_{\boldsymbol{\theta}\in\Theta}|\boldsymbol{A}_t^{(2)}(\boldsymbol{\theta},\hat{\boldsymbol{f}}_1^{(0:1)}) - \boldsymbol{A}_t^{(2)}(\boldsymbol{\theta})|$ converge e.a.s. to zero.
The first term follows by noting that $\partial\alpha/\partial\boldsymbol{\theta}$ is constant, and

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial s_t(\hat{f}_t)}{\partial\boldsymbol{\theta}} - \frac{\partial s_t(f_t)}{\partial\boldsymbol{\theta}}\right| \le \sup_{f^*}\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial^2 s_t(f^*)}{\partial f\partial\boldsymbol{\theta}}\right| \times \sup_{\boldsymbol{\theta}\in\Theta}|f_t - \hat{f}_t|. \tag{F.7}$$

The result now follows from Lemma 2.1 in SM06 due to the e.a.s. convergence $\hat{f}_t \overset{e.a.s.}{\to} f_t$ uniformly over Θ, the SE nature of the term involving the $\sup_{f^*}$, and the existence of a small positive moment for the $\sup_{f^*}$, which implies the existence of a log moment. The e.a.s. convergence for the second term follows by a similar argument.

The third term follows directly from the e.a.s. convergence $\hat{\boldsymbol{f}}_t^{(1)} \overset{e.a.s.}{\to} \boldsymbol{f}_t^{(1)}$ uniformly over Θ.

Type II: The terms

$$\frac{\partial \alpha}{\partial \boldsymbol{\theta}} \frac{\partial s_t}{\partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}, \qquad \alpha \frac{\partial^2 s_t}{\partial \boldsymbol{\theta} \partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}.$$

Both terms follow by a similar argument as the first set of terms, combined with Corollary TA.16. For instance for the first term, we have $\partial \alpha / \partial \boldsymbol{\theta}$ is constant, and

$$\frac{\partial s_t(\hat{f}_t)}{\partial f_t} \overset{e.a.s.}{\to} \frac{\partial s_t(f_t)}{\partial f_t} \tag{F.8}$$

uniformly over Θ, given the arguments under terms of Type I. Given the uniform e.a.s. convergence of both $\partial s_t(\hat{f})/\partial f_t$ and $\hat{\boldsymbol{f}}_t^{(1)}$, the results follow directly from Corollary TA.16.

Type III: The term

$$\alpha \frac{\partial^2 s_t}{\partial f_t^2} \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}.$$

The uniform e.a.s. convergence of each of the elements in $(\partial \hat{f}_t / \boldsymbol{\theta})(\partial \hat{f}_t / \boldsymbol{\theta}^\top)$ follows from Corollary TA.16 given a log moment for each of the elements of $\partial f_t / \boldsymbol{\theta}$, which is implied by $n_{f_{\boldsymbol{\theta}}} > 0$. Note that the latter also implies a log moment for $(\partial f_t / \boldsymbol{\theta})(\partial f_t / \boldsymbol{\theta}^\top)$. Next, we have by the mean value theorem that

$$\sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{\partial^2 s_t(\hat{f}_t)}{\partial f_t^2} - \frac{\partial^2 s_t(f_t)}{\partial f_t^2} \right| \leq \sup_{f^*} \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{\partial^3 s_t(f^*)}{\partial f_t^3} \right| \times \sup_{\boldsymbol{\theta} \in \Theta} \left| \hat{f}_t - f_t \right| \overset{e.a.s.}{\to} 0\,.$$

The existence of a log moment for $\sup_{f^*} \sup_{\boldsymbol{\theta} \in \Theta} |\partial^3 s_t / \partial f_t^3|$ is implied by $\bar{n}_s^{fff} > 0$. This again implies the uniform e.a.s. convergence of $\partial^2 s_t / \partial f_t^2$ via Lemma 2.1 in SM06. Also note that $\{\partial^2 s_t(f_t)/\partial f_t^2\}$ is SE and has a bounded log moment, because $\bar{n}_s^{ff} > 0$. So the final result follows by applying Corollary TA.16 to $\{\partial^2 s_t(f_t)/\partial f_t^2\}$ and each of the elements of $\{(\partial f_t / \boldsymbol{\theta})(\partial f_t / \boldsymbol{\theta}^\top)\}$.

■

G More Results on Moment Preserving Functions

Checking the moment conditions needed for a number of the propositions and theorems based on low-level conditions can be considerably simplified by working with the concept of moment preserving maps.

The final technical lemma presented below provides simple moment preserving properties for several common functions of random variables. For notational simplicity we let $h^{(k)}$ denote the kth order derivative of a function h. The moment properties on h or $h^{(k)}$ can now easily be derived from moment conditions on the inputs of h and the moment preserving properties through its membership of the set $\mathbb{M}_{\Theta,\Theta}(n,m)$.

Lemma TA.18. (Catalog of $\mathbb{M}^{\boldsymbol{k}}_{\Theta,\Theta}(\boldsymbol{n},m)$ Moment Preserving Maps) *For every $\boldsymbol{\theta}\in\Theta$, let $h(\cdot;\boldsymbol{\theta}):\mathcal{X}\to\mathbb{R}$ and $w(\cdot,\cdot,\boldsymbol{\theta}):\mathcal{X}\times\mathcal{V}\to\mathbb{R}$ be measurable functions.*

(a) Let $h(\cdot;\boldsymbol{\theta})$ be an affine function,

$$h(x;\boldsymbol{\theta})=\theta_0+\theta_1 x\ \forall\ (x,\boldsymbol{\theta})\in\mathcal{X}\times\Theta,\quad \boldsymbol{\theta}=(\theta_0,\theta_1)\in\Theta\subseteq\mathbb{R}^2.$$

Then, $h(\cdot;\boldsymbol{\theta})\in\mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m)$ with $n=m\ \forall\ \boldsymbol{\theta}\in\Theta$, and $h^{(k)}(\cdot;\boldsymbol{\theta})\in\mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m)$ for all $(\boldsymbol{\theta},n,m,k)\in\Theta\times\mathbb{R}^+_0\times\mathbb{R}^+_0\times\mathbb{N}$. If Θ is compact, then $h\in\mathbb{M}^k_{\Theta,\Theta}(n,m)$ with $n=m$ for $k=0$ and $h^{(k)}(\cdot;\boldsymbol{\theta})\in\mathbb{M}_{\Theta,\Theta}(n,m)\ \forall\ (n,m,k)\in\mathbb{R}^+_0\times\mathbb{R}^+_0\times\mathbb{N}$.

(b) Let $h(\cdot;\boldsymbol{\theta})$ be a polynomial function,

$$h(x;\boldsymbol{\theta})=\sum_{j=0}^{J}\theta_j x^j\ \forall\ (x,\boldsymbol{\theta})\in\mathcal{X}\times\Theta,\quad \boldsymbol{\theta}=(\theta_0,\dots,\theta_J)\in\Theta\subseteq\mathbb{R}^J,\ J\geq 1.$$

Then $h^{(k)}(\cdot;\boldsymbol{\theta})\in\mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m=n/(J-k)\ \forall\ (k,\boldsymbol{\theta})\in\mathbb{N}_0\times\Theta$. If Θ is compact, then $h^{(k)}\in\mathbb{M}_{\Theta,\Theta}(n,m)$ with $m=n/(J-k)\ \forall\ k\in\mathbb{N}_0$.

(c) Let

$$h(x;\boldsymbol{\theta})=\sum_{j=0}^{J}\theta_j x^{r_j}\ \forall\ (x,\boldsymbol{\theta})\in\mathcal{X}\times\Theta,\boldsymbol{\theta}=(\theta_0,\dots,\theta_J)\in\Theta\subseteq\mathbb{R}^J,$$

where $r_j\geq 0$. Then $h^{(k)}(\cdot;\boldsymbol{\theta})\in\mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m=n/(\max_j r_j-k)\ \forall\ (\boldsymbol{\theta},k)\in\Theta\in\mathbb{N}_0: k\leq\min_j r_j$. If Θ is compact, then $h^{(k)}\in\mathbb{M}_{\Theta,\Theta}(n,m)$ with $m=n/(\max_j r_j-k)\ \forall\ k\in\mathbb{N}_0: k\leq\min_j r_j$.

(d) Let

$$\sup_{x\in\mathcal{X}}|h(x;\boldsymbol{\theta})|\leq\bar{h}(\boldsymbol{\theta})<\infty\ \forall\ \boldsymbol{\theta}\in\Theta.$$

Then $h(\cdot;\boldsymbol{\theta})\in\mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m)\ \forall\ (n,m,\boldsymbol{\theta})\in\Theta\times\mathbb{R}^+_0\times\mathbb{R}^+_0$. If additionally, $\sup_{\boldsymbol{\theta}\in\Theta}\bar{h}(\boldsymbol{\theta})\leq\bar{\bar{h}}<\infty$, then $h\in\mathbb{M}_{\Theta,\Theta}(n,m)\ \forall\ (n,m)\in\mathbb{R}^+_0\times\mathbb{R}^+_0$.

(e) Let

$$h(\cdot;\boldsymbol{\theta})\in\mathbb{C}^k(\mathcal{X})$$

and

$$\sup_{x\in\mathcal{X}}|h^{(k)}(x;\boldsymbol{\theta})|\leq\bar{h}_k(\boldsymbol{\theta})<\infty\ \forall\ \boldsymbol{\theta}\in\Theta.$$

Then $h^{(k)}(\cdot;\boldsymbol{\theta})\in\mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m=n/k\ \forall\ \boldsymbol{\theta}\in\Theta$. If furthermore, $\sup_{\boldsymbol{\theta}\in\Theta}\bar{h}_k(\boldsymbol{\theta})\leq\bar{\bar{h}}<\infty$, then $h^{(k)}\in\mathbb{M}_{\Theta,\Theta}(n,m)$ with $m=n/k$.

(f) *Let*

$$w(x,v;\boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 v, \quad (\theta_0,\theta_1,\theta_2,x,v) \in \mathbb{R}^3 \times \mathcal{X} \times \mathcal{V}.$$

Then $w^{(k_x,k_v)}(\cdot,\cdot,\boldsymbol{\theta}) \in \mathbb{M}_{\Theta,\boldsymbol{\theta}}(\boldsymbol{n},m)\ \forall\ (k_x,k_v,\boldsymbol{\theta}) \in \mathbb{N}_0 \times \mathbb{N}_0 \times \Theta$ *with* $\boldsymbol{n} = (n_x,n_v)$ *and* $m = \min\{n_x,n_v\}$. *If furthermore* Θ *is compact, then*

$$w^{(k_x,k_v)} \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n},m)\ \forall\ (k_x,k_v) \in \mathbb{N}_0 \times \mathbb{N}_0,$$

with $m = \min\{n_x,n_v\}$;

(g) *If*

$$w(x,v,\boldsymbol{\theta}) = \theta_0 + \theta_1 xv, \quad (\theta_0,\theta_1) \in \mathbb{R}^2,$$

then $w^{(k_x,k_v)}(\cdot,\cdot,\boldsymbol{\theta}) \in \mathbb{M}_{\Theta,\boldsymbol{\theta}}(\boldsymbol{n},m)\ \forall\ (k_x,k_v,\boldsymbol{\theta}) \in \mathbb{N}_0 \times \mathbb{N}_0 \times \Theta$ *with* $\boldsymbol{n} = (n_x,n_v)$ *where* $m = n_x n_v/(n_x+n_v)$. *If furthermore,* Θ *is compact, then*

$$w^{(k_x,k_v)} \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n},m)\ \forall\ (k_x,k_v) \in \mathbb{N}_0 \times \mathbb{N}_0,$$

with $\boldsymbol{n} = (n_x,n_v)$ *where* $m = n_x n_v/(n_x+n_v)$.

Proof.
Part (a): By the C_r-inequality in (Loève, 1977, p.157), for *(a)* we have for some c that

$$\begin{aligned}\mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n &= \mathbb{E}|\theta_0 + \theta_1 x_t(\boldsymbol{\theta})|^n \\ &\leq c\mathbb{E}|\theta_0|^n + c\mathbb{E}|\theta_1 x_t(\boldsymbol{\theta})|^n \\ &\leq c|\theta_0|^n + c|\theta_1|^n \mathbb{E}|x_t(\boldsymbol{\theta})|^n,\end{aligned}$$

and hence, $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m)$ with $n = m\ \forall\ \boldsymbol{\theta} \in \Theta$ because

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty\ \forall\ \boldsymbol{\theta} \in \Theta \Rightarrow \mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n < \infty\ \forall\ \boldsymbol{\theta} \in \Theta.$$

Also, $h^{(k)}(\cdot;\boldsymbol{\theta}) \in \mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m)\ \forall\ (m,n,k,\boldsymbol{\theta}) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{N} \times \Theta$ as $h^{(1)}(x_t(\boldsymbol{\theta}),\boldsymbol{\theta}) = \theta_1$ and $h^{(i)}(x_t(\boldsymbol{\theta}),\boldsymbol{\theta}) = 0\ \forall\ i \geq 2$. Furthermore,

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n &= \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_0 + \theta_1 x_t(\boldsymbol{\theta})|^n \\ &\leq c\,\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_0|^n + c\,\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_1 x_t(\boldsymbol{\theta})|^n \\ &\leq c\sup_{\boldsymbol{\theta}\in\Theta}|\theta_0|^n + c\sup_{\boldsymbol{\theta}\in\Theta}|\theta_1|^n \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n,\end{aligned}$$

and as a result, if Θ is compact, we have $h \in \mathbb{M}_{\Theta,\Theta}(n,m)$ with $n = m$ because $\sup_{\boldsymbol{\theta}\in\Theta}|\theta_0|^n < \infty$ and $\sup_{\boldsymbol{\theta}\in\Theta}|\theta_1|^n < \infty$, and hence, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h\ (x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n < \infty$. Again, $h^{(k)} \in \mathbb{M}_{\Theta,\Theta}(n,m)\ \forall\ (m,n,k) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{N}$ follows from having $h^{(1)}(x_t(\boldsymbol{\theta}),\boldsymbol{\theta}) = \theta_1$ and $h^{(i)}(x_t(\boldsymbol{\theta}),\boldsymbol{\theta}) = 0\ \forall\ i \geq 2$.

Part (b): We have that for some c

$$\begin{aligned}\mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n &= \mathbb{E}|\sum_{j=0}^{J}\theta_j x_t^j(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^{J}\mathbb{E}|\theta_j x_t^j(\boldsymbol{\theta})|^n \\ &\leq c\sum_{j=0}^{J}|\theta_j|^n \mathbb{E}|x_t(\boldsymbol{\theta})|^{jn},\end{aligned}$$

and hence, $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m = n/J \ \forall\ \boldsymbol{\theta} \in \Theta$ because

$$\begin{gathered}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \\ \mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall\ \boldsymbol{\theta} \in \Theta \Rightarrow \\ \mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/J} \leq c\sum_{j=0}^{J}|\theta_j|^n \mathbb{E}|x_t(\boldsymbol{\theta})^j|^{n/J} \leq c \cdot J \cdot \mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall\ \boldsymbol{\theta} \in \Theta.\end{gathered}$$

Also, $h^{(k)}(\cdot;\boldsymbol{\theta}) \in \mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m) \ \forall\ (k,\boldsymbol{\theta}) \in \mathbb{N}_0 \times \Theta$ with $m = n/(J-k)$, because

$$h^{(k)}(x_t(\boldsymbol{\theta}),\boldsymbol{\theta}) = \sum_{j=k}^{J}\theta_j^* x^{j-k}$$

and hence

$$\begin{gathered}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \\ \mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall\ \boldsymbol{\theta} \in \Theta \Rightarrow \\ \mathbb{E}|h^{(k)}(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/(J-k)} \leq c\sum_{j=0}^{J}\mathbb{E}|\theta_j^* x_t(\boldsymbol{\theta})^{j-k}|^{n/(J-k)} \\ \leq c\sum_{j=0}^{J}|\theta_j^*|^{n/(J-k)}\mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall\ \boldsymbol{\theta} \in \Theta.\end{gathered}$$

Furthermore,

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n &= \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\sum_{j=0}^{J}\theta_j x_t^j(\boldsymbol{\theta})|^n \\ &\leq c\sum_{j=0}^{J}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_j x_t^j(\boldsymbol{\theta})|^n \\ &\leq c\sum_{j=0}^{J}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_j|^n \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^{jn},\end{aligned}$$

and hence, if Θ is compact, we have $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}_{\Theta,\Theta}(n,m)$ with $m = n/J$ because

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/J} < \infty.$$

and $h^{(k)}(\cdot;\boldsymbol{\theta}) \in \mathbb{M}_{\Theta,\Theta}(n,m)$ with $n = m/(J-k) \ \forall\ (\boldsymbol{\theta},k) \in \Theta \times \mathbb{N}_0$ because

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h^{(k)}(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/(J-k)} < \infty$$

by the same argument.

Part (c): For some c,

$$\begin{aligned}\mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n &= \mathbb{E}|\sum_{j=0}^{J}\theta_j x_t^{r_j}(\boldsymbol{\theta})|^n \\ &\leq c\sum_{j=0}^{J}\mathbb{E}|\theta_j x_t^{r_j}(\boldsymbol{\theta})|^n \\ &\leq c\sum_{j=0}^{J}|\theta_j|^n \mathbb{E}|x_t(\boldsymbol{\theta})|^{r_j n}.\end{aligned}$$

Hence, $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m = n/\max_j r_j \ \forall\ \boldsymbol{\theta} \in \Theta$ because

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow$$
$$\mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall\ \boldsymbol{\theta} \in \Theta \Rightarrow$$
$$\mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/\max_j r_j} \leq c\sum_{j=0}^{J}|\theta_j|^n\mathbb{E}|x_t(\boldsymbol{\theta})|^{r_j n/\max_j r_j} < \infty \ \forall\ \boldsymbol{\theta} \in \Theta.$$

Similarly, $h^{(k)}(\cdot;\boldsymbol{\theta}) \in \mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m = n/(\max_j r_j - k) \ \forall\ (\boldsymbol{\theta},k) \in \Theta \times \mathbb{N}_0 : k \leq \min_j r_j$, because we have

$$\begin{aligned}
\mathbb{E}|h^{(k)}(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n &= \mathbb{E}|\sum_{j=0}^{J}\theta_j^* x_t^{r_j-k}(\boldsymbol{\theta})|^n \\
&\leq c\sum_{j=0}^{J}\mathbb{E}|\theta_j^* x_t^{r_j-k}(\boldsymbol{\theta})|^n \\
&\leq c\sum_{j=0}^{J}|\theta_j^*|^n\mathbb{E}|x_t(\boldsymbol{\theta})|^{(r_j-k)n},
\end{aligned}$$

and hence it follows that

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow$$
$$\mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall\ \boldsymbol{\theta} \in \Theta \Rightarrow$$
$$\mathbb{E}|h^{(k)}(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/(\max_j r_j)} \leq c\sum_{j=0}^{J}|\theta_j^*|^n\mathbb{E}|x_t(\boldsymbol{\theta})|^{(r_j-k)n/(\max_j r_j-k)} < \infty.$$

Furthermore,

$$\begin{aligned}
\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n &= \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\sum_{j=0}^{J}\theta_j x_t^{r_j}(\boldsymbol{\theta})|^n \\
&\leq c \ \times\sum_{j=0}^{J}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_j x_t^{r_j}(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^{J}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_j|^n\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^{r_j n}.
\end{aligned}$$

Hence, if Θ is compact, we have $h \in \mathbb{M}_{\Theta,\Theta}(n,m)$ with $m = n/\max_j r_j$ because $\sup_{\boldsymbol{\theta}\in\Theta}|\theta_j|^n < \infty \ \forall\ j$, and hence it follows that

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/\max_j r_j} < \infty.$$

Similarly, we have $h^{(k)} \in \mathbb{M}_{\Theta,\Theta}(n,m)$ with $m = n/\max_j(r_j - k)$ because we have

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/(\max_j r_j-k)} < \infty$$

by the same argument.
Part (d): We have that

$$h(x_t(\boldsymbol{\theta});\boldsymbol{\theta}) \leq \bar{h}(\boldsymbol{\theta}) \ \forall\ \boldsymbol{\theta} \in \Theta \Rightarrow$$
$$\mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n \leq \bar{h}(\boldsymbol{\theta})^n \ \forall\ (\boldsymbol{\theta},n) \in \Theta \times \mathbb{R}_0^+,$$

and hence, $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m) \forall\ (n,m,\boldsymbol{\theta}) \in \Theta \times \mathbb{R}_0^+ \times \mathbb{R}_0^+$ because

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow$$
$$\mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^m \leq \bar{h}(\boldsymbol{\theta})^m < \infty \ \forall\ (n,m,\boldsymbol{\theta}) \in \Theta \times \mathbb{R}_0^+ \times \mathbb{R}_0^+.$$

Furthermore, if $\sup_{\boldsymbol{\theta}\in\Theta} \bar{h}(\boldsymbol{\theta}) \leq \bar{\bar{h}}$, then

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n \leq \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} \bar{h}(\boldsymbol{\theta})^n \;\forall\; n \in \mathbb{R}_0^+.$$

Hence, $h \in \mathbb{M}_{\Theta,\Theta}(n,m) \forall\; (n,m) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+$ as

$$\begin{gathered}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \\ \sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^m \leq \sup_{\boldsymbol{\theta}\in\Theta} \bar{h}(\boldsymbol{\theta})^m \leq \bar{\bar{h}}^m < \infty \;\forall\; (n,m) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+.\end{gathered}$$

Part (e): We have for some c and by an exact k^{th}-order Taylor expansion around a point $x \in \text{int}(\mathcal{X})$ that

$$\begin{aligned}\mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n &\leq \mathbb{E}|\sum_{j=0}^{k} \theta_j x_t^j(\boldsymbol{\theta})|^n \\ &\leq c \sum_{j=0}^{J} \mathbb{E}|\theta_j x_t^j(\boldsymbol{\theta})|^n \\ &\leq c \sum_{j=0}^{J} |\theta_j|^n \mathbb{E}|x_t(\boldsymbol{\theta})|^{jn},\end{aligned}$$

where $\infty > \theta_k \geq \bar{h}_k(\boldsymbol{\theta}) \geq \sup_{x\in\mathcal{X}} |h^{(k)}(x\boldsymbol{\theta})| \;\forall\; \boldsymbol{\theta} \in \Theta$, and hence, $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m = n/k \;\forall\; \boldsymbol{\theta} \in \Theta$ because

$$\begin{gathered}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \\ \mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \;\forall\; \boldsymbol{\theta} \in \Theta \Rightarrow \\ \mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/k} \leq c \textstyle\sum_{j=0}^{J} |\theta_j|^n \mathbb{E}|x_t(\boldsymbol{\theta})|^{jn/k} < \infty \;\forall\; \boldsymbol{\theta} \in \Theta.\end{gathered}$$

Furthermore,

$$\begin{aligned}\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n &= \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |\sum_{j=0}^{J} \theta_j x_t^j(\boldsymbol{\theta})|^n \\ &\leq c \sum_{j=0}^{J} \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |\theta_j \times x_t^j(\boldsymbol{\theta})|^n \\ &\leq c \sum_{j=0}^{J} \sup_{\boldsymbol{\theta}\in\Theta} |\theta_j|^n \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |x_t(\boldsymbol{\theta})|^{jn},\end{aligned}$$

and hence, if Θ is compact, we have $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}_{\Theta,\Theta}(n,m)$ with $m = n/k$ because $\sup_{\boldsymbol{\theta}\in\Theta} |\theta_j|^n < \infty \;\forall\; j$, and hence,

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/k} < \infty$$

by a similar argument.

Part (f): We have for some c that

$$\begin{aligned}\mathbb{E}|w(x_t(\boldsymbol{\theta}), v_t;\boldsymbol{\theta})|^n &= \mathbb{E}|\theta_0 + \theta_1 x_t(\boldsymbol{\theta}) + \theta_2 v_t|^n \\ &\leq |\theta_0|^n + |\theta_1|^n \mathbb{E}|x_t(\boldsymbol{\theta})|^n + |\theta_2|^n \mathbb{E}|v_t|^n.\end{aligned}$$

Hence, $w^{(k_x,k_v)} \in \mathbb{M}_{\Theta,\boldsymbol{\theta}}(\boldsymbol{n},m) \forall\ (k_x,k_v) \in \mathbb{N}_0 \times \mathbb{N}_0$ with $\boldsymbol{n} = (n_x, n_v)$ and $m = \min\{n_x, n_v\}$ because

$$\begin{gathered}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x_t(\boldsymbol{\theta})|^{n_x} < \infty\ \wedge\ \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |v_t|^{n_v} < \infty \Rightarrow \\ \mathbb{E}|x_t(\boldsymbol{\theta})|^{n_x} < \infty\ \wedge\ \mathbb{E}|v_t|^{n_v} < \infty\end{gathered}$$

implies

$$\begin{aligned}\mathbb{E}|w(x_t(\boldsymbol{\theta}), v_t; \boldsymbol{\theta})|^{\min\{n_x,n_v\}} &\leq |\theta_0|^{\min\{n_x,n_v\}} + |\theta_1|^{\min\{n_x,n_v\}}\mathbb{E}|x_t(\boldsymbol{\theta})|^{\min\{n_x,n_v\}} + \\ &\qquad |\theta_2|^{\min\{n_x,n_v\}}\mathbb{E}|v_t|^{\min\{n_x,n_v\}} < \infty\end{aligned}$$

and $\mathbb{E}|w^{(1,0)}(x_t(\boldsymbol{\theta}), v_t; \boldsymbol{\theta})|^{\min\{n_x,n_v\}} = |\theta_1|^n < \infty$. Similarly for v we have

$$\mathbb{E}|w^{(0,1)}(x_t(\boldsymbol{\theta}), v_t; \boldsymbol{\theta})|^{\min\{n_x,n_v\}} = |\theta_2|^n < \infty,$$

and for any derivative we have

$$\mathbb{E}|w^{(k_x,k_v)}(x_t(\boldsymbol{\theta}), v_t; \boldsymbol{\theta})|^{\min\{n_x,n_v\}} = 0 < \infty\ \forall\ (k_x,k_v) : k_x + k_v > 1.$$

Furthermore, if Θ is compact, then

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |w(x_t(\boldsymbol{\theta}), v_t; \boldsymbol{\theta})|^n &= \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\theta_0 + \theta_1 x_t(\boldsymbol{\theta}) + \theta_2 v_t|^n \\ &\leq \sup_{\boldsymbol{\theta}\in\Theta} |\theta_0|^n + \sup_{\boldsymbol{\theta}\in\Theta} |\theta_1|^n \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x_t(\boldsymbol{\theta})|^n + \\ &\qquad \sup_{\boldsymbol{\theta}\in\Theta} |\theta_2|^n \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |v_t|^n,\end{aligned}$$

and hence, $w^{(k_x,k_v)} \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n},m)\ \forall\ (k_x,k_v) \in \mathbb{N}_0 \times \mathbb{N}_0$ with $\boldsymbol{n} = (n_x, n_v)$ and $m = \min\{n_x, n_v\}$ because

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x_t(\boldsymbol{\theta})|^{n_x} < \infty\ \wedge\ \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |v_t|^{n_v} < \infty$$

implies by a similar argument the bound

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |w^{(k_x,k_v)}(x_t(\boldsymbol{\theta}), v_t; \boldsymbol{\theta})|^{\min\{n_1,n_2\}} < \infty.$$

Part (g): We have $\mathbb{E}|w(x_t(\boldsymbol{\theta}), v_t; \boldsymbol{\theta})|^n < \infty$ if and only if $(\mathbb{E}|w(x_t(\boldsymbol{\theta}), v_t;\ \boldsymbol{\theta})|^n)^{1/n} < \infty$. By the generalized Hölder's inequality

$$\begin{aligned}(\mathbb{E}|w(x_t(\boldsymbol{\theta}), v_t; \boldsymbol{\theta})|^n)^{1/n} &= (\mathbb{E}|\theta_0 + \theta_1 x_t(\boldsymbol{\theta}) v_t|^n)^{1/n} \\ &\leq |\theta_0| + |\theta_1|(\mathbb{E}|x_t(\boldsymbol{\theta}) v_t|^n)^{1/n} \\ &\leq |\theta_0| + |\theta_1|(\mathbb{E}|x_t(\boldsymbol{\theta})|^r)^{1/r}(\mathbb{E}|v_t|^s)^{1/s},\end{aligned}$$

with $1/r + 1/s = 1/n$, and hence, $w^{(k_x,k_v)} \in \mathbb{M}_{\Theta,\boldsymbol{\theta}}(\boldsymbol{n},m)\ \forall\ (k_x,k_v) \in \mathbb{N}_0 \times \mathbb{N}_0$ with $\boldsymbol{n} = (n_x, n_v)$ if

$$1/m = 1/n_x + 1/n_v \Leftrightarrow m = n_x n_v/(n_x + n_v),$$

because then

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x_t(\boldsymbol{\theta})|^{n_x} < \infty\ \wedge\ \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |v_t|^{n_v} < \infty,$$

which implies

$$\mathbb{E}|x_t(\boldsymbol{\theta})|^{n_x} < \infty\ \wedge\ \mathbb{E}|v_t|^{n_v} < \infty \Rightarrow \mathbb{E}|w(x_t(\boldsymbol{\theta}), v_t; \boldsymbol{\theta})|^{\frac{n_x n_v}{n_x + n_v}} < \infty.$$

Furthermore, if Θ is compact, then

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|w(x_t(\boldsymbol{\theta}),v_t;\boldsymbol{\theta})|^n<\infty$$

if and only if

$$(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|w(x_t(\boldsymbol{\theta}),v_t;\boldsymbol{\theta})|^n)^{1/n}<\infty,$$

and since we have

$$\begin{aligned}(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|w(x_t(\boldsymbol{\theta}),v_t;\boldsymbol{\theta})|^n)^{1/n} &= (\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_0+\theta_1 x_t(\boldsymbol{\theta})v_t|^n)^{1/n}\\ &\leq \sup_{\boldsymbol{\theta}\in\Theta}|\theta_0|+\sup_{\boldsymbol{\theta}\in\Theta}|\theta_1|(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})v_t|^n)^{1/n}\\ &\leq \sup_{\boldsymbol{\theta}\in\Theta}|\theta_0|+\sup_{\boldsymbol{\theta}\in\Theta}|\theta_1|(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^r)^{1/r}(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|v_t|^s)^{1/s},\end{aligned}$$

with r and s satisfying $1/r+1/s=1/n$ by the generalized Hölder's inequality, and hence, $w^{(k_x,k_v)}\in\mathbb{M}_{\Theta,\Theta}(\boldsymbol{n},m)\ \forall\ (k_x,k_v)\in\mathbb{N}_0\times\mathbb{N}_0$ with $\boldsymbol{n}=(n_x,n_v)$ if $m=n_x n_v/(n_x+n_v)$ by a similar argument. ■