EXACT MAXIMUM LIKELIHOOD ESTIMATION FOR NON-GAUSSIAN MOVING AVERAGES

Nan-Jung Hsu and F. Jay Breidt

National Tsing-Hua University and Colorado State University

Supplementary Material

This note contains proofs for Theorem 1 and some required Lemmas.

In Lii and Rosenblatt (1992), the asymptotic distribution of the MLE was derived based on the following "quasi" likelihood function

$$\ell_n(\boldsymbol{\psi}) = -n \log |\theta_s^*| + \sum_{t=-q+1}^{n} \log f_\sigma(z_t(\boldsymbol{\theta})), \tag{S1}$$

where $z_t(\boldsymbol{\theta})$ is a function of $\{X_t : t = 0, \pm 1, \pm 2, \ldots,\}$ satisfying

$$z_t(\boldsymbol{\theta}) = \left[\theta^\dagger(B)\right]^{-1} \left[\theta_s^* B^s \tilde{\theta}(B^{-1})\right]^{-1} X_t \equiv \pi_\theta(B) X_t = \sum_{j=-\infty}^{\infty} \pi_j(\boldsymbol{\theta}) X_{t-j}, \tag{S2}$$

in which $|\pi_j(\boldsymbol{\theta})|$ decays to zero exponentially as $j \to \infty$ for each $\boldsymbol{\theta}$ (unit root cases are excluded). This "quasi" likelihood is ideal but cannot be computed since the residuals $\{z_t(\boldsymbol{\theta})\}$ depend on the infinite series of $\{X_t\}$ but only finite numbers of them are observed in practice. However, the "quasi" likelihood can be viewed as the exact likelihood asymptotically. The corresponding maximizer, called the "quasi-MLE" by Lii and Rosenblatt (1992), is asymptotically equivalent to the exact MLE. Lii and Rosenblatt (1992) further approximated this "quasi" likelihood by truncating $z_t(\boldsymbol{\theta})$ in (S2) such that it only depends on the observed data, so that the corresponding truncated likelihood can be implemented in practice. They showed that the maximizer of the truncated likelihood, which we refer to as the LR MLE, is asymptotically equivalent to the ideal "quasi-MLE" and therefore is asymptotically equivalent to the MLE.

In this paper, we consider the following objective function:

$$\hat{\ell}_n(\boldsymbol{\psi}) = -n \log |\theta_s^*| + \sum_{t=-q+1}^{n} \log f_\sigma(\hat{z}_t(\boldsymbol{\theta})), \tag{S3}$$

where $\hat{z}_t(\boldsymbol{\theta})$ depends on some plug-in initial variables $(\hat{\boldsymbol{z}}_r, \hat{\boldsymbol{w}}_s)$. When $(\hat{\boldsymbol{z}}_r, \hat{\boldsymbol{w}}_s)' = \mathbf{0}$, the corresponding maximizer is the conditional MLE defined in (12). The joint MLE defined in (13) is another special case which maximizes $\hat{\ell}_n(\boldsymbol{\psi})$ with some particular $(\hat{\boldsymbol{z}}_r, \hat{\boldsymbol{w}}_s)$ as functions of $\boldsymbol{X}_n$. In the following, we shall show that the maximizer of (S3) is asymptotically equivalent to the "quasi-MLE" when the initial variables $(\hat{\boldsymbol{z}}_r, \hat{\boldsymbol{w}}_s)$ are bounded away from infinity.

Similar to Lii and Rosenblatt (1992), we first define a neighborhood of the true parameter vector $\boldsymbol{\psi}_0 \equiv (\boldsymbol{\theta}_0, \sigma_0)'$:

$$Q_\epsilon = \left\{ \boldsymbol{\psi} \in I\!R^{q+1} : |\boldsymbol{\psi} - \boldsymbol{\psi}_0| \le \epsilon \right\},$$

where $|\cdot|$ is the max norm on $I\!R^{q+1}$ (i.e., $|\boldsymbol{x}| = \max_j\{|x_j|\}$ for $\boldsymbol{x} = (x_1, x_2, \ldots, x_{q+1})'$). Then, for small $\epsilon > 0$, there exists $d \in (0,1)$ such that

$$\max_{\boldsymbol{\psi} \in Q_\epsilon} \left\{ \text{all roots in } \theta^*(z) \text{ and all roots in } \theta^\dagger(z^{-1}) \text{ in absolute value} \right\} < d < 1, \tag{S4}$$

which is a direct result from (3.2) in Lii and Rosenblatt (1992). In order to prove Theorem 1, we need the following lemmas.

Lemma S1 (Exercise 8.8 in Chapter 1, Durrett (2005)) *Assume* $\{X_t; t = 1, 2, \ldots\}$ *are iid non-zero random variables, then* $\sum_{t=1}^n X_t z^t$ *converges almost surely for* $|z| < 1$ *if and only if* $E \log^+ |X_t| < \infty$*, where* $\log^+(x) = \max\{\log(x), 0\}$.

Lemma S2 (Proposition 3.1.1 in Brockwell and Davis (1991)) *Assume* $\{X_t\}$ *is any sequence of random variables such that* $\sup_t E|X_t| < \infty$*. If* $\sum_j |\psi_j| < \infty$*, then* $\sum_j \psi_j X_{t-j}$ *converges almost surely.*

Lemma S3 *There exists some constant* $C > 0$ *such that*

$$\sup_{\boldsymbol{\psi} \in Q_\epsilon} |z_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta}_0)| \le C\,\epsilon^{1/2} \sum_{j=-\infty}^{\infty} d^{|j|} |X_{t-j}|.$$

Proof: According to Equation (3.3)' in Lii and Rosenblatt (1992), there exists some constant $C > 0$ such that

$$\sup_{\boldsymbol{\psi} \in Q_\epsilon} |\pi_j(\boldsymbol{\theta}) - \pi_j(\boldsymbol{\theta}_0)| \le C \epsilon^{1/2} d^{|j|},$$

for all j. The lemma follows directly from the expression for $z_t(\boldsymbol{\theta})$ in (S2). □

Lemma S4 *Given the observed data* $\boldsymbol{X}_n = \boldsymbol{x}_n$*, the actual values of the initial variables* $(\boldsymbol{z}_r, \boldsymbol{w}_s)$ *associated with* $\boldsymbol{x}_n$ *and some plug-in values* $(\hat{\boldsymbol{z}}_r, \hat{\boldsymbol{w}}_s)$*, we have*

$$\begin{aligned}
\sup_{\boldsymbol{\psi} \in Q_\epsilon} |\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})| &\le C_1 (n-t)^{s-1} d^{n-t} |\hat{\boldsymbol{w}}_s - \boldsymbol{w}_s| + C_2 (t+q)^{r-1} d^t |\hat{\boldsymbol{z}}_r - \boldsymbol{z}_r|, \\
\sup_{\boldsymbol{\psi} \in Q_\epsilon} \left| \frac{\partial}{\partial \boldsymbol{\theta}} (\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})) \right| &\le C_3 (n-t)^{s} d^{n-t} |\hat{\boldsymbol{w}}_s - \boldsymbol{w}_s| + C_4 (t+q)^{r} d^t |\hat{\boldsymbol{z}}_r - \boldsymbol{z}_r|, \\
\sup_{\boldsymbol{\psi} \in Q_\epsilon} \left| \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} (\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})) \right| &\le C_5 (n-t)^{s+1} d^{n-t} |\hat{\boldsymbol{w}}_s - \boldsymbol{w}_s| + C_6 (t+q)^{r+1} d^t |\hat{\boldsymbol{z}}_r - \boldsymbol{z}_r|,
\end{aligned}$$

for $t = -q+1, 2, \ldots, n$, *where* $|\cdot|$ *is the max norm on the vector space (i.e.,* $|\boldsymbol{x}| = \max_j |x_j|$*), and* C_1*–*C_6 *are some positive constants.*

Sketch of Proof: For a general MA(q) process, though $z_t(\boldsymbol{\theta}, \boldsymbol{X}_n, \boldsymbol{Z}_r, \boldsymbol{W}_s)$ can be computed recursively based on the backward and forward recursions (6) and (4), their expressions in terms of the initials $(\boldsymbol{w}_s, \boldsymbol{z}_r)$ and $\boldsymbol{\theta}$ are quite messy. For simplicity, we only sketch the proof for this lemma by showing the results for a special case: a non-purely non-invertible MA(2) (i.e., $r = s = 1$) satisfying $\theta^\dagger(B) = 1 + \theta_1^\dagger B$ and $\theta^*(B) = 1 + \theta_1^* B$ with $|\theta_1^\dagger| < 1$ and $|\theta_1^*| > 1$. The parameter vector considered in this particular case is $\boldsymbol{\psi} = (\theta_1^\dagger, \theta_1^*, \sigma)'$. Define $W_t = \theta^\dagger(B) Z_t$, we have $X_t = \theta^*(B) W_t$. According to the forward and backward recursions (4) and (6), the residuals defined in (7) can be expressed explicitly:

$$
\begin{aligned}
W_{n-j} &= -\sum_{\ell=1}^{j} \left(-\frac{1}{\theta_1^*}\right)^{\ell} X_{n-j+\ell} + \left(-\frac{1}{\theta_1^*}\right)^{j} W_n, \quad j = 1, 2, \ldots, n, \\
z_t &= \sum_{i=0}^{t} \left(-\theta_1^\dagger\right)^{i} W_{t-i} + \left(-\theta_1^\dagger\right)^{t+1} Z_{-1} \\
&= -\sum_{i=0}^{t} \sum_{\ell=1}^{n-t+i} \left(-\theta_1^\dagger\right)^{i} \left(-\frac{1}{\theta_1^*}\right)^{\ell} X_{t-i+\ell} + \sum_{i=0}^{t} \left(-\theta_1^\dagger\right)^{i} \left(-\frac{1}{\theta_1^*}\right)^{n-t+i} W_n + \left(-\theta_1^\dagger\right)^{t+1} Z_{-1} \\
&= -\sum_{i=0}^{t} \sum_{\ell=1}^{n-t+i} \left(-\theta_1^\dagger\right)^{i} \left(-\frac{1}{\theta_1^*}\right)^{\ell} X_{t-i+\ell} \\
&\quad + \left(-\frac{1}{\theta_1^*}\right)^{n-t} \frac{1 - (\theta_1^\dagger/\theta_1^*)^{t+1}}{1 - \theta_1^\dagger/\theta_1^*} W_n + \left(-\theta_1^\dagger\right)^{t+1} Z_{-1},
\end{aligned} \tag{S5}
$$

for $t = 0, 1, \ldots, n$. Note that the residuals $\{z_t\}$ in (S5) are functions of $\boldsymbol{\theta}$, $\boldsymbol{X}_n$ and (Z_{-1}, W_n). We use the notation $z_t(\boldsymbol{\theta})$ for $z_t(\boldsymbol{\theta}, \boldsymbol{X}_n, z_{-1}, w_n)$ where (z_{-1}, w_n) are the actual values of initial conditions to generate the data $\boldsymbol{X}_n = \boldsymbol{x}_n$ and the notation $\hat{z}_t(\boldsymbol{\theta})$ for $z_t(\boldsymbol{\theta}, \boldsymbol{X}_n, \hat{z}_{-1}, \hat{w}_n)$ where $(\hat{z}_{-1}, \hat{w}_n)$ are some plug-in values of initial conditions.

According to (S5), we have

$$
\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta}) = \frac{1 - (\theta_1^\dagger/\theta_1^*)^{t+1}}{1 - \theta_1^\dagger/\theta_1^*} \left(-\frac{1}{\theta_1^*}\right)^{n-t} (\hat{w}_n - w_n) + \left(-\theta_1^\dagger\right)^{t+1} (\hat{z}_{-1} - z_{-1}).
$$

For this particular MA(2) process, the inequality in (S4) becomes

$$
\max_{\boldsymbol{\psi} \in Q_\epsilon} \left\{ |\theta_1^\dagger|, \, |\theta_1^*|^{-1} \right\} < d < 1. \tag{S6}
$$

Therefore,

$$
\sup_{\boldsymbol{\psi} \in Q_\epsilon} |\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})| \le C_1 d^{n-t} |\hat{w}_n - w_n| + C_2 d^t |\hat{z}_{-1} - z_{-1}|.
$$

For this particular example, the other two inequalities can be easily obtained by taking the derivatives of $\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and applying (S6) to them. For a general MA(q), the linear expression for $z_t(\boldsymbol{\theta}, \boldsymbol{X}_n, \boldsymbol{Z}_r, \boldsymbol{W}_s)$ can also be derived explicitly but the coefficients are very complicated in terms of $\boldsymbol{\theta}$. However, the coefficients associated with $\boldsymbol{W}_s$ and $\boldsymbol{Z}_r$ have the orders $O((n-t)^{s-1}d^{n-t})$ and $O((t+q)^{r-1}d^t)$ for $\boldsymbol{\psi} \in Q_\epsilon$. Consequently, the coefficients of $\boldsymbol{W}_s$ and $\boldsymbol{Z}_r$ have the orders $O((n-t)^s d^{n-t})$ and $O((t+q)^r d^t)$ in the first derivative $\partial z_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and have the orders $O((n-t)^{s+1}d^{n-t})$ and $O((t+q)^{r+1}d^t)$ in the second derivative $\partial^2 z_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$, respectively.
□

Lemma S5 *All of the elements in the derivatives $\partial z_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\partial^2 z_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ have the form $\sum_{j=-\infty}^{\infty} \gamma_j Z_{t-j}$ with some $\{\gamma_j\}$ where $|\gamma_j|$ decays exponentially as $|j| \to \infty$.*

Proof: Define $V_t = \theta^*(B) Z_t$ and

$$[\theta^*(B)]^{-1} = \left[\theta_s^* B^s \left(1 + \frac{\theta_{s-1}^*}{\theta_s^*} B^{-1} + \cdots + \frac{1}{\theta_s^*} B^{-s}\right)\right]^{-1} = (\theta_s^*)^{-1} B^{-s} \left[\tilde{\theta}(B^{-1})\right]^{-1}.$$

Under the expression of $z_t(\boldsymbol{\theta})$ in (S2), the elements of partial derivatives satisfy

$$\begin{aligned}
\frac{\partial z_t(\boldsymbol{\theta})}{\partial \theta_j^\dagger} &= \frac{\partial}{\partial \theta_j^\dagger}[\theta^\dagger(B)]^{-1} W_t = \frac{-B^j}{[\theta^\dagger(B)]^2} W_t = \frac{-1}{\theta^\dagger(B)} (\theta^\dagger(B))^{-1} W_{t-j} = -\left[\theta^\dagger(B)\right]^{-1} Z_{t-j}, \\
\frac{\partial z_t(\boldsymbol{\theta})}{\partial \theta_j^*} &= \frac{\partial}{\partial \theta_j^*}[\theta^*(B)]^{-1} V_t = \frac{-B^j}{[\theta^*(B)]^2} V_t = -\left[\theta^*(B)\right]^{-1} Z_{t-j}, \\
\frac{\partial^2 z_t(\boldsymbol{\theta})}{\partial \theta_j^\dagger \partial \theta_k^\dagger} &= \frac{\partial}{\partial \theta_k^\dagger}\left(\frac{-1}{\theta^\dagger(B)}\right) Z_{t-j} = \left[\theta^\dagger(B)\right]^{-2} Z_{t-j-k}, \\
\frac{\partial^2 z_t(\boldsymbol{\theta})}{\partial \theta_j^* \partial \theta_k^*} &= \frac{\partial}{\partial \theta_k^*}\left(\frac{-1}{\theta^*(B)}\right) Z_{t-j} = \left[\theta^*(B)\right]^{-2} Z_{t-j-k}, \\
\frac{\partial^2 z_t(\boldsymbol{\theta})}{\partial \theta_j^\dagger \partial \theta_k^*} &= \frac{\partial}{\partial \theta_k^*}\left(\frac{-1}{\theta^\dagger(B)}\right) Z_{t-j} = \left[\theta^\dagger(B)\theta^*(B)\right]^{-1} Z_{t-j-k},
\end{aligned}$$

for all possible j and k. Since both $[\theta^\dagger(B)]^{-1}$ and $[\theta^*(B)]^{-1}$ can be expressed as polynomials of B with infinite order and the coefficients of B^k decay exponentially to zero as $|k| \to \infty$, therefore all the partial derivatives have the form $\sum_{j=-\infty}^{\infty} \gamma_j Z_{t-j}$ with some $\{\gamma_j\}$ where $|\gamma_j|$ decay exponentially as $|j| \to \infty$. □

Lemma S6 *Under the assumptions (A1)–(A9) given in Theorem 1,*

$$\sup_{\boldsymbol{\psi} \in Q_\epsilon} \frac{1}{n} |\ell_n(\boldsymbol{\psi}) - \hat{\ell}_n(\boldsymbol{\psi})| \to 0, \quad a.s.$$

Proof: Given $\ell_n(\boldsymbol{\psi})$ and $\hat{\ell}_n(\boldsymbol{\psi})$ in (S1) and (S3), we have

$$\frac{1}{n}|\ell_n(\boldsymbol{\psi}) - \hat{\ell}_n(\boldsymbol{\psi})| = \frac{1}{n}\left|\sum_{t=-q+1}^{n} \log f_\sigma(z_t(\boldsymbol{\theta})) - \sum_{t=-q+1}^{n} \log f_\sigma(\hat{z}_t(\boldsymbol{\theta}))\right|$$

$$
\begin{aligned}
&\leq \frac{1}{n}\sum_{t=-q+1}^{n} |\log f_\sigma(z_t(\boldsymbol{\theta})) - \log f_\sigma(\hat{z}_t(\boldsymbol{\theta}))| \\
&= \frac{1}{n}\sum_{t=-q+1}^{n} \left|(z_t(\boldsymbol{\theta}) - \hat{z}_t(\boldsymbol{\theta}))\frac{f'_\sigma}{f_\sigma}(\hat{z}^*_t(\boldsymbol{\theta}))\right| \\
&\leq \frac{1}{n}\sum_{t=-q+1}^{n} |z_t(\boldsymbol{\theta}) - \hat{z}_t(\boldsymbol{\theta})| \left|\frac{f'_\sigma}{f_\sigma}(z_t(\boldsymbol{\theta}_0)) + \left(\frac{f'_\sigma}{f_\sigma}(\hat{z}^*_t(\boldsymbol{\theta})) - \frac{f'_\sigma}{f_\sigma}(z_t(\boldsymbol{\theta}_0))\right)\right| \\
&\leq \frac{1}{n}\sum_{t=-q+1}^{n} |z_t(\boldsymbol{\theta}) - \hat{z}_t(\boldsymbol{\theta})| \left|\frac{f'_\sigma}{f_\sigma}(z_t(\boldsymbol{\theta}_0))\right| + \frac{1}{n}\sum_{t=-q+1}^{n} |z_t(\boldsymbol{\theta}) - \hat{z}_t(\boldsymbol{\theta})| \left|\frac{f'_\sigma}{f_\sigma}(\hat{z}^*_t(\boldsymbol{\theta})) - \frac{f'_\sigma}{f_\sigma}(z_t(\boldsymbol{\theta}_0))\right| \\
&\equiv \frac{1}{n}(A_1 + A_2),
\end{aligned}
$$

where the first-order Taylor expansion is used with $\hat{z}^*_t(\boldsymbol{\theta}) = z_t(\boldsymbol{\theta}) + u_t(\boldsymbol{\theta})(\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta}))$ for some $u_t(\boldsymbol{\theta}) \in [0, 1]$ (i.e., $\hat{z}^*_t(\boldsymbol{\theta})$ is between $z_t(\boldsymbol{\theta})$ and $\hat{z}_t(\boldsymbol{\theta})$). According to Lemma S4, we have

$$
\begin{aligned}
\sup_{\boldsymbol{\psi}\in Q_\epsilon} A_1 &= \sum_{t=-q+1}^{n} \left|\frac{f'_\sigma}{f_\sigma}(z_t(\boldsymbol{\theta}_0))\right| \sup_{\boldsymbol{\psi}\in Q_\epsilon} |z_t(\boldsymbol{\theta}) - \hat{z}_t(\boldsymbol{\theta})| \\
&\leq \sum_{t=-q+1}^{n} \left|\frac{f'_\sigma}{f_\sigma}(z_t(\boldsymbol{\theta}_0))\right| \left\{C_1(n-t)^{s-1}d^{n-t}|\hat{\boldsymbol{w}}_s - \boldsymbol{w}_s| + C_2(t+q)^{r-1}d^t|\hat{\boldsymbol{z}}_r - \boldsymbol{z}_r|\right\} \\
&= C_1|\hat{\boldsymbol{w}}_s - \boldsymbol{w}_s| \sum_{t=-q+1}^{n} (n-t)^{s-1}d^{n-t} \left|\frac{f'_\sigma}{f_\sigma}(z_t(\boldsymbol{\theta}_0))\right| \\
&\quad + C_2|\hat{\boldsymbol{z}}_r - \boldsymbol{z}_r| \sum_{t=-q+1}^{n} (t+q)^{r-1}d^t \left|\frac{f'_\sigma}{f_\sigma}(z_t(\boldsymbol{\theta}_0))\right|. \qquad \text{(S7)}
\end{aligned}
$$

Since $f_\sigma(z) = \sigma^{-1}f(z/\sigma)$ and Assumption (A8), we have

$$
E\left[\frac{f'_\sigma}{f_\sigma}(z_t(\boldsymbol{\theta}_0))\right]^2 = \int \left[\frac{f'_\sigma}{f_\sigma}(z)\right]^2 f_\sigma(z)dz = \int \left[\frac{1}{\sigma}\frac{f'}{f}\left(\frac{z}{\sigma}\right)\right]^2 \frac{1}{\sigma}f\left(\frac{z}{\sigma}\right)dz = \frac{1}{\sigma^2}\int \frac{[f'(z)]^2}{f(z)}dz < \infty,
$$

and therefore $\left\{\frac{f'_\sigma}{f_\sigma}(z_t(\boldsymbol{\theta}_0))\,; t = -1, \ldots, n\right\}$ are iid random variables with finite second moment. Moreover,

$$
\sum_{t=-q+1}^{n} (n-t)^{s-1}d^{n-t} < \infty \quad \text{and} \quad \sum_{t=-q+1}^{n} (t+q)^{r-1}d^t < \infty,
$$

so that both summations of the coefficients in (S7) converge almost surely and consequently A_1 converges almost surely by Lemma S2.

For simplifying A_2, we first write

$$
\hat{z}^*_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta}_0) = (z_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta}_0)) + u_t(\boldsymbol{\theta})\left(\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})\right),
$$

which implies

$$
|\hat{z}^*_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta}_0)| \leq |z_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta}_0)| + |\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})|.
$$

Under Assumption (A9) with $u(\cdot) = \frac{f'}{f}(\cdot)$, A_2 satisfies

$$
\begin{aligned}
A_2 &= \sum_{t=-q+1}^{n} |z_t(\boldsymbol{\theta}) - \hat{z}_t(\boldsymbol{\theta})| \frac{1}{\sigma} \left| \frac{f'}{f}\left(\frac{\hat{z}_t^*(\boldsymbol{\theta})}{\sigma}\right) - \frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta}_0)}{\sigma}\right) \right| \\
&\leq \frac{1}{\sigma} \sum_{t=-q+1}^{n} |z_t(\boldsymbol{\theta}) - \hat{z}_t(\boldsymbol{\theta})| \, A \left\{ \left(1 + |z_t(\boldsymbol{\theta}_0)/\sigma|^k\right) \sigma^{-1} |\hat{z}_t^*(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta}_0)| + \sigma^{-\ell} |\hat{z}_t^*(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta}_0)|^{\ell} \right\} \\
&\leq \frac{A}{\sigma} \sum_{t=-q+1}^{n} |z_t(\boldsymbol{\theta}) - \hat{z}_t(\boldsymbol{\theta})| \left\{ \left(1 + |z_t(\boldsymbol{\theta}_0)/\sigma|^k\right) \sigma^{-1} \left(|z_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta}_0)| + |\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})|\right) \right. \\
&\qquad \left. + \, \sigma^{-\ell} \left(|z_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta}_0)| + |\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})|\right)^{\ell} \right\}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\sup_{\boldsymbol{\psi} \in Q_\epsilon} A_2 &\leq \frac{A}{(\sigma_0 - \epsilon)^2} \sum_{t=-q+1}^{n} \left(1 + \frac{|z_t(\boldsymbol{\theta}_0)|^k}{(\sigma_0 - \epsilon)^k}\right) \sup_{\boldsymbol{\psi} \in Q_\epsilon} |z_t(\boldsymbol{\theta}) - \hat{z}_t(\boldsymbol{\theta})| \sup_{\boldsymbol{\psi} \in Q_\epsilon} |z_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta}_0)| \\
&\quad + \frac{A}{(\sigma_0 - \epsilon)^2} \sum_{t=-q+1}^{n} \left(1 + \frac{|z_t(\boldsymbol{\theta}_0)|^k}{(\sigma_0 - \epsilon)^k}\right) \sup_{\psi \in Q_\epsilon} |z_t(\boldsymbol{\theta}) - \hat{z}_t(\boldsymbol{\theta})|^2 \\
&\quad + \frac{A}{(\sigma_0 - \epsilon)^{\ell+1}} \sum_{t=-q+1}^{n} \sup_{\boldsymbol{\psi} \in Q_\epsilon} |z_t(\boldsymbol{\theta}) - \hat{z}_t(\boldsymbol{\theta})| \left(2^{\ell} \sup_{\boldsymbol{\psi} \in Q_\epsilon} |z_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta}_0)|^{\ell} + 2^{\ell} \sup_{\boldsymbol{\psi} \in Q_\epsilon} |\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})|^{\ell}\right) \\
&\equiv A\left(A_{21} + A_{22} + A_{23}\right),
\end{aligned}
$$

where A_{21}, A_{22} and A_{23} can be further bounded as follows. According to Lemmas S3 and S4,

$$
\begin{aligned}
A_{21} &\leq \frac{C\,\epsilon^{1/2}}{(\sigma_0 - \epsilon)^2} \sum_{t=-q+1}^{n} \left\{ \left(1 + \frac{|z_t(\boldsymbol{\theta}_0)|^k}{(\sigma_0 - \epsilon)^k}\right) \left(\sum_{j=-\infty}^{\infty} d^{|j|} |X_{t-j}|\right) \right. \\
&\qquad \left. \times \left\{C_1 (n-t)^{s-1} d^{n-t} |\hat{\boldsymbol{w}}_s - \boldsymbol{w}_s| + C_2 (t+q)^{r-1} d^t |\hat{\boldsymbol{z}}_r - \boldsymbol{z}_r|\right\} \right\} \\
&= \epsilon^{1/2} C_3 |\hat{\boldsymbol{w}}_s - \boldsymbol{w}_s| \sum_{t=-q+1}^{n} (n-t)^{s-1} d^{n-t} \left(1 + \frac{|z_t(\boldsymbol{\theta}_0)|^k}{(\sigma_0 - \epsilon)^k}\right) \left(\sum_{j=-\infty}^{\infty} d^{|j|} |X_{t-j}|\right) \\
&\quad + \epsilon^{1/2} C_4 |\hat{\boldsymbol{z}}_r - \boldsymbol{z}_r| \sum_{t=-q+1}^{n} (t+q)^{r-1} d^t \left(1 + \frac{|z_t(\boldsymbol{\theta}_0)|^k}{(\sigma_0 - \epsilon)^k}\right) \left(\sum_{j=-\infty}^{\infty} d^{|j|} |X_{t-j}|\right),
\end{aligned}
$$

where C_i's are some positive constants. In the last equation, the second summation in A_{21} satisfies

$$
\begin{aligned}
&\sum_{t=-q+1}^{n} (t+q)^{r-1} d^t \left(1 + \frac{|z_t(\boldsymbol{\theta}_0)|^k}{(\sigma_0 - \epsilon)^k}\right) \left(\sum_{j=-\infty}^{\infty} d^{|j|} |X_{t-j}|\right) \\
&\leq \left\{ \sum_{t=-q+1}^{n} (t+q)^{r-1} d^t \left(1 + \frac{|z_t(\boldsymbol{\theta}_0)|^k}{(\sigma_0 - \epsilon)^k}\right)^2 \right\}^{1/2} \left\{ \sum_{t=-q+1}^{n} (t+q)^{r-1} d^t \left(\sum_{j=-\infty}^{\infty} d^{|j|} |X_{t-j}|\right)^2 \right\}^{1/2} \\
&\leq \left(\sum_{t=-q+1}^{n} (t+q)^{2(r-1)} d^t \right)^{1/2} \left(\sum_{t=-q+1}^{n} d^t \left(1 + \frac{|z_t(\boldsymbol{\theta}_0)|^k}{(\sigma_0 - \epsilon)^k}\right)^4 \right)^{1/4}
\end{aligned}
$$

$$\times\left(\sum_{t=-q+1}^{n} d^t\left(\sum_j d^{|j|}|X_{t-j}|\right)^4\right)^{1/4}, \tag{S8}$$

where the series in the first parenthesis is finite and the series in the second parenthesis converges by Lemma S1, in which the required condition is satisfied:

$$\begin{aligned}
E\log^+\left(1+\frac{|z_t(\boldsymbol{\theta}_0)|^k}{(\sigma_0-\epsilon)^k}\right)^4 &= 4E\log\left(1+\frac{|z_t(\boldsymbol{\theta}_0)|^k}{(\sigma_0-\epsilon)^k}\right)\\
&\le 4E\left[\log\left(1+\frac{|z_t(\boldsymbol{\theta}_0)|^k}{(\sigma_0-\epsilon)^k}\right)1_{\{|z_t(\boldsymbol{\theta}_0)|\le\sigma_0-\epsilon\}}\right]+4E\left[\log\left(1+\frac{|z_t(\boldsymbol{\theta}_0)|^k}{(\sigma_0-\epsilon)^k}\right)1_{\{|z_t(\boldsymbol{\theta}_0)|>\sigma_0-\epsilon\}}\right]\\
&\le 4(\log 2)P(|z_t(\boldsymbol{\theta}_0)|\le\sigma_0-\epsilon)+4E\left[\log\left(2\frac{|z_t(\boldsymbol{\theta}_0)|^k}{(\sigma_0-\epsilon)^k}\right)1_{\{|z_t(\boldsymbol{\theta}_0)|>\sigma_0-\epsilon\}}\right]\\
&\le 4(\log 2)+4k\frac{E|z_t(\boldsymbol{\theta}_0)|}{\sigma_0-\epsilon}\\
&< \infty,
\end{aligned}$$

(we use $\log x < x$ for $x > 0$ to obtain the third inequality). The series in the third parenthesis of (S8) converges almost surely by Lemma S2 which satisfies the required condition:

$$E\left(\sum_{j=-\infty}^{\infty} d^{|j|}|X_{t-j}|\right)^4 \le EX_t^4\left(\sum_{j=-\infty}^{\infty} d^{|j|}\right)^4 < \infty.$$

Under similar derivations, the first summation in A_{21} and each term in A_{22} and A_{23} can also be shown to converge almost surely and therefore A_2 converges almost surely. To complete the proof for Lemma S6, we have

$$\sup_{\boldsymbol{\psi}\in Q_\epsilon}\frac{1}{n}|\ell_n(\boldsymbol{\psi})-\hat{\ell}_n(\boldsymbol{\psi})| \le n^{-1}\sup_{\boldsymbol{\psi}\in Q_\epsilon}(A_1+A_2)\to 0, \quad \text{a.s.}$$

□

Lemma S7 *Under the assumptions (A1)–(A9) given in Theorem 1,*

$$n^{-1/2}\sum_{t=-q+1}^{n}\left[\frac{\partial}{\partial\boldsymbol{\psi}}\log f_\sigma(\hat{z}_t(\boldsymbol{\theta}))-\frac{\partial}{\partial\boldsymbol{\psi}}\log f_\sigma(z_t(\boldsymbol{\theta}))\right]_{\boldsymbol{\psi}=\boldsymbol{\psi}_0}\to\mathbf{0}. \tag{S9}$$

Proof: The parameter vector $\boldsymbol{\psi}$ contains $\boldsymbol{\theta}$ and σ. We first consider the partial derivative with respect to $\boldsymbol{\theta}$:

$$\begin{aligned}
\left|\frac{\partial}{\partial\boldsymbol{\theta}}\log f_\sigma(\hat{z}_t(\boldsymbol{\theta}))-\frac{\partial}{\partial\boldsymbol{\theta}}\log f_\sigma(z_t(\boldsymbol{\theta}))\right| &= \left|\frac{f'_\sigma}{f_\sigma}(\hat{z}_t(\boldsymbol{\theta}))\frac{\partial}{\partial\boldsymbol{\theta}}\hat{z}_t(\boldsymbol{\theta})-\frac{f'_\sigma}{f_\sigma}(z_t(\boldsymbol{\theta}))\frac{\partial}{\partial\boldsymbol{\theta}}z_t(\boldsymbol{\theta})\right|\\
&\le \frac{1}{\sigma}\left|\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta})}{\sigma}\right)\right|\left|\frac{\partial}{\partial\boldsymbol{\theta}}(\hat{z}_t(\boldsymbol{\theta})-z_t(\boldsymbol{\theta}))\right|+\frac{1}{\sigma}\left|\frac{\partial}{\partial\boldsymbol{\theta}}z_t(\boldsymbol{\theta})\right|\left|\frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta})}{\sigma}\right)-\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta})}{\sigma}\right)\right|\\
&\quad+\frac{1}{\sigma}\left|\frac{\partial}{\partial\boldsymbol{\theta}}(\hat{z}_t(\boldsymbol{\theta})-z_t(\boldsymbol{\theta}))\right|\left|\frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta})}{\sigma}\right)-\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta})}{\sigma}\right)\right|\\
&\equiv B_{1t}+B_{2t}+B_{3t}.
\end{aligned} \tag{S10}$$

Accordingly, the derivative with respect to $\boldsymbol{\theta}$ in (S9) can be bounded by the sum of $\{B_{1t}\}, \{B_{2t}\}$ and $\{B_{3t}\}$ subject to the decomposition in (S10). By Lemma S4, the first term of (S9) subject to B_{1t} in (S10) evaluated at $\boldsymbol{\psi}_0$ satisfies

$$
\begin{aligned}
n^{-1/2}\left[\sum_{t=-q+1}^{n} B_{1t}\right]_{\boldsymbol{\psi}=\boldsymbol{\psi}_0} &= n^{-1/2}\sigma_0^{-1}\sum_{t=-q+1}^{n}\left|\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta}_0)}{\sigma_0}\right)\right|\left|\frac{\partial}{\partial\boldsymbol{\theta}}\left(\hat{z}_t(\boldsymbol{\theta})-z_t(\boldsymbol{\theta})\right)\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&\le n^{-1/2}\sigma_0^{-1}\sum_{t=-q+1}^{n}\left|\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta}_0)}{\sigma_0}\right)\right|\left\{C_3(n-t)^s d^{n-t}|\hat{\boldsymbol{w}}_s-\boldsymbol{w}_s| + C_4(t+q)^r d^t|\hat{\boldsymbol{z}}_r-\boldsymbol{z}_r|\right\},
\end{aligned}
$$

in which, by Lemma S2, the two series converge almost surely since

$$
\sum_{t=-q+1}^{n}(n-t)^s d^{n-t}<\infty, \quad \sum_{t=-q+1}^{n}(t+q)^r d^t<\infty, \quad \text{and} \quad E\left|\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta}_0)}{\sigma_0}\right)\right|<\infty.
$$

Therefore, $n^{-1/2}[\sum_{t=-1}^{n} B_{1t}]_{\boldsymbol{\psi}=\boldsymbol{\psi}_0}$ converges to zero almost surely. According to Lemma S5 and Assumption (A9), the part of (S9) subject to B_{2t} in (S10) satisfies

$$
\begin{aligned}
n^{-1/2}\left[\sum_{t=-q+1}^{n} B_{2t}\right]_{\boldsymbol{\psi}=\boldsymbol{\psi}_0} &= n^{-1/2}\sigma_0^{-1}\sum_{t=-q+1}^{n}\left|\frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta}_0)}{\sigma_0}\right)-\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta}_0)}{\sigma_0}\right)\right|\left|\frac{\partial}{\partial\boldsymbol{\theta}}z_t(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&\le n^{-1/2}\sigma_0^{-1}A\sum_{t=-q+1}^{n}\left\{\left(1+\frac{|z_t(\boldsymbol{\theta}_0)|^k}{\sigma_0^k}\right)\frac{|\hat{z}_t(\boldsymbol{\theta}_0)-z_t(\boldsymbol{\theta}_0)|}{\sigma_0}+\frac{|\hat{z}_t(\boldsymbol{\theta}_0)-z_t(\boldsymbol{\theta}_0)|^\ell}{\sigma_0^\ell}\right\}\left\{\sum_{j=-\infty}^{\infty}|\gamma_j Z_{t-j}|\right\}.
\end{aligned}
$$

Following the arguments similar to those in simplifying (S8), one can show that $n^{-1/2}[\sum_{t=-q+1}^{n} B_{2t}]_{\boldsymbol{\psi}=\boldsymbol{\psi}_0}$ converges to zero almost surely by Lemma S2 since

$$
E\left(\sum_{j=-\infty}^{\infty}|\gamma_j||Z_{t-j}|\right)^4 = (E|Z_t|)^4\left(\sum_{j=-\infty}^{\infty}|\gamma_j|\right)^4<\infty.
$$

Finally, the remaining part of (S9), $n^{-1/2}[\sum_{t=-q+1}^{n} B_{3t}]_{\boldsymbol{\psi}=\boldsymbol{\psi}_0}$, subject to B_{3t} satisfies

$$
\begin{aligned}
n^{-1/2}\left[\sum_{t=-q+1}^{n} B_{3t}\right]_{\boldsymbol{\psi}=\boldsymbol{\psi}_0} &= n^{-1/2}\sigma_0^{-1}\sum_{t=-q+1}^{n}\left|\frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta}_0)}{\sigma_0}\right)-\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta}_0)}{\sigma_0}\right)\right|\left|\frac{\partial}{\partial\boldsymbol{\theta}}\left(\hat{z}_t(\boldsymbol{\theta})-z_t(\boldsymbol{\theta})\right)\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&\le A n^{-1/2}\sigma_0^{-1}\sum_{t=-q+1}^{n}\left\{\left(1+\frac{|z_t(\boldsymbol{\theta}_0)|^k}{\sigma_0^k}\right)\sigma_0^{-1}\left|\hat{z}_t(\boldsymbol{\theta}_0)-z_t(\boldsymbol{\theta}_0)\right|+\sigma_0^{-\ell}\left|\hat{z}_t(\boldsymbol{\theta}_0)-z_t(\boldsymbol{\theta}_0)\right|^\ell\right\} \\
&\qquad \times\left\{C_3(n-t)^s d^{n-t}|\hat{\boldsymbol{w}}_s-\boldsymbol{w}_s|+C_4(t+q)^r d^t|\hat{\boldsymbol{z}}_r-\boldsymbol{z}_r|\right\} \\
&\to \mathbf{0}, \quad a.s.,
\end{aligned}
$$

in which every cross product series can be shown to converge almost surely using the techniques given in the proof of Lemma S6.

To complete the proof for Lemma S7, we then consider the partial derivative with respect to σ:

$$\begin{aligned}
&\left|\frac{\partial}{\partial\sigma}\log f_\sigma(\hat{z}_t(\boldsymbol{\theta})) - \frac{\partial}{\partial\sigma}\log f_\sigma(z_t(\boldsymbol{\theta}))\right| = \frac{1}{\sigma^2}\left|\frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta})}{\sigma}\right)\hat{z}_t(\boldsymbol{\theta}) - \frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta})}{\sigma}\right)z_t(\boldsymbol{\theta})\right| \\
\leq\ & \frac{1}{\sigma^2}\left|\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta})}{\sigma}\right)\right||\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})| + \frac{1}{\sigma^2}|z_t(\boldsymbol{\theta})|\left|\frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta})}{\sigma}\right) - \frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta})}{\sigma}\right)\right| \\
& + \frac{1}{\sigma^2}|\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})|\left|\frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta})}{\sigma}\right) - \frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta})}{\sigma}\right)\right| \\
\equiv\ & B_{4t} + B_{5t} + B_{6t}. \qquad \text{(S11)}
\end{aligned}$$

According to this decomposition (S11), the derivative with respect to σ in Equation (S9) can be bounded by three series subject to $\{B_{4t}\}, \{B_{5t}\}$ and $\{B_{6t}\}$. Again, under similar derivations for showing the convergence of A_1 and B_{2t}, one can show that $\sum_t B_{4t}$, $\sum_t B_{5t}$ and $\sum_t B_{6t}$ converge almost surely at $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. Therefore, the proof is complete. □

Lemma S8 *Under the assumptions (A1)–(A9) given in Theorem 1,*

$$n^{-1}\left(\hat{\boldsymbol{B}}(\boldsymbol{\psi}_n^*) - \boldsymbol{B}(\boldsymbol{\psi}_0)\right) \xrightarrow{P} \mathbf{0},$$

where $\{\boldsymbol{\psi}_n^*\}$ *is a sequence converging to* $\boldsymbol{\psi}_0$*,*

$$\boldsymbol{B}(\boldsymbol{\psi}) \equiv \begin{pmatrix} \boldsymbol{B}_{\theta\theta}(\boldsymbol{\psi}) & \boldsymbol{B}_{\theta\sigma}(\boldsymbol{\psi}) \\ \boldsymbol{B}_{\sigma\theta}(\boldsymbol{\psi}) & \boldsymbol{B}_{\sigma\sigma}(\boldsymbol{\psi}) \end{pmatrix} \equiv \sum_{t=-q+1}^{n} \frac{\partial^2}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\log f_\sigma(z_t(\boldsymbol{\theta})),$$

and $\hat{\boldsymbol{B}}(\boldsymbol{\psi})$ *is the second derivative of* $\sum_t \log f_\sigma(\hat{z}_t(\boldsymbol{\theta}))$*.*

Proof: First, $\hat{\boldsymbol{B}}(\boldsymbol{\psi}_n^*) - \boldsymbol{B}(\boldsymbol{\psi}_0)$ is decomposed as two parts:

$$\hat{\boldsymbol{B}}(\boldsymbol{\psi}_n^*) - \boldsymbol{B}(\boldsymbol{\psi}_0) = (\boldsymbol{B}(\boldsymbol{\psi}_n^*) - \boldsymbol{B}(\boldsymbol{\psi}_0)) + (\hat{\boldsymbol{B}}(\boldsymbol{\psi}_n^*) - \boldsymbol{B}(\boldsymbol{\psi}_n^*)).$$

Lii and Rosenblatt (1992) have shown that

$$n^{-1}(\boldsymbol{B}(\boldsymbol{\psi}_n^*) - \boldsymbol{B}(\boldsymbol{\psi}_0)) \xrightarrow{P} \mathbf{0}.$$

Therefore, it is enough to show

$$n^{-1}(\hat{\boldsymbol{B}}(\boldsymbol{\psi}_n^*) - \boldsymbol{B}(\boldsymbol{\psi}_n^*)) \xrightarrow{P} \mathbf{0}.$$

Let $\boldsymbol{\psi}_n^* = (\boldsymbol{\theta}_n^*, \sigma_n^*)'$. Since

$$\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log f_\sigma(z_t(\boldsymbol{\theta})) = \frac{1}{\sigma}\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta})}{\sigma}\right)\frac{\partial^2 z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \frac{1}{\sigma^2}\left(\frac{f'}{f}\right)'\left(\frac{z_t(\boldsymbol{\theta})}{\sigma}\right)\frac{\partial z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'},$$

we have

$$\left|\hat{\boldsymbol{B}}_{\theta\theta}(\boldsymbol{\psi}_n^*) - \boldsymbol{B}_{\theta\theta}(\boldsymbol{\psi}_n^*)\right| \le \sum_{t=-q+1}^{n} \left| \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \log f_\sigma(\hat{z}_t(\boldsymbol{\theta})) - \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \log f_\sigma(z_t(\boldsymbol{\theta}))\right|_{\boldsymbol{\psi}=\boldsymbol{\psi}_n^*}$$

$$\le \frac{1}{\sigma_n^*}\sum_{t=-q+1}^{n}\left|\frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\left[\frac{\partial^2\,\hat{z}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*} - \frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\left[\frac{\partial^2\,z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}\right|$$

$$+\frac{1}{\sigma_n^{*2}}\sum_{t=-q+1}^{n}\left|\left(\frac{f'}{f}\right)'\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\left[\frac{\partial\,\hat{z}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\,\hat{z}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*} - \left(\frac{f'}{f}\right)'\left(\frac{z_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\left[\frac{\partial\,z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\,z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}\right|$$

$$\le \frac{1}{\sigma_n^*}\sum_{t=-q+1}^{n}\left|\frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\right|\left|\frac{\partial^2\left(\hat{z}_t(\boldsymbol{\theta})-z_t(\boldsymbol{\theta})\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}$$

$$+\frac{1}{\sigma_n^*}\sum_{t=-q+1}^{n}\left|\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right) - \frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\right|\left|\frac{\partial^2\,z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}$$

$$+\frac{1}{\sigma_n^{*2}}\sum_{t=-q+1}^{n}\left\{\left|\left(\frac{f'}{f}\right)'\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\right|\left|\frac{\partial\left(\hat{z}_t(\boldsymbol{\theta})-z_t(\boldsymbol{\theta})\right)}{\partial\boldsymbol{\theta}}\frac{\partial\left(\hat{z}_t(\boldsymbol{\theta})-z_t(\boldsymbol{\theta})\right)}{\partial\boldsymbol{\theta}'} + \frac{\partial\left(\hat{z}_t(\boldsymbol{\theta})-z_t(\boldsymbol{\theta})\right)}{\partial\boldsymbol{\theta}}\frac{\partial\,z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right.\right.$$

$$\left.\left. + \frac{\partial\,z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\left(\hat{z}_t(\boldsymbol{\theta})-z_t(\boldsymbol{\theta})\right)}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}\right\}$$

$$+\frac{1}{\sigma_n^{*2}}\sum_{t=-q+1}^{n}\left|\left(\frac{f'}{f}\right)'\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right) - \left(\frac{f'}{f}\right)'\left(\frac{z_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\right|\left|\frac{\partial\,z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\,z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}. \tag{S12}$$

Each series in (S12) can be shown to converge almost surely using similar techniques in proving previous lemmas. Moreover, for large enough n such that $\boldsymbol{\psi}_n^* \in Q_\epsilon$, since

$$\frac{\partial^2}{\partial\sigma^2}\log f_\sigma(z_t(\boldsymbol{\theta})) = \frac{z_t^2(\boldsymbol{\theta})}{\sigma^4}\left(\frac{f'}{f}\right)'\left(\frac{z_t(\boldsymbol{\theta})}{\sigma}\right) + \frac{2\,z_t(\boldsymbol{\theta})}{\sigma^3}\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta})}{\sigma}\right),$$

we have

$$|\hat{\boldsymbol{B}}_{\sigma\sigma}(\boldsymbol{\psi}_n^*) - \boldsymbol{B}_{\sigma\sigma}(\boldsymbol{\psi}_n^*)| \le \frac{1}{\sigma_n^{*4}}\sum_{n=-q+1}^{n}\left|\left(\frac{f'}{f}\right)'\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\hat{z}_t^2(\boldsymbol{\theta}_n^*) - \left(\frac{f'}{f}\right)'\left(\frac{z_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)z_t^2(\boldsymbol{\theta}_n^*)\right|$$

$$+\frac{2}{\sigma_n^{*3}}\sum_{n=-q+1}^{n}\left|\frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\hat{z}_t(\boldsymbol{\theta}_n^*) - \frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)z_t(\boldsymbol{\theta}_n^*)\right|. \tag{S13}$$

Ignoring the scalar $1/\sigma_n^{*4}$, the first term on the right hand side of (S13) is bounded by the following:

$$\sum_{n=-q+1}^{n}\left|\left(\frac{f'}{f}\right)'\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right) - \left(\frac{f'}{f}\right)'\left(\frac{z_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\right|z_t^2(\boldsymbol{\theta}_n^*) + \left|\left(\frac{f'}{f}\right)'\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\right||\hat{z}_t^2(\boldsymbol{\theta}_n^*) - z_t^2(\boldsymbol{\theta}_n^*)|$$

$$\le \sum_{n=-q+1}^{n}\left\{\left|\left(\frac{f'}{f}\right)'\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right) - \left(\frac{f'}{f}\right)'\left(\frac{z_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\right|z_t^2(\boldsymbol{\theta}_n^*) + \left\{(\hat{z}_t(\boldsymbol{\theta}_n^*) - z_t(\boldsymbol{\theta}_n^*))^2 + 2|z_t(\boldsymbol{\theta}_n^*)||\hat{z}_t(\boldsymbol{\theta}_n^*) - z_t(\boldsymbol{\theta}_n^*)|\right\}\right.$$

$$\left.\times\left|\left(\frac{f'}{f}\right)'\left(\frac{z_t(\boldsymbol{\theta}_0)}{\sigma_0}\right) + \left\{\left(\frac{f'}{f}\right)'\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right) - \left(\frac{f'}{f}\right)'\left(\frac{z_t(\boldsymbol{\theta}_0)}{\sigma_0}\right)\right\}\right|\right\},$$

in which every series in the decomposition converges almost surely under analogous proofs given in previous lemmas. The second term on the right hand side of (S13) is identical to

$$\frac{2}{\sigma_n^*}\sum_{t=-q+1}^{n}\left|\frac{\partial}{\partial\sigma}\log f_\sigma(\hat{z}_t(\boldsymbol{\theta}))-\frac{\partial}{\partial\sigma}\log f_\sigma(z_t(\boldsymbol{\theta}))\right|_{\boldsymbol{\psi}=\boldsymbol{\psi}_n^*},$$

which also converges with probability one based on the same decomposition given in (S11).

Finally, we consider $\boldsymbol{B}_{\theta\sigma}$. Since

$$\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\sigma}\log f_\sigma(z_t(\boldsymbol{\theta}))=-\frac{1}{\sigma^3}\left(\frac{f'}{f}\right)'\left(\frac{z_t(\boldsymbol{\theta})}{\sigma}\right)z_t(\boldsymbol{\theta})\frac{\partial z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}-\frac{1}{\sigma^2}\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta})}{\sigma}\right)\frac{\partial z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}},$$

we have

$$\begin{aligned}
|\hat{\boldsymbol{B}}_{\theta\sigma}(\boldsymbol{\psi}_n^*)-\boldsymbol{B}_{\theta\sigma}(\boldsymbol{\psi}_n^*)| &\le \sum_{t=-q+1}^{n}\left|\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\sigma}\log f_\sigma(\hat{z}_t(\boldsymbol{\theta}))-\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\sigma}\log f_\sigma(z_t(\boldsymbol{\theta}))\right|_{\boldsymbol{\psi}=\boldsymbol{\psi}_n^*}\\
\le\ & \frac{1}{\sigma_n^{*3}}\sum_{t=-q+1}^{n}\left|\left(\frac{f'}{f}\right)'\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\hat{z}_t(\boldsymbol{\theta}_n^*)\left[\frac{\partial\hat{z}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}-\left(\frac{f'}{f}\right)'\left(\frac{z_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)z_t(\boldsymbol{\theta}_n^*)\left[\frac{\partial z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}\right|\\
&+\frac{1}{\sigma_n^{*2}}\sum_{t=-q+1}^{n}\left|\frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\left[\frac{\partial\hat{z}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}-\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\left[\frac{\partial z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}\right|\\
\le\ & \frac{1}{\sigma_n^{*3}}\sum_{t=-q+1}^{n}\left|\left(\frac{f'}{f}\right)'\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)(\hat{z}_t(\boldsymbol{\theta}_n^*)-z_t(\boldsymbol{\theta}_n^*))\right|\left|\frac{\partial(\hat{z}_t(\boldsymbol{\theta})-z_t(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}\\
&+\frac{1}{\sigma_n^{*3}}\sum_{t=-q+1}^{n}\left|\left(\frac{f'}{f}\right)'\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\right|\left|z_t(\boldsymbol{\theta}_n^*)\frac{\partial(\hat{z}_t(\boldsymbol{\theta})-z_t(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}}+(\hat{z}_t(\boldsymbol{\theta})-z_t(\boldsymbol{\theta}))\frac{\partial z_t(\boldsymbol{\theta}_n^*)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}\\
&+\frac{1}{\sigma_n^{*3}}\sum_{t=-q+1}^{n}\left|\left(\frac{f'}{f}\right)'\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)-\left(\frac{f'}{f}\right)'\left(\frac{z_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\right||z_t(\boldsymbol{\theta}_n^*)|\left|\frac{\partial z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}\\
&+\frac{1}{\sigma_n^{*2}}\sum_{t=-q+1}^{n}\left|\frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\right|\left|\frac{\partial(\hat{z}_t(\boldsymbol{\theta})-z_t(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}\\
&+\frac{1}{\sigma_n^{*2}}\sum_{t=-q+1}^{n}\left|\frac{f'}{f}\left(\frac{\hat{z}_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)-\frac{f'}{f}\left(\frac{z_t(\boldsymbol{\theta}_n^*)}{\sigma_n^*}\right)\right|\left|\frac{\partial z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}. \qquad \text{(S14)}
\end{aligned}$$

Analogously, each series in the decomposition (S14) converges with probability one. Consequently, $\hat{\boldsymbol{B}}(\boldsymbol{\psi}_n^*)-\boldsymbol{B}(\boldsymbol{\psi}_n^*)$ converges with probability one and $n^{-1}(\hat{\boldsymbol{B}}(\boldsymbol{\psi}_n^*)-\boldsymbol{B}(\boldsymbol{\psi}_n^*))\xrightarrow{P}\boldsymbol{0}$. □

Proof of Theorem 1: According to Lemma S6, as $n\to\infty$, there exists a sequence of solutions $\{\hat{\boldsymbol{\psi}}_n\}$ to the proposed likelihood equations

$$\frac{\partial}{\partial\boldsymbol{\psi}}\hat{\ell}_n(\boldsymbol{\psi})=\boldsymbol{0},$$

which satisfy $\boldsymbol{\psi}_n\in Q_\epsilon$ and $\boldsymbol{\psi}_n\to\boldsymbol{\psi}_0$ with probability one. By Taylor expansion, we have

$$\boldsymbol{0} \;=\; n^{-1/2}\left[\frac{\partial}{\partial\boldsymbol{\psi}}\hat{\ell}_n(\boldsymbol{\psi})\right]_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}_n}=n^{-1/2}\sum_{t=-q+1}^{n}\left[\frac{\partial}{\partial\boldsymbol{\psi}}\log f_\sigma(\hat{z}_t(\boldsymbol{\theta}))\right]_{\boldsymbol{\psi}=\boldsymbol{\psi}_0}+n^{-1}\hat{\boldsymbol{B}}(\boldsymbol{\psi}_n^*)\,n^{1/2}(\hat{\boldsymbol{\psi}}_n-\boldsymbol{\psi}_0)$$

$$
\begin{aligned}
&= n^{-1/2} \sum_{t=-q+1}^{n} \left[\frac{\partial}{\partial \boldsymbol{\psi}} \log f_\sigma(z_t(\boldsymbol{\theta})) \right]_{\boldsymbol{\psi}=\boldsymbol{\psi}_0} + n^{-1/2} \sum_{t=-q+1}^{n} \left[\frac{\partial}{\partial \boldsymbol{\psi}} \log f_\sigma(\hat{z}_t(\boldsymbol{\theta})) - \frac{\partial}{\partial \boldsymbol{\psi}} \log f_\sigma(z_t(\boldsymbol{\theta})) \right]_{\boldsymbol{\psi}=\boldsymbol{\psi}_0} \\
&\quad + n^{-1} \boldsymbol{B}(\boldsymbol{\psi}_0)\, n^{1/2}(\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0) + n^{-1} \left[\hat{\boldsymbol{B}}(\boldsymbol{\psi}_n^*) - \boldsymbol{B}(\boldsymbol{\psi}_0) \right] n^{1/2}(\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0),
\end{aligned}
$$

where $\boldsymbol{\psi}_n^*$ is between $\hat{\boldsymbol{\psi}}_n$ and $\boldsymbol{\psi}_0$. According to Lemmas S7 and S8,

$$
\mathbf{0} = n^{-1/2} \sum_{t=-q+1}^{n} \left[\frac{\partial}{\partial \boldsymbol{\psi}} \log f_\sigma(z_t(\boldsymbol{\theta})) \right]_{\boldsymbol{\psi}=\boldsymbol{\psi}_0} + n^{-1} \boldsymbol{B}(\boldsymbol{\psi}_0)\, n^{1/2}(\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0) + o_p(1),
$$

which implies

$$
n^{1/2}(\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0) = \left[n^{-1} \boldsymbol{B}(\boldsymbol{\psi}_0) \right]^{-1} \left\{ -n^{-1/2} \sum_{t=-q+1}^{n} \left[\frac{\partial}{\partial \boldsymbol{\psi}} \log f_\sigma(z_t(\boldsymbol{\theta})) \right]_{\boldsymbol{\psi}=\boldsymbol{\psi}_0} \right\} + o_p(1).
$$

The first term in the above equation is the dominant term and it converges to the limit distribution of the normalized quasi-MLE which has been shown to be $N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-1}\right)$ (Lii and Rosenblatt, 1992), where $\boldsymbol{\Sigma}$ is given in Equation (1.7) in Lii and Rosenblatt (1992). Since the joint MLE $\hat{\boldsymbol{\psi}}_J$ and the conditional MLE $\hat{\boldsymbol{\psi}}_c$ are the maximizers of the form $\hat{\ell}_n(\boldsymbol{\psi})$ with different initial setups, they are both asymptotically equivalent to the quasi-MLE. Namely,

$$
\begin{aligned}
n^{1/2}\left(\hat{\boldsymbol{\psi}}_c - \boldsymbol{\psi}_0\right) &\xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-1}\right), \\
n^{1/2}\left(\hat{\boldsymbol{\psi}}_J - \boldsymbol{\psi}_0\right) &\xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-1}\right).
\end{aligned}
$$

□

References

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd ed., Springer-Verlag, New York.

Durrett, R. (2005). *Probability: Theory and Examples*, 3rd ed., Duxbury.

Lii, K.-S. and Rosenblatt, M. (1992). An approximate maximum likelihood estimation for non-Gaussian non-minimum phase moving average processes. *Journal of Multivariate Analysis*, **43**, 272–299.