Anders Rahbek
Rasmus Søndergaard Pedersen
University of Copenhagen

October 2024

Part VI
Statistical inference for time series models

In Part II we considered the properties of the maximum likelihood (ML) estimators for AR and VAR models. A neat feature of these estimators is that they can be written in closed-form as a solution to the problem of maximizing the log-likelihood function. Given this closed form, limit theorems, such as a LLN and CLT, can be applied directly to show consistency and asymptotic normality of the estimators. In this chapter we consider a much more general setting, where an estimator is a maximizer of a criterion function. This class of estimators is labelled *extremum estimators*, and covers a wide range of estimators such as least squares and ML. We state conditions ensuring consistency and asymptotic normality of such estimators. The estimators may not be given in closed-form, as is the case for the ML estimators for ARCH models, and the properties of the estimator are derived from the properties of the criterion function and its derivatives. We consider the AR(1) and ARCH(1) models as running examples and prove that the estimators are consistent and asymptotically normal under the assumption that the data-generating process (DGP) is stationary and ergodic with certain moments finite. To do so, we rely on the limit theorems presented in Part V. *We emphasize that, alternatively, one may consider geometrically ergodic DGPs and apply the limit theory presented in Part I.* Lastly we discuss how the results can be used for hypothesis testing and present potential extensions. The general results for extremum estimators presented are based on Newey and McFadden (1994) [NM henceforth]. We emphasize that several of the assumptions and arguments can be relaxed and refined, and we refer to Amemiya (1985) for many more details as well as Francq and Zakoïan (2019) with particular emphasis on estimation of conditional heteroskedasticity models such as the ARCH(1).

VI.1 Extremum estimators

Suppose that we seek to estimate a vector of parameters in an econometric model. We assume that this parameter vector, θ, is of finite dimension and belongs to a parameter space $\Theta \subseteq \mathbb{R}^k$ for some $k \geq 1$. Given a data set $\{X_t\}_{t=1,\dots,T}$ of length $T \geq 1$, we consider an objective function $Q_T(\theta) := f(\{X_t\}_{t=1}^T; \theta)$, with f measurable. The extremum estimator of θ, labelled $\hat{\theta}_T$, is defined as any solution to

$$\hat{\theta}_T = \arg\max_{\theta \in \Theta} Q_T(\theta). \tag{VI.1}$$

Example VI.1.1 (AR(1)) *Consider the AR(1) model given by*

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z},$$

with $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an i.i.d.$N(0, \sigma^2)$ process. As considered in Part II, for a set of observations $\{x_t\}_{t=0}^T$, the log-likelihood function (conditional on x_0) is given by

$$\frac{1}{T}\sum_{t=1}^{T} \log\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_t - \rho x_{t-1})^2}{2\sigma^2}\right)\right].$$

Treating σ^2 as known (fixed), the single parameter of interest is $\rho =: \theta$, and the parameter space is given by $\Theta = \mathbb{R}$. In this case, the objective function is given by

$$Q_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} q_t(\theta), \quad q_t(\theta) = -\frac{1}{2}(x_t - \theta x_{t-1})^2. \tag{VI.2}$$

Recall also from Part II that the ML estimator is given by,

$$\hat{\theta}_T = \frac{T^{-1}\sum_{t=1}^{T} x_{t-1}x_t}{T^{-1}\sum_{t=1}^{T} x_{t-1}^2}. \tag{VI.3}$$

Although the estimator is given in closed form, for illustrative purposes, we consider the estimation of AR(1) models as an ongoing example and consider the stochastic properties of the objective function $Q_T(\theta)$ and its derivatives later on.

Example VI.1.2 (ARCH(1)) *Consider the ARCH(1) model given by*

$$\begin{aligned} x_t &= \sigma_t z_t, \quad t \in \mathbb{Z} \\ \sigma_t^2 &= \omega + \alpha x_{t-1}^2, \end{aligned}$$

with $\{z_t\}_{t\in\mathbb{Z}}$ *an i.i.d. process with* $z_t \overset{D}{=} N(0,1)$*. Moreover, the parameter vector is given by* $\theta = (\omega, \alpha)'$ *with* $\omega > 0$ *and* $\alpha \geq 0$*. Recall from Part I that* x_t *has conditional density*

$$f(x_t|x_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{x_t^2}{2\sigma_t^2}\right).$$

Consequently, for a sample $\{x_t\}_{t=0}^T$*, the log-likelihood function (conditional on* x_0*) is given by*

$$Q_T(\theta) = \sum_{t=1}^T q_t(\theta), \quad q_t(\theta) = \log\left[\frac{1}{\sqrt{2\pi\sigma_t^2(\theta)}} \exp\left(-\frac{x_t^2}{2\sigma_t^2(\theta)}\right)\right], \tag{VI.4}$$

with $\sigma_t^2(\theta) = \omega + \alpha x_{t-1}^2$*. The parameter space* $\Theta \subseteq (0,\infty) \times [0,\infty)$ *is considered in more detail later.*

VI.2 Consistency

In this section, we consider conditions ensuring that the extremum estimator in (VI.1) has a non-stochastic probability limit, θ_0, as the sample size T diverges. Suppose that $Q_T(\theta)$ has a non-stochastic probability limit $Q(\theta)$ as $T \to \infty$, and suppose that $Q(\theta)$ is uniquely maximized at some point θ_0. Then we say that θ_0 is the true value of θ. This value, inherently, depends on (or is determined from) the underlying DGP. Likewise, $Q(\theta)$ depends on the DGP, as discussed in more detail in the next examples. As mentioned, throughout we consider only DGPs that are stationary and ergodic, which for notational convenience we label $\{x_t\}_{t\in\mathbb{Z}}$ (and not $\{x_t^*\}_{t\in\mathbb{Z}}$).

Example VI.2.1 (AR(1) ctd.) *Consider the objective function in (VI.2) with* $\Theta = \mathbb{R}$*. Suppose that the DGP is given in terms of* $\theta_0 = \rho_0$ *with* $|\rho_0| < 1$ *and that* $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ *is an i.i.d. process with* $\mathbb{E}[\varepsilon_t] = 0$ *and* $\mathbb{V}[\varepsilon_t] = \sigma_0^2 < \infty$*. From Example V.4.1, we know that the DGP is stationary and ergodic, and we have that* $\mathbb{E}[x_t] = 0$*,* $\mathbb{V}[x_t] = \sigma_0^2/(1-\theta_0^2)$ *and* $\mathbb{E}[x_t x_{t-1}] = \theta_0\sigma_0^2/(1-\theta_0^2)$ *for all* t*. Note that for any* $\theta \in \Theta$*,*

$$\begin{aligned}
\mathbb{E}[|q_t(\theta)|] &= \frac{1}{2}\mathbb{E}[|x_t^2 + \theta^2 x_{t-1}^2 - 2\theta x_t x_{t-1}|] \\
&\leq \frac{1}{2}\left(\mathbb{E}[x_t^2] + \theta^2\mathbb{E}[x_{t-1}^2] + 2|\theta|\mathbb{E}[|x_t x_{t-1}|]\right) \quad \textit{(triangle)} \\
&\leq \frac{1}{2}\left(\mathbb{E}[x_t^2] + \theta^2\mathbb{E}[x_{t-1}^2] + 2|\theta|\sqrt{\mathbb{E}[x_t^2]\mathbb{E}[x_{t-1}^2]}\right) \quad \textit{(Cauchy-Schwarz)} \\
&= \frac{1}{2}\left(1 + \theta^2 + 2|\theta|\right)\mathbb{E}[x_t^2] < \infty,
\end{aligned}$$

where the last equality follows by stationarity. Moreover,

$$\begin{aligned}\mathbb{E}\left[q_t(\theta)\right] &= -\frac{1}{2}\mathbb{E}[x_t^2 + \theta^2 x_{t-1}^2 - 2\theta x_t x_{t-1}] \\ &= -\frac{1}{2}\left(\mathbb{E}[x_t^2] + \theta^2\mathbb{E}[x_{t-1}^2] - 2\theta\mathbb{E}[x_t x_{t-1}]\right) \\ &= -\frac{1}{2}\left(\frac{\sigma_0^2}{1-\theta_0^2}\right)\left(1+\theta^2 - 2\theta_0\theta\right) =: Q(\theta)\end{aligned}$$

which is maximized at $\theta = \theta_0$.
Using Theorem V.2.1, noting for a fixed θ, $q_t(\theta)$ is a function of the stationary and ergodic $\{x_t\}_{t\in\mathbb{Z}}$ with $\mathbb{E}[|q_t(\theta)|] < \infty$, we have that $\{q_t(\theta)\}_{t\in\mathbb{Z}}$ itself is stationary and ergodic. Lastly, by Theorem V.2.2, as $T \to \infty$,

$$Q_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} q_t(\theta) \xrightarrow{p} \mathbb{E}\left[q_t(\theta)\right] = Q(\theta).$$

Example VI.2.2 (ARCH(1) ctd.) *Consider the log-likelihood function $Q_T(\theta)$ in (VI.4) with $\Theta \subseteq (0,\infty)\times[0,\infty)$ and*

$$\begin{aligned}q_t(\theta) &= -\frac{1}{2}\left[\log(2\pi) + \log(\sigma_t^2(\theta)) + \frac{x_t^2}{\sigma_t^2(\theta)}\right] \\ &= -\frac{1}{2}\left[\log(2\pi) + \log(\omega + \alpha x_{t-1}^2) + \frac{x_t^2}{\omega + \alpha x_{t-1}^2}\right]. \qquad \text{(VI.5)}\end{aligned}$$

Suppose that the DGP $\{x_t\}_{t\in\mathbb{Z}}$ has $\theta_0 = (\omega_0, \alpha_0)'$ with $\alpha_0 < 1$. Then from Examples V.4.2 and V.4.4, the DGP is stationary and ergodic with $\mathbb{E}[x_t^2] < \infty$. For any fixed $\theta \in \Theta$,

$$\mathbb{E}[|q_t(\theta)|] = \frac{1}{2}\mathbb{E}\left[\left|\log(2\pi) + \log(\omega + \alpha x_{t-1}^2) + \frac{x_t^2}{\omega + \alpha x_{t-1}^2}\right|\right].$$

Using that $\omega > 0$, $\alpha x_{t-1}^2 \geq 0$ and $\log(x) \leq x - 1$ for $x > 0$, we have that

$$\log(\omega) \leq \log\left(\omega + \alpha x_{t-1}^2\right) \leq \omega + \alpha x_{t-1}^2,$$

such that

$$|\log\left(\omega + \alpha x_{t-1}^2\right)| \leq |\log(\omega)| + \left(\omega + \alpha x_{t-1}^2\right).$$

Hence,

$$\begin{aligned}\mathbb{E}\left[|\log\left(\omega + \alpha x_{t-1}^2\right)|\right] &\leq |\log(\omega)| + \mathbb{E}[\omega + \alpha x_{t-1}^2] \\ &= |\log(\omega)| + \omega + \alpha\mathbb{E}[x_{t-1}^2] < \infty.\end{aligned}$$

Likewise, using that $\alpha x_{t-1}^2 \geq 0$

$$0 \leq \mathbb{E}\left[\frac{x_t^2}{\omega + \alpha x_{t-1}^2}\right] \leq \mathbb{E}\left[\frac{x_t^2}{\omega}\right] = \omega^{-1}\mathbb{E}[x_t^2] < \infty.$$

We then have, using the triangle inequality,

$$\mathbb{E}[|q_t(\theta)|] \leq \frac{1}{2}\left[\log(2\pi) + \mathbb{E}\left[|\log\left(\omega + \alpha x_{t-1}^2\right)|\right] + \mathbb{E}\left[\frac{x_t^2}{\omega + \alpha x_{t-1}^2}\right]\right] < \infty.$$

Using Theorem V.2.1, noting for any fixed θ, $q_t(\theta)$ *is a function of the stationary and ergodic* $\{x_t\}_{t\in\mathbb{Z}}$ *with* $\mathbb{E}[|q_t(\theta)|] < \infty$, *we have that* $\{q_t(\theta)\}_{t\in\mathbb{Z}}$ *itself is stationary and ergodic. Moreover, Theorem V.2.2 implies that*

$$Q_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} q_t(\theta) \xrightarrow{p} \mathbb{E}\left[q_t(\theta)\right] =: Q(\theta).$$

Lastly, we seek to show that θ_0 *is the unique maximizer of* $Q(\theta)$. *To do so, note that*

$$\mathbb{E}\left[\frac{x_t^2}{\omega_0 + \alpha_0 x_{t-1}^2}\right] = \mathbb{E}\left[\frac{z_t^2(\omega_0 + \alpha_0 x_{t-1}^2)}{\omega_0 + \alpha_0 x_{t-1}^2}\right] = \mathbb{E}[z_t^2] = 1,$$

and

$$\mathbb{E}\left[\frac{x_t^2}{\omega + \alpha x_{t-1}^2}\right] = \mathbb{E}\left[\frac{z_t^2(\omega_0 + \alpha_0 x_{t-1}^2)}{\omega + \alpha x_{t-1}^2}\right] = \mathbb{E}\left[z_t^2\right]\mathbb{E}\left[\frac{\omega_0 + \alpha_0 x_{t-1}^2}{\omega + \alpha x_{t-1}^2}\right] = \mathbb{E}\left[\frac{\omega_0 + \alpha_0 x_{t-1}^2}{\omega + \alpha x_{t-1}^2}\right].$$

Then

$$\begin{aligned}
Q(\theta_0) - Q(\theta) &= \mathbb{E}\left[q_t(\theta_0)\right] - \mathbb{E}\left[q_t(\theta)\right] \\
&= -\frac{1}{2}\mathbb{E}\left[\log\left(\omega_0 + \alpha_0 x_{t-1}^2\right) + 1 - \log\left(\omega + \alpha x_{t-1}^2\right) - \frac{\omega_0 + \alpha_0 x_{t-1}^2}{\omega + \alpha x_{t-1}^2}\right] \\
&= \frac{1}{2}\mathbb{E}\left[\log\left(\frac{\omega + \alpha x_{t-1}^2}{\omega_0 + \alpha_0 x_{t-1}^2}\right) + \frac{\omega_0 + \alpha_0 x_{t-1}^2}{\omega + \alpha x_{t-1}^2} - 1\right].
\end{aligned}$$

Note again that for $x > 0$, $\log(x) \leq x - 1$, *or equivalently,* $0 \leq \log(x^{-1}) + x - 1$, *with equality if and only if* $x = 1$. *We then have that* $Q(\theta_0) - Q(\theta) \geq 0$ *with equality if and only if* $\omega_0 + \alpha_0 x_{t-1}^2 = \omega + \alpha x_{t-1}^2$ *almost surely, or equivalently,* $(\alpha_0 - \alpha)x_{t-1}^2 = \omega - \omega_0$ *almost surely. This can only hold if* $\theta = \theta_0$ *or if* x_{t-1}^2 *has a degenerate distribution. The latter is ruled out by the assumption that* z_{t-1} *is Gaussian. Consequently,* $Q(\theta_0) - Q(\theta) \geq 0$ *with equality if and only if* $\theta = \theta_0$. *We conclude that* θ_0 *is the unique maximizer of* $Q(\theta)$.

Given that $\hat{\theta}_T$ maximizes $Q_T(\theta)$, that $Q_T(\theta)$ has limit $Q(\theta)$, and that θ_0 maximizes $Q(\theta)$, it seems natural that $\hat{\theta}_T$ converges to θ_0 as the sample size diverges. This is indeed correct under additional technical conditions, stated in the following theorem:

Theorem VI.2.1 (NM, Theorem 2.7) *Suppose that*

1. *Θ is a convex set,*
2. *there exists a function $Q(\theta)$ that has unique maximum at θ_0 on Θ,*
3. *θ_0 is an interior point of Θ,*
4. *$Q_T(\theta)$ is concave, and*
5. *$Q_T(\theta) \xrightarrow{p} Q(\theta)$ for all $\theta \in \Theta$ as $T \to \infty$.*

Then $\hat{\theta}_T$, given by (VI.1), exists with probability approaching one, and $\hat{\theta}_T \xrightarrow{p} \theta_0$ as $T \to \infty$.

Proof. See NM (pp. 2133–2134). ■

Example VI.2.3 (AR(1) ctd.) *Clearly $\Theta = \mathbb{R}$ is convex. Moreover, $Q(\theta)$ is uniquely maximized at any θ_0 with $|\theta_0| < 1$ which is an interior point of Θ. We have that*

$$\frac{\partial^2 Q_T(\theta)}{\partial\theta\partial\theta} = -\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2 \leq 0, \tag{VI.6}$$

such that $Q_T(\theta)$ is concave. Lastly, as showed earlier, it holds that $Q_T(\theta) \xrightarrow{p} Q(\theta)$ for all $\theta \in \Theta$. By Theorem VI.2.1, we conclude that $\hat{\theta}_T \xrightarrow{p} \theta_0$. Note that this conclusion was reached in a much more direct way in Part II. In particular, recall that (VI.1) has solution (VI.3). Noting that $x_t = \theta_0 x_{t-1} + \varepsilon_t$, we have that

$$\hat{\theta}_T = \theta_0 + \frac{T^{-1}\sum_{t=1}^{T} x_{t-1}\varepsilon_t}{T^{-1}\sum_{t=1}^{T} x_{t-1}^2}.$$

Applying Theorem V.2.2 to both the numerator and denominator, as $T \to \infty$,

$$\left(T^{-1}\sum_{t=1}^{T} x_{t-1}\varepsilon_t, T^{-1}\sum_{t=1}^{T} x_{t-1}^2\right) \xrightarrow{p} \left(\mathbb{E}[x_{t-1}\varepsilon_t], \mathbb{E}[x_t^2]\right) = \left(0, \frac{\sigma_0^2}{1-\theta_0^2}\right).$$

Consequently, $\hat{\theta}_T \xrightarrow{p} \theta_0 + 0 = \theta_0$.

When considering non-linear models where the parameters might be constrained, one or more of the conditions in Theorem VI.2.1 may be violated. For instance, the objective function may not be concave, or θ_0 may not be an interior point. Instead, it is customary to assume that the parameter space Θ is compact, which under continuity of the objective function, automatically ensures existence of a maximizer. In this case, we have the following result.

Theorem VI.2.2 (NM, Theorem 2.1) *Suppose that*

1. *Θ is compact,*
2. *there exists a function $Q(\theta)$ that has unique maximum at θ_0 on Θ,*
3. *$Q(\theta)$ is continuous in θ, and*
4. *$\sup_{\theta\in\Theta} |Q_T(\theta) - Q(\theta)| \xrightarrow{p} 0$ as $T \to \infty$.*

Then with $\hat{\theta}_T$ given by (VI.1), $\hat{\theta}_T \xrightarrow{p} \theta_0$ as $T \to \infty$.

The first three conditions in Theorem VI.2.2 are typically straightforward to verify. The fourth condition states that the objective function $Q_T(\theta)$ converges uniformly in probability to $Q(\theta)$. Essentially, this condition ensures that the maximizer of $Q_T(\theta)$ should lie close to the the maximizer of $Q(\theta)$ as the sample size increases, as demonstrated in equation (VI.22) in the proof provided in the Appendix. The uniform convergence of the objective function can typically be shown to hold by applying the Uniform Law of Large Numbers (ULLN) for stationary and ergodic processes stated in Theorem A.1 in the Appendix, or – in the case of geometrically processes – Lemma A.2

Example VI.2.4 (AR(1) ctd.) *Consider the compact parameter space $\Theta = [-1, 1]$ and consider the stationary and ergodic setting with the true value $|\theta_0| < 1$. From earlier, we have that*

$$\begin{aligned}
Q(\theta) &= \mathbb{E}[q_t(\theta)] \\
&= \mathbb{E}\left[-\frac{1}{2}(x_t - \theta x_{t-1})^2\right] \\
&= -\frac{1}{2}\left(\frac{\sigma_0^2}{1-\theta_0^2}\right)\left(1 + \theta^2 - 2\theta_0\theta\right),
\end{aligned}$$

which is continuous in θ and uniquely maximized at $\theta = \theta_0$. Clearly, $q_t(\theta)$ is measurable and continuous in θ. Furthermore, for any $\theta \in \Theta = [-1, 1]$,

$$\begin{aligned}
|q_t(\theta)| &\leq \frac{1}{2}|x_t^2 + \theta^2 x_{t-1}^2 - 2\theta x_t x_{t-1}| \\
&\leq \frac{1}{2}\left(x_t^2 + x_{t-1}^2 + 2|x_t x_{t-1}|\right),
\end{aligned}$$

such that

$$\mathbb{E}[\sup_{\theta\in\Theta}|q_t(\theta)|] \leq \frac{1}{2}\mathbb{E}\left[x_t^2 + x_{t-1}^2 + 2|x_t x_{t-1}|\right] < \infty,$$

as $\mathbb{E}[x_t^2] < \infty$*. By Theorem A.1,* $\sup_{\theta\in\Theta}|Q_T(\theta) - Q(\theta)| \xrightarrow{p} 0$ *as* $T \to \infty$*, and consequently, by Theorem VI.2.2, we have that* $\hat{\theta}_T \xrightarrow{p} \theta_0$ *as* $T \to \infty$*.*

Example VI.2.5 (ARCH(1) ctd.) *We seek to find assumptions such that Conditions 1-4 of Theorem VI.2.2 hold. We initially elaborate on the parameter space* Θ *for the estimation of the parameters,* $\theta = (\omega, \alpha) \in (0,\infty)\times[0,\infty)$*. Note that Condition 4 requires that the log-likelihood function converges uniformly in probability on* Θ*, and we seek to show this by applying Theorem A.1. To do so we need that* $\mathbb{E}[\sup_{\theta\in\Theta}|q_t(\theta)|] < \infty$*, with* $q_t(\theta)$ *the log-likelihood contribution given in (VI.5). Note initially that for any* $x, y \in \mathbb{R}$*,*

$$\sup_{\theta\in(0,\infty)\times[0,\infty)}\left|\log(2\pi) + \log(\omega + \alpha x^2) + \frac{x^2}{\omega + \alpha y^2}\right| = \infty.$$

In particular, this suggests that $\omega > 0$ *should be bounded away from zero and bounded from above, whereas* $\alpha \geq 0$ *should be bounded from above. Consequently, we consider the compact parameter space (Condition 1)*

$$\Theta = [\omega_L, \omega_U] \times [0, \alpha_U], \tag{VI.7}$$

with $0 < \omega_L < \omega_U < \infty$ *and* $0 < \alpha_U < \infty$*. Then for any* $x, y \in \mathbb{R}$*, using the same type of bounds as in Example VI.2.2,*

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|\log(2\pi) + \log(\omega + \alpha y^2) + \frac{x^2}{\omega + \alpha y^2}\right| \\
&\leq \sup_{\theta\in[\omega_L,\omega_U]\times[0,\alpha_U]}\left(\log(2\pi) + |\log(\omega)| + \left(\omega + \alpha y^2\right) + \frac{x^2}{\omega + \alpha y^2}\right) \\
&\leq \log(2\pi) + |\log(\omega_L)| + |\log(\omega_U)| + \left(\omega_U + \alpha_U y^2\right) + \frac{x^2}{\omega_L}.
\end{aligned}$$

Maintaining the assumption that $\alpha_0 < 1$ *such that* $\{x_t\}_{t\in\mathbb{Z}}$ *is stationary and ergodic with* $\mathbb{E}[x_t^2] < \infty$*, we have that*

$$\mathbb{E}[\sup_{\theta\in\Theta}|q_t(\theta)|] \leq \log(2\pi) + |\log(\omega_L)| + |\log(\omega_U)| + \omega_U + \alpha_U\mathbb{E}(x_{t-1}^2) + \frac{\mathbb{E}[x_t^2]}{\omega_L} < \infty.$$

Clearly, $q_t(\theta)$ *is finite almost surely for any* $\theta \in \Theta$ *and continuous in* θ*. Consequently, by Theorem A.1,*

$$\sup_{\theta\in\Theta}|Q_T(\theta) - \mathbb{E}\left[q_t(\theta)\right]| \xrightarrow{p} 0 \quad \textit{as } T \to \infty,$$

such that Condition 4 holds. Moreover, by dominated convergence $Q(\theta) = \mathbb{E}[q_t(\theta)]$ is continuous in θ (Condition 3). Lastly, assuming that $\theta_0 \in \Theta$, from Example VI.2.2, it holds that $Q(\theta)$ is uniquely maximized at θ_0 (Condition 2). We conclude that all Conditions 1-4 of Theorem VI.2.2 hold with Θ given by (VI.7) provided that $\theta_0 \in \Theta$ and $\alpha_0 < 1$. Under these conditions, with $\hat{\theta}_T$ given by (VI.1), $\hat{\theta}_T \xrightarrow{p} \theta_0$ as $T \to \infty$.

Remark VI.2.1 *Example VI.2.5 showed that for a compact parameter space, the MLE $\hat{\theta}_T$ converges to θ_0 provided that the DGP is stationary and ergodic with $\mathbb{E}[x_t^2] < \infty$. The assumption about finite second moment can be relaxed such that $\mathbb{E}[|x_t|^\delta] < \infty$ for some $\delta > 0$; see e.g. Francq and Zakoïan (2019, Chapter 7). One can also relax the assumption about the compactness of the parameter space, as done in Kristensen and Rahbek (2005), who assume that the DGP $\{x_t\}_{t=0,1,...}$ is geometrically ergodic (but not necessarily stationary). Lastly, it is possible to show that the MLE for α is consistent, even if $\mathbb{E}[\log(\alpha_0 z_t^2)] > 0$, such that no stationary solution to the ARCH(1) process exists (Example V.4.2). This relies on rather technical arguments, and we refer to Jensen and Rahbek (2004a) and Francq and Zakoïan (2012) for details.*

VI.3 Asymptotic normality

In this section, we present conditions such that the estimator $\hat{\theta}_T$ is asymptotically normal. Let $\|\cdot\|$ denote any matrix norm; cf. the matrix results provided in the appendix to Part I.

We have the following result.

Theorem VI.3.1 *Suppose that with $\hat{\theta}_T$ given by (VI.1),*

1. *$\hat{\theta}_T \xrightarrow{p} \theta_0$ as $T \to \infty$,*

2. *the true value θ_0 is an interior point of the parameter space Θ,*

3. *$Q_T(\theta_0)$ is twice continuously differentiable in a neighborhood $\mathcal{N}(\theta_0)$ of θ_0 almost surely,*

4. *the score,*

$$\sqrt{T}\frac{\partial Q_T(\theta_0)}{\partial \theta} := \sqrt{T}\left.\frac{\partial Q_T(\theta)}{\partial \theta}\right|_{\theta=\theta_0} \xrightarrow{D} N(0, \Omega_0) \quad \text{as } T \to \infty,$$

with Ω_0 a positive definite $(k \times k)$ matrix, and

5. *there exists a matrix-valued function* $\Sigma(\theta)$ *that is continuous at* θ_0*, such that*

$$\sup_{\theta\in\mathcal{N}(\theta_0)}\left\|\frac{\partial^2 Q_T(\theta)}{\partial\theta\partial\theta'}-\Sigma(\theta)\right\|\xrightarrow{p}0\quad \text{as } T\to\infty,$$

and $\Sigma_0 := \Sigma(\theta_0)$ *invertible.*

Then as $T\to\infty$*,*

$$\sqrt{T}\left(\hat\theta_T-\theta_0\right)\xrightarrow{D}N(0,\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}).$$

Proof. Note that Condition 1 implies that $\hat\theta_T\in\mathcal{N}(\theta_0)$ with probability approaching one. This combined with Conditions 2 and 3 yield that, with probability approaching one,

$$\frac{\partial Q_T(\hat\theta_T)}{\partial\theta}=0_{k\times1}.$$

Mean value expansions (element-by-element) of $\partial Q_T(\hat\theta_T)/\partial\theta$ at θ_0 yield (with probability approaching one)

$$0_{k\times1}=\frac{\partial Q_T(\theta_0)}{\partial\theta}+\frac{\partial^2 Q_T(\theta_T^*)}{\partial\theta\partial\theta'}(\hat\theta_T-\theta_0), \tag{VI.8}$$

where the mean value θ_T^* lies between[1] $\hat\theta_T$ and θ_0. Note that Condition 1 implies that $\theta_T^*\xrightarrow{p}\theta_0$ as $T\to\infty$. This combined with Condition 5 and Lemma A.1 yield that

$$\frac{\partial^2 Q_T(\theta_T^*)}{\partial\theta\partial\theta'}=\Sigma_0+o_p(1). \tag{VI.9}$$

Combining (VI.8) and (VI.9) yield that

$$0_{k\times1}=\sqrt{T}\frac{\partial Q_T(\theta_0)}{\partial\theta}+(\Sigma_0+o_p(1))\sqrt{T}(\hat\theta_T-\theta_0).$$

Since Σ_0 is invertible, re-arranging yields

$$\sqrt{T}(\hat\theta_T-\theta_0)=-\left(\Sigma_0+o_p(1)\right)^{-1}\sqrt{T}\frac{\partial Q_T(\theta_0)}{\partial\theta}\xrightarrow{D}N(0,\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}),$$

where the convergence follows by Condition 4 and Slutsky's lemma. ■

[1]Note that θ_T^* may vary across the rows of $\partial^2 Q_T(\theta_T^*)/\partial\theta\partial\theta'$; see, e.g., Jensen and Rahbek (2004b) for the precise details.

Example VI.3.1 (AR(1) ctd.) *Consider Example VI.2.4 with the compact parameter space $\Theta = [-1, 1]$, and recall that $|\theta_0| < 1$. We seek to verify Conditions 1–5 of Theorem VI.3.1. The estimator is consistent as shown in Example VI.2.4 (Condition 1). Since $|\theta_0| < 1$, θ_0 is an interior point of the parameter space (Condition 2). From VI.6,*

$$\frac{\partial^2 Q_T(\theta)}{\partial \theta^2} = -\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2,$$

which is clearly continuous in θ (Condition 3), and clearly by Theorem V.2.2,

$$\frac{\partial^2 Q_T(\theta)}{\partial \theta^2} \xrightarrow{p} -\mathbb{E}[x_t^2] =: \Sigma_0 < 0$$

as $T \to \infty$ uniformly in θ (Condition 5). It remains to show Condition 4. We have that

$$\begin{aligned}
\frac{\partial Q_T(\theta)}{\partial \theta} &= \frac{1}{T}\sum_{t=1}^{T} \frac{\partial q_t(\theta)}{\partial \theta} \\
&= \frac{1}{T}\sum_{t=1}^{T} \frac{\partial}{\partial \theta}\left(-\frac{1}{2}(x_t - \theta x_{t-1})^2\right) \\
&= \frac{1}{T}\sum_{t=1}^{T} (x_t - \theta x_{t-1}) x_{t-1},
\end{aligned}$$

such that

$$\sqrt{T}\frac{\partial Q_T(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \varepsilon_t x_{t-1}.$$

Clearly $\varepsilon_t x_{t-1}$ forms a martingale difference with $\mathbb{E}[\varepsilon_t^2 x_{t-1}^2] = \mathbb{E}[\varepsilon_t^2]\mathbb{E}[x_{t-1}^2] < \infty$. By Theorem V.2.3 we have that

$$\sqrt{T}\frac{\partial Q_T(\theta_0)}{\partial \theta} \xrightarrow{D} N(0, \Omega_0),$$

with $\Omega_0 = \mathbb{E}[\varepsilon_t^2]\mathbb{E}[x_{t-1}^2]$. By Theorem VI.3.1 we have that

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{D} N(0, \Sigma_0^{-1}\Omega_0\Sigma_0^{-1}).$$

Noting that $\Sigma_0^{-1}\Omega_0\Sigma_0^{-1} = \mathbb{E}[\varepsilon_t^2]/\mathbb{E}[x_t^2] = (1 - \theta_0^2)$

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{D} N(0, (1 - \theta_0^2)),$$

identical to the conclusion for the ML estimator in (II.6).

VI.3.1 Asymptotic normality of the ML estimator for the ARCH(1) model

As in Example VI.2.5, consider the parameter space Θ in (VI.7), and assume throughout that the true $\alpha_0 < 1$ such that the DGP is stationary and ergodic with $\mathbb{E}[x_t^2] < \infty$. We seek to show that Conditions 1–5 of Theorem VI.3.1 hold.

Condition 1:

From Example VI.2.5 we have that $\hat{\theta}_T$ is consistent, that is, Condition 1 holds.

Condition 2:

We assume that θ_0 is an interior point of Θ, so that Condition 2 holds.

Condition 3:

Recall that $Q_T(\theta) = T^{-1}\sum_{t=1}^{T} q_t(\theta)$ with $q_t(\theta)$ given by (VI.5). Clearly $q_t(\theta)$, and hence $Q_T(\theta)$, is twice continuously differentiable (almost surely), and we have that Condition 3 holds.

Condition 4:

We have that

$$\begin{aligned}\frac{\partial Q_T(\theta)}{\partial\theta} &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial q_t(\theta)}{\partial\theta}, \\ &= -\frac{1}{2T}\sum_{t=1}^{T}\frac{\partial}{\partial\theta}\left[\log(2\pi) + \log(\omega + \alpha x_{t-1}^2) + \frac{x_t^2}{\omega + \alpha x_{t-1}^2}\right]\end{aligned}$$

with

$$\begin{aligned}\frac{\partial q_t(\theta)}{\partial\theta} &= -\frac{1}{2}\frac{\partial}{\partial\theta}\left[\log(2\pi) + \log(\omega + \alpha x_{t-1}^2) + \frac{x_t^2}{\omega + \alpha x_{t-1}^2}\right] \\ &= -\frac{1}{2}\frac{1}{\omega + \alpha x_{t-1}^2}\left(1 - \frac{x_t^2}{\omega + \alpha x_{t-1}^2}\right) w_t.\end{aligned}$$

with

$$w_t := \begin{bmatrix} 1 \\ x_{t-1}^2 \end{bmatrix}.$$

Use that $x_t^2 = (\omega_0 + \alpha_0 x_{t-1}^2) z_t^2$, such that

$$\frac{\partial q_t(\theta_0)}{\partial \theta} = \frac{1}{2} \frac{1}{\omega_0 + \alpha_0 x_{t-1}^2} \left(z_t^2 - 1 \right) w_t.$$

With $\mathcal{F}_t = \sigma(x_t, x_{t-1}, \dots)$, note that (provided that $\mathbb{E}[\| \partial q_t(\theta_0)/\partial\theta \|] < \infty$, which is shown below), almost surely

$$\mathbb{E}\left[\frac{\partial q_t(\theta_0)}{\partial \theta} | \mathcal{F}_{t-1} \right] = 0_{2\times 1}.$$

Hence $\partial q_t(\theta_0)/\partial\theta$ is a martingale difference, and we seek to prove Condition 4 by applying the CLT in Theorem V.2.3. To do so, we show that the CLT applies to any linear combination of $\partial q_t(\theta_0)/\partial\theta$. For any fixed non-zero $\lambda := (\lambda_1, \lambda_2)' \in \mathbb{R}^2$, let

$$\begin{aligned} Y_t^{(\lambda)} &:= \lambda' \frac{\partial q_t(\theta_0)}{\partial \theta} \\ &= \frac{1}{2} \left(z_t^2 - 1 \right) \frac{\lambda' w_t}{\omega_0 + \alpha_0 x_{t-1}^2} \\ &= \frac{1}{2} \left(z_t^2 - 1 \right) \frac{\lambda_1 + \lambda_2 x_{t-1}^2}{\omega_0 + \alpha_0 x_{t-1}^2}. \end{aligned} \tag{VI.10}$$

Clearly, $\mathbb{E}[Y_t^{(\lambda)} | \mathcal{F}_{t-1}] = 0$ almost surely. Moreover, by independence,

$$\begin{aligned} \mathbb{E}\left[\left(Y_t^{(\lambda)} \right)^2 \right] &= \mathbb{E}\left[\left(\frac{1}{2} \left(z_t^2 - 1 \right) \frac{\lambda_1 + \lambda_2 x_{t-1}^2}{\omega_0 + \alpha_0 x_{t-1}^2} \right)^2 \right] \\ &= \frac{1}{4} \mathbb{E}\left[\left(z_t^2 - 1 \right)^2 \right] \mathbb{E}\left[\left(\frac{\lambda_1 + \lambda_2 x_{t-1}^2}{\omega_0 + \alpha_0 x_{t-1}^2} \right)^2 \right], \\ &= \frac{1}{2} \mathbb{E}\left[\left(\frac{\lambda_1 + \lambda_2 x_{t-1}^2}{\omega_0 + \alpha_0 x_{t-1}^2} \right)^2 \right], \end{aligned}$$

where we have used that $\mathbb{E}\left[\left(z_t^2 - 1 \right)^2 \right] = \mathbb{E}[z_t^4 + 1 - 2z_t^2] = 3 + 1 - 2 = 2$ since $z_t \overset{D}{=} N(0,1)$. Since θ_0 is an interior point of Θ_0, we have that both $\omega_0, \alpha_0 > 0$, so that

$$\left| \frac{\lambda_1}{\omega_0 + \alpha_0 x_{t-1}^2} \right| = \frac{|\lambda_1|}{\omega_0 + \alpha_0 x_{t-1}^2} \leq \frac{|\lambda_1|}{\omega_0} < \infty,$$

and

$$\left| \frac{\lambda_2 x_{t-1}^2}{\omega_0 + \alpha_0 x_{t-1}^2} \right| = \frac{|\lambda_2| x_{t-1}^2}{\omega_0 + \alpha_0 x_{t-1}^2} \leq \frac{|\lambda_2|}{\alpha_0} < \infty.$$

These bounds yield that

$$\mathbb{E}\left[\left(\frac{\lambda_1+\lambda_2 x_{t-1}^2}{\omega_0+\alpha_0 x_{t-1}^2}\right)^2\right] \leq \left(\frac{|\lambda_1|}{\omega_0}\right)^2 + \left(\frac{|\lambda_2|}{\alpha_0}\right)^2 + 2\left(\frac{|\lambda_1|}{\omega_0}\right)\left(\frac{|\lambda_2|}{\alpha_0}\right) < \infty,$$

and we have that

$$\mathbb{E}\left[\left(Y_t^{(\lambda)}\right)^2\right] < \infty.$$

By Theorem V.2.3, we then have that for any non-zero λ,

$$\frac{1}{\sqrt{T}}\lambda'\frac{\partial Q_T(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\lambda'\frac{\partial q_t(\theta_0)}{\partial\theta} \xrightarrow{D} N(0,\gamma_\lambda),$$

with $\gamma_\lambda = \mathbb{E}[(Y_t^{(\lambda)})^2] < \infty$ provided that $\gamma_\lambda > 0$. Note that

$$\begin{aligned}
\gamma_\lambda &= \mathbb{E}\left[\left(\lambda'\frac{\partial q_t(\theta_0)}{\partial\theta}\right)^2\right] \\
&= \lambda'\mathbb{E}\left[\frac{\partial q_t(\theta_0)}{\partial\theta}\frac{\partial q_t(\theta_0)}{\partial\theta'}\right]\lambda \\
&= \lambda'\Omega_0\lambda,
\end{aligned}$$

with

$$\begin{aligned}
\Omega_0 &:= \mathbb{E}\left[\frac{\partial q_t(\theta_0)}{\partial\theta}\frac{\partial q_t(\theta_0)}{\partial\theta'}\right] \\
&= \mathbb{E}\left[\frac{1}{4}\frac{1}{(\omega_0+\alpha_0 x_{t-1}^2)^2}\left(z_t^2-1\right)^2 w_t w_t'\right] \\
&= \frac{1}{2}\mathbb{E}\left[\frac{1}{(\omega_0+\alpha_0 x_{t-1}^2)^2} w_t w_t'\right].
\end{aligned}$$

Clearly, $\gamma_\lambda > 0$ for all non-zero λ, if and only if Ω_0 is positive definite. We prove the latter by contradiction. Note that $\lambda'\Omega_0\lambda \geq 0$ for all λ. Suppose that Ω_0 is not positive definite. Then there exists a non-zero λ such that

$$\lambda'\Omega_0\lambda = \mathbb{E}\left[\left(\frac{\lambda' w_t}{\omega_0+\alpha_0 x_{t-1}^2}\right)^2\right] = 0.$$

Since $\left(\lambda' w_t/(\omega_0+\alpha_0 x_{t-1}^2)\right)^2$ is non-negative and $\omega_0+\alpha_0 x_{t-1}^2 > 0$ it holds that $\lambda' w_t = 0$ almost surely, that is, $\mathbb{P}\left(\lambda_1+\lambda_2 x_{t-1}^2 = 0\right) = 1$. Since z_{t-1} is Gaussian, we rule out $\mathbb{P}\left(\lambda_1+\lambda_2 x_{t-1}^2 = 0\right) = 1$ unless $(\lambda_1,\lambda_2) = (0,0)$, that is, λ is the zero vector. We conclude that Ω_0 is positive definite, and we have that

$$\frac{1}{\sqrt{T}}\lambda'\frac{\partial Q_T(\theta_0)}{\partial\theta} \xrightarrow{D} N(0, \lambda'\Omega_0\lambda),$$

for all non-zero λ. The Cramér-Wold theorem yields that

$$\frac{1}{\sqrt{T}}\frac{\partial Q_T(\theta_0)}{\partial\theta} \xrightarrow{D} N(0, \Omega_0),$$

and we conclude that Condition 4 holds.

Condition 5:

We set up for an application of the ULLN in Theorem A.1 to each entry of $\partial^2 Q_T(\theta)/\partial\theta\partial\theta'$. Since θ_0 is an interior point of Θ, $\alpha_0 > 0$, and we have that there exists a α_L satisfying $0 < \alpha_L < \alpha_0 < \alpha_U$, such that $\theta_0 \in [\omega_L, \omega_U] \times [\alpha_L, \alpha_U] =: \mathcal{N}(\theta_0)$. We have that

$$\frac{\partial^2 Q_T(\theta)}{\partial\theta\partial\theta'} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 q_t(\theta)}{\partial\theta\partial\theta'},$$

with

$$\begin{aligned}
\frac{\partial^2 q_t(\theta)}{\partial\theta\partial\theta'} &= \begin{bmatrix} \frac{\partial^2}{\partial\omega^2}q_t(\theta) & \frac{\partial^2}{\partial\omega\partial\alpha}q_t(\theta) \\ \frac{\partial^2}{\partial\omega\partial\alpha}q_t(\theta) & \frac{\partial^2}{\partial\alpha^2}q_t(\theta) \end{bmatrix} \\
&= \frac{1}{2}\frac{1}{\sigma_t^4(\theta)}\left(1 - \frac{2x_t^2}{\sigma_t^2(\theta)}\right) w_t w_t'.
\end{aligned}$$

It suffices to show that

$$\mathbb{E}\left[\sup_{\theta\in\mathcal{N}(\theta_0)}\left\|\frac{\partial^2 q_t(\theta)}{\partial\theta\partial\theta'}\right\|\right] < \infty. \tag{VI.11}$$

Note that for some constant $C > 0$,

$$
\begin{aligned}
\left\| \frac{\partial^2 q_t(\theta)}{\partial\theta\partial\theta'} \right\| &= \left\| \frac{1}{2} \frac{1}{\sigma_t^4(\theta)} \left(1 - \frac{2x_t^2}{\sigma_t^2(\theta)} \right) w_t w_t' \right\| \\
&\leq \left(\frac{1}{2} \frac{1}{\sigma_t^4(\theta)} \left(1 + \frac{2x_t^2}{\sigma_t^2(\theta)} \right) \right) \| w_t w_t' \| \\
&\leq C \left(\frac{1}{2} \frac{1}{\sigma_t^4(\theta)} \left(1 + \frac{2x_t^2}{\sigma_t^2(\theta)} \right) \right) w_t' w_t \\
&= \frac{C}{2} \left(\frac{1}{\sigma_t^4(\theta)} \left(1 + \frac{2x_t^2}{\sigma_t^2(\theta)} \right) \right) (1 + x_{t-1}^4) \\
&= \frac{C}{2} \frac{(1 + x_{t-1}^4)}{\sigma_t^4(\theta)} + \frac{C}{2} \frac{(1 + x_{t-1}^4) 2x_t^2}{\sigma_t^6(\theta)} \\
&= \frac{C}{2} \frac{(1 + x_{t-1}^4)}{\sigma_t^4(\theta)} + C \frac{z_t^2 (1 + x_{t-1}^4)(\omega_0 + \alpha_0 x_{t-1}^2)}{\sigma_t^6(\theta)},
\end{aligned}
$$

so that (VI.11) holds provided that

$$
\mathbb{E}\left[\sup_{\theta \in \mathcal{N}(\theta_0)} \left| \frac{C}{2} \frac{(1 + x_{t-1}^4)}{\sigma_t^4(\theta)} \right| \right] + \mathbb{E}\left[\sup_{\theta \in \mathcal{N}(\theta_0)} \left| C \frac{z_t^2 (1 + x_{t-1}^4)(\omega_0 + \alpha_0 x_{t-1}^2)}{\sigma_t^6(\theta)} \right| \right] < \infty.
$$

We restrict our attention to the second term, and note that the finiteness of the first term follows by similar arguments. We have (ignoring the constant $C > 0$)

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\theta \in \mathcal{N}(\theta_0)} \left| \frac{z_t^2 (1 + x_{t-1}^4)(\omega_0 + \alpha_0 x_{t-1}^2)}{\sigma_t^6(\theta)} \right| \right] \\
&= \mathbb{E}\left[\sup_{\theta \in \mathcal{N}(\theta_0)} \frac{(1 + x_{t-1}^4)(\omega_0 + \alpha_0 x_{t-1}^2)}{\sigma_t^6(\theta)} \right] \qquad \text{(independence)} \\
&\leq \mathbb{E}\left[\frac{(1 + x_{t-1}^4)(\omega_0 + \alpha_0 x_{t-1}^2)}{(\omega_L + \alpha_L x_{t-1}^2)^3} \right] \\
&= \mathbb{E}\left[\frac{(\omega_0 + \alpha_0 x_{t-1}^2)}{(\omega_L + \alpha_L x_{t-1}^2)^3} + \frac{x_{t-1}^4 (\omega_0 + \alpha_0 x_{t-1}^2)}{(\omega_L + \alpha_L x_{t-1}^2)^3} \right] \\
&\leq \frac{\omega_0}{\omega_L^3} + \frac{\alpha_0}{\omega_L^2 \alpha_L} + \frac{\omega_0}{\omega_L \alpha_L^2} + \frac{\alpha_0}{\alpha_L^3} < \infty.
\end{aligned}
$$

We conclude that (VI.11) holds such that, by Theorem A.1,

$$
\sup_{\theta \in \mathcal{N}(\theta_0)} \left\| T^{-1} \sum_{t=1}^{T} \frac{\partial^2 q_t(\theta)}{\partial\theta\partial\theta'} - \Sigma(\theta) \right\| \xrightarrow{p} 0 \quad \text{as } T \to \infty,
$$

with

$$\Sigma(\theta) = \mathbb{E}\left[\frac{\partial^2 q_t(\theta)}{\partial\theta\partial\theta'}\right].$$

Clearly, $\Sigma(\theta)$ is continuous at θ_0, and it remains to show that $\Sigma_0 = \Sigma(\theta_0)$ is invertible. But this is immediate, noting that

$$\begin{aligned}
\Sigma_0 &= \mathbb{E}\left[\frac{1}{2}\frac{1}{\sigma_t^4(\theta_0)}\left(1 - \frac{2x_t^2}{\sigma_t^2(\theta_0)}\right) w_t w_t'\right] \\
&= \mathbb{E}\left[\frac{1}{2}\frac{1}{\sigma_t^4(\theta_0)}\left(1 - 2z_t^2\right) w_t w_t'\right] \\
&= -\frac{1}{2}\mathbb{E}\left[\frac{1}{\sigma_t^4(\theta_0)} w_t w_t'\right] \\
&= -\Omega_0.
\end{aligned}$$

We conclude that Condition 5 holds.

To sum up:

If $\alpha_0 \in (0,1)$, it holds that the ML estimator $\hat{\theta}_T$ satisfies

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{D} N(0, -\Sigma_0^{-1}),$$

using that $\Omega_0 = -\Sigma_0$. Note that the latter equality is the so-called information equality that holds for ($\sqrt{T}$-consistent) ML estimators.

VI.4 Hypothesis testing

Similar to Section II.4.4, we may consider testing hypotheses about the parameter vector θ. We restrict our attention to linear hypotheses of the form

$$H : R'\theta = r, \tag{VI.12}$$

with R a known $(k \times l)$ matrix with rank $0 < l \leq k$, and r a known l-dimensional vector.[2]

Example VI.4.1 (ARCH(1) ctd) *Following Example V.4.6 we may test the hypothesis*

$$H : \alpha = \frac{1}{\sqrt{3}}, \tag{VI.13}$$

such that the unconditional distribution of x_t has tail index $\kappa = 4$. In this case, recalling that $\theta = (\omega, \alpha)'$, we have that $R = (0,1)'$ and $r = 1/\sqrt{3}$.

[2] The following considerations may be extended to more general hypotheses of the form $F(\theta) = r$ for certain known functions F; see e.g. Section 9 of NM for additional details.

VI.4.1 Likelihood Ratio (LR) tests

Similar to Part II, we consider a likelihood ratio (LR) test for the hypothesis. Specifically, let $\tilde{\theta}_T$ denote the maximizer of $Q_T(\theta)$ over Θ subject to the constraint (VI.12). Then the LR statistic for the hypothesis in (VI.12) is given by

$$LR_T(H) := 2T[Q_T(\hat{\theta}_T) - Q_T(\tilde{\theta}_T)]. \tag{VI.14}$$

In the special case of the hypothesis $\theta = \theta_0$ (that is, $l = k$, $R = I_k$ and $r = \theta_0$), $\tilde{\theta}_T = \theta_0$. In this case, under the assumptions of Theorem VI.3.1, a second-order mean-value expansion of $Q_T(\theta_0)$ around $\hat{\theta}_T$ yields that

$$Q_T(\theta_0) = Q_T(\hat{\theta}_T) + \frac{\partial Q_T(\hat{\theta}_T)}{\partial \theta'}(\theta_0 - \hat{\theta}_T) + \frac{1}{2}(\hat{\theta}_T - \theta_0)'\hat{\Sigma}_T(\hat{\theta}_T - \theta_0) + o_p(T^{-1}),$$

with

$$\hat{\Sigma}_T := \frac{\partial^2 Q_T(\hat{\theta}_T)}{\partial \theta \partial \theta'}.$$

Hence, using the first-order condition $\partial Q_T(\hat{\theta}_T)/\partial \theta = 0_{k \times 1}$ (with probability approaching one), we have that the LR statistic satisfies

$$LR_T\,(\theta = \theta_0) = 2T[Q_T(\hat{\theta}_T) - Q_T(\theta_0)] = W + o_p(1), \tag{VI.15}$$

where $W := T(\hat{\theta}_T - \theta_0)'(-\hat{\Sigma}_T)(\hat{\theta}_T - \theta_0)$ corresponds to the Wald statistic derived in Part II. By Lemma VI.4.1 provided below, $\hat{\Sigma}_T \xrightarrow{p} \Sigma_0$ as $T \to \infty$. In the case where the information equality holds, that is,

$$\Omega_0 = -\Sigma_0, \tag{VI.16}$$

such that $\sqrt{T}(\hat{\theta} - \theta_0)$ has asymptotic covariance $-\Sigma_0^{-1}$, we then have that

$$LR_T\,(\theta = \theta_0) \xrightarrow{D} \chi^2_k \quad \text{as } T \to \infty.$$

Under a more general hypothesis of the form (VI.12), we have the following result, where the consistency of $\tilde{\theta}_T$ may be shown to hold under the assumptions of Theorem VI.2.2.

Theorem VI.4.1 *Suppose that the assumptions of Theorem VI.3.1 hold. Moreover, assume that the information equality (VI.16) holds, and that under the hypothesis H in (VI.12), $\tilde{\theta}_T \xrightarrow{p} \theta_0$ as $T \to \infty$. Then under the hypothesis H in (VI.12), the LR statistic in (VI.14) satisfies*

$$LR_T(H) \xrightarrow{D} \chi^2_l, \quad \text{as } T \to \infty.$$

Proof. See Appendix. ■

VI.4.2 Wald tests

If the information equality does not hold – which is for instance the case for so-called *quasi* ML estimators for ARCH models, as considered later – $LR_T(H)$ does not have a standard chi-squared distribution. In such a case, one may alternatively make use of a (robustified) Wald test, provided that one has a consistent estimator for the asymptotic covariance matrix of $\sqrt{T}(\hat{\theta}_T - \theta_0)$. Specifically, suppose that there exists a matrix $\hat{\Psi}_T$, such that

$$\hat{\Psi}_T \xrightarrow{p} \Sigma_0^{-1}\Omega_0\Sigma_0^{-1} \quad \text{as } T \to \infty. \tag{VI.17}$$

Then a Wald statistic for the hypothesis (VI.12) is given by

$$W_{\text{rob}} = T\left(R'\hat{\theta}_T - r\right)'\left(R'\hat{\Psi}_T R\right)^{-1}\left(R'\hat{\theta}_T - r\right). \tag{VI.18}$$

We have the following result.

Theorem VI.4.2 *Under the assumptions of Theorem VI.3.1, suppose that (VI.17) holds. Then under the hypothesis H in (VI.12), the Wald statistic W_{rob} in (VI.18) satisfies*

$$W_{\text{rob}} \xrightarrow{D} \chi^2(l).$$

Proof. Under the hypothesis H in (VI.12), we have that

$$R'\theta_0 = r,$$

such that by Theorem VI.3.1 and the continuous mapping theorem

$$R'\hat{\theta}_T - r = R'(\hat{\theta}_T - \theta_0) \xrightarrow{D} N(0, R'\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}R). \tag{VI.19}$$

The $l \times l$ matrix $R'\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}R$ is positive definite, since $\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}$ and R have full rank. Likewise, from (VI.17),

$$R'\hat{\Psi}_T R \xrightarrow{p} R'\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}R. \tag{VI.20}$$

Combining (VI.19) and (VI.20) together with the continuous mapping theorem yields the desired result. ■

Remark VI.4.1 *By similar arguments, if one is interested in testing a hypothesis about the jth entry of θ, that is,*

$$H : \theta_j = r,$$

then one may use that t-statistic

$$\tau_{\theta_j=r} = \frac{\hat{\theta}_{T,j} - r}{\sqrt{(\hat{\Psi}_T)_{jj}/T}},$$

with $\hat{\theta}_{T,j}$ *and* $(\hat{\Psi}_T)_{jj}$ *the* j *entry of* $\hat{\theta}_T$ *and the diagonal of* $\hat{\Psi}_T$*, respectively. Under the assumptions of Theorem VI.4.2,*

$$\tau_{\theta_j=r} \xrightarrow{D} N(0,1).$$

For instance in terms of the AR(1) model, testing the hypothesis $\theta = \theta_0$ *with* $|\theta_0| < 1$*, we may from Example VI.3.1 apply the t-statistic*

$$\tau_{\theta=\theta_0} = \frac{\hat{\theta}_T - \theta_0}{\sqrt{(1-\hat{\theta}_T^2)/T}} \xrightarrow{D} N(0,1).$$

Example VI.4.2 (ARCH(1) ctd.) *We seek to apply Theorem VI.4.2 for constructing a Wald test for the hypothesis in (VI.13) based on the ML estimator* $\hat{\theta}_T = (\hat{\omega}_T, \hat{\alpha}_T)'$*. Note that under the hypothesis,* $\alpha_0 < 1$*, and as carefully argued in Section VI.3.1, all the conditions of Theorem VI.3.1 hold under this condition. It remains to find a consistent estimator of the asymptotic covariance of the ML estimator. Recall from Section VI.3.1,*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{D} N(0, -\Sigma_0^{-1}),$$

with $\Sigma_0 = \Sigma(\theta_0)$ *and*

$$\Sigma(\theta) = \mathbb{E}\left[\frac{\partial^2 q_t(\theta)}{\partial\theta\partial\theta'}\right],$$

$$\frac{\partial^2 q_t(\theta)}{\partial\theta\partial\theta'} = \frac{1}{2}\frac{1}{\omega + \alpha x_{t-1}^2}\left(1 - \frac{2x_t^2}{\omega + \alpha x_{t-1}^2}\right) w_t w_t', \quad w_t = (1, x_{t-1}^2)'.$$

From Condition 5 of Theorem VI.3.1,

$$\sup_{\theta\in\mathcal{N}(\theta_0)} \left\| T^{-1}\sum_{t=1}^{T} \frac{\partial^2 q_t(\theta)}{\partial\theta\partial\theta'} - \Sigma(\theta) \right\| \xrightarrow{p} 0,$$

and this combined with Condition 1 and Lemma A.1, gives that

$$T^{-1}\sum_{t=1}^{T} \frac{\partial^2 q_t(\hat{\theta}_T)}{\partial\theta\partial\theta'} \xrightarrow{p} \Sigma(\theta_0) = \Sigma_0.$$

Hence a consistent estimator of the asymptotic covariance is

$$\hat{\Psi}_T = \left(-T^{-1} \sum_{t=1}^{T} \frac{\partial^2 q_t(\hat{\theta}_T)}{\partial\theta\partial\theta'} \right)^{-1}$$
$$= \left(T^{-1} \sum_{t=1}^{T} \frac{1}{2} \frac{1}{\hat{\omega}_T + \hat{\alpha}_T x_{t-1}^2} \left(1 - \frac{2x_t^2}{\hat{\omega}_T + \hat{\alpha}_T x_{t-1}^2} \right) w_t w_t' \right)^{-1},$$

that satisfies

$$\hat{\Psi}_T \xrightarrow{p} -\Sigma_0^{-1}.$$

For this choice of estimator, under the hypothesis in (VI.13), the Wald statistic satisfies

$$W_{\text{rob}} = \frac{T(\hat{\alpha}_T - 1/\sqrt{3})^2}{(\hat{\Psi}_T)_{22}} \xrightarrow{D} \chi^2(1).$$

We end this section by considering and estimator for $\hat{\Psi}_T$. As demonstrated in the previous Example VI.4.2, in the case where $\sqrt{T}(\hat{\theta} - \theta_0)$ has asymptotic covariance given by $-\Sigma_0^{-1}$, we use

$$\hat{\Psi}_T = \left(-\frac{\partial^2 Q_T(\hat{\theta}_T)}{\partial\theta\partial\theta'} \right)^{-1},$$

which, as already argued, converges in probability to $-\Sigma_0^{-1}$. In the more general case where the asymptotic covariance is given by $\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}$ one would in addition need a consistent estimator for Ω_0. We restrict here our attention to estimators given as maximizers of criterion functions of the form

$$Q_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} q_t(\theta). \tag{VI.21}$$

This class of estimators includes the estimators for the parameters in AR and ARCH models already considered. We have the following result, which follows directly from applications of Lemma A.1.

Lemma VI.4.1 *Under the assumptions of Theorem VI.3.1,*

$$\hat{\Sigma}_T := \frac{\partial^2 Q_T(\hat{\theta}_T)}{\partial\theta\partial\theta'} \xrightarrow{p} \Sigma_0, \quad \text{as } T \to \infty.$$

Suppose in addition that the criterion function is given by (VI.21), and that there exists a matrix-valued function $\Omega(\theta)$ that is continuous at θ_0, such that

$$\sup_{\theta\in\mathcal{N}(\theta_0)} \left\| T^{-1} \sum_{t=1}^{T} \frac{\partial q_t(\theta)}{\partial\theta} \frac{\partial q_t(\theta)}{\partial\theta'} - \Omega(\theta) \right\| \xrightarrow{p} 0, \quad \text{as } T \to \infty,$$

and $\Omega(\theta_0) = \Omega_0$. *Then*

$$\hat{\Omega}_T := T^{-1} \sum_{t=1}^{T} \frac{\partial q_t(\hat{\theta}_T)}{\partial \theta} \frac{\partial q_t(\hat{\theta}_T)}{\partial \theta'} \xrightarrow{p} \Omega_0, \quad \text{as } T \to \infty,$$

and, consequently,

$$\hat{\Psi}_T := \hat{\Sigma}_T^{-1} \hat{\Omega}_T \hat{\Sigma}_T^{-1}$$

satisfies (VI.17).

Remark VI.4.2 *Note that the above lemma essentially restricts the criterion function to settings where the score contributions* $\partial q_t(\theta_0)/\partial\theta$ *are uncorrelated in time – or martingale differences – as considered for the AR and ARCH models. This is true, as the CLT for martingale differences, e.g., Theorem V.2.3, is compatible with a limiting covariance matrix given by* $\mathbb{E}[(\partial q_t(\theta_0)/\partial\theta)(\partial q_t(\theta_0)/\partial\theta')] = \Omega_0$. *If* $\partial q_t(\theta_0)/\partial\theta$ *is autocorrelated, the structure of the asymptotic covariance looks different; see, e.g.,Theorem I.4.3 in the context of geometrically ergodic processes. In such a case alternative covariance estimators are needed, see e.g. Francq and Zakoïan (2019, Chapter 5).*

VI.5 Concluding remarks

We conclude this note by emphasizing that the results for consistency and asymptotic normality applies to a general range of estimators. As demonstrated, the high-level conditions of Theorems VI.2.2 and VI.3.1 are typically manageable to verify in cases where the criterion function is of the form (VI.21) and the DGP is stationary and ergodic by relying on appropriate LLNs and CLTs.

An important assumption made in Theorem VI.3.1 is that the true parameter value θ_0 is an interior point of the parameter space. For instance, this rules out the case where $\alpha_0 = 0$ in the ARCH(1) model, as the parameter $\alpha \geq 0$. In such a case one can show that the limiting distribution of $\sqrt{T}(\hat{\alpha}_T - \alpha_0) = \sqrt{T}\hat{\alpha}_T \geq 0$ is non-Gaussian, and alternative arguments are needed. We refer to Cavaliere et al. (2022) for additional details.

References

Amemiya, T., 1985, *Advanced Econometrics*, Harvard University Press, 1st edition.

Cavaliere, G., Nielsen, H.B., Pedersen, R.S., & Rahbek, A., 2022, "Bootstrap inference on the boundary of the parameter space with application to conditional volatility models", *Journal of Econometrics*, Vol. 227, pp. 241–263.

Francq, C., & Zakoïan, J.-M., 2012, "Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models", *Econometrica*, Vol. 80, pp. 821–861.

Francq, C., & Zakoïan, J.-M., 2019, *GARCH Models: Structure, Statistical Inference and Financial Applications*, 2nd edition, Wiley.

Jensen, S.T., & Rahbek, A., 2004a, "Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case", *Econometrica*, Vol. 72, pp. 641–646.

Jensen, S.T., & Rahbek, A., 2004b, "Asymptotic inference for nonstationary GARCH", *Econometric Theory*, Vol. 20, pp. 1203–1226.

Kristensen, D., & Rahbek, A., 2005, "Asymptotics of the QMLE for a class of ARCH(q) models", *Econometric Theory*, Vol. 21, pp. 946–961.

Lange, T., Rahbek, A., & Jensen, S.T., 2011, "Estimation and asymptotic inference in the AR-ARCH model", *Econometric Reviews*, Vol. 30, pp. 129–153.

Newey, W.K., & McFadden, D., 1994, "Large sample estimation and hypothesis testing", in R.F. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Vol. 4, pp.2111-2245.

Ranga Rao, R., 1962, "Relations between weak and uniform convergence of measures with applications", *The Annals of Mathematical Statistics*, Vol. 33, 659–680.

Straumann, D., & Mikosch, T., 2006, "Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach", *The Annals of Statistics*, Vol. 34, pp. 2449–2495.

A Additional limit results

Theorem A.1 *Let $\{X_t\}_{t\in\mathbb{Z}}$ be stationary and ergodic with $X_t \in \mathbb{R}^d$. Let $\Theta \subset \mathbb{R}^k$ be compact, and let f be a measurable function from $(\mathbb{R}^d)^\infty \times \Theta$ to $\mathbb{R}$. Define*

$$q_t(\theta) = f(\dots, X_{t-1}, X_t, X_{t+1}, \dots; \theta).$$

If $q_t(\theta)$ is finite almost surely, then $\{q_t(\theta)\}_{t\in\mathbb{Z}}$ is stationary and ergodic. Furthermore, assume that $q_t(\theta)$ is continuous in θ almost surely and that $\mathbb{E}[\sup_{\theta\in\Theta} |q_t(\theta)|] < \infty$. Then

$$\sup_{\theta\in\Theta} \left| T^{-1} \sum_{t=1}^{T} q_t(\theta) - \mathbb{E}\left[q_t(\theta)\right] \right| \xrightarrow{p} 0 \quad \text{as } T \to \infty.$$

Proof. The stationarity and ergodicity of $\{q_t(\theta)\}_{t\in\mathbb{Z}}$ follows by Theorem V.2.1. The uniform convergence follows by Ranga Rao (1962); see also Mikosch and Straumann (2006, Section 2.3) for additional details. ■

Lemma A.1 *Suppose that*

1. $X_T \xrightarrow{p} x_0 \in \mathbb{R}^k$,
2. $\sup_{x\in B(x_0,\epsilon)} |g_T(\theta) - g(\theta)| \xrightarrow{p} 0$ *with* $B(x_0, \epsilon) := \{x : \|x - x_0\| < \epsilon\}$ *for some* $\epsilon > 0$, *and*
3. *the non-stochastic function $g(x)$ is continuous at x_0.*

Then $g_T(X_T) \xrightarrow{p} Q(x_0)$.

Proof. We have with probability approaching one,

$$\begin{aligned} |g_T(X_T) - g(x_0)| &\leq |g_T(X_T) - g(X_T)| + |g(X_T) - g(x_0)| \\ &\leq \sup_{x\in B(x_0,\epsilon)} |g_T(x) - g(x)| + |g(X_T) - g(x_0)|, \end{aligned}$$

where the first inequality follows by the triangle inequality, and second equality follows from the fact that $X_T \in B(x_0, \epsilon)$ with probability approaching one. Conditions 1 and 3 imply that $|g(X_T) - g(x_0)| \xrightarrow{p} 0$. This combined with Condition 2 yield that $|g_T(X_T) - g(x_0)| \xrightarrow{p} 0$. ■

Lemma A.2 (Lange et al., 2011, Lemma 3) *Let $\{x_t\}_{t=-k,-k+1,\dots}$ be a geometrically ergodic Markov chain for some finite $k \geq 0$. Let $w_t := (x_{t-k}, \dots, x_t)$ and let $q(w, \theta)$ be a real-valued measurable function which is continuous in θ*

for all w. Let Θ be a compact set of the same dimension as θ. Assume that $E[q(w_t^*, \theta)] = 0$ for all $\theta \in \Theta$, and that $E[\sup_{\theta\in\Theta} |q(w_t^*, \theta)|] < \infty$. Then

$$\sup_{\theta\in\Theta} \left| T^{-1} \sum_{t=1}^{T} q(w_t, \theta) \right| \xrightarrow{p} 0 \quad \textit{as } T \to \infty.$$

B Proofs

Proof of Theorem VI.2.2

The proof follows NM (pp.2121–2122). Let $Q(\hat{\theta}_T) := E[q_t(\theta)]|_{\theta=\hat{\theta}_T}$. For any $\epsilon > 0$ we have with probability approaching one that

$$\begin{aligned} Q_T(\hat{\theta}_T) &> Q_T(\theta_0) - \epsilon/3, \\ Q(\hat{\theta}_T) &> Q_T(\hat{\theta}_T) - \epsilon/3, \\ Q_T(\theta_0) &> Q(\theta_0) - \epsilon/3, \end{aligned}$$

where the first inequality follows by the fact that $\hat{\theta}_T$ is a maximizer of $Q_T(\theta)$ on Θ, and the second and third follow by Condition 4. Combining these inequalities, we have with probability approaching one,

$$Q(\hat{\theta}_T) > Q_T(\hat{\theta}_T) - \epsilon/3 > Q_T(\theta_0) - 2\epsilon/3 > Q(\theta_0) - \epsilon.$$

Hence, for any $\epsilon > 0$, with probability approaching one,

$$Q(\hat{\theta}_T) > Q(\theta_0) - \epsilon. \tag{VI.22}$$

Let $\mathcal{N}(\theta_0)$ be *any* open subset of Θ containing θ_0, and note that $\Theta \cap \mathcal{N}(\theta_0)^{\complement}$ is compact. Hence, using continuity of $Q(\theta)$, there exists a $\theta^* \in \Theta \cap \mathcal{N}(\theta_0)^{\complement}$ such that

$$Q(\theta^*) = \sup_{\theta\in\Theta\cap\mathcal{N}(\theta_0)^{\complement}} Q(\theta).$$

Since θ_0 is the unique maximizer of $Q(\theta)$ on Θ, and $\theta_0 \notin \Theta \cap \mathcal{N}(\theta_0)^{\complement}$, it must hold that

$$Q(\theta^*) < Q(\theta_0).$$

Choosing $\epsilon = Q(\theta_0) - Q(\theta^*) > 0$, it follows that with probability approaching one,

$$Q(\hat{\theta}_T) > Q(\theta_0) - \epsilon > Q(\theta^*) = \sup_{\theta\in\Theta\cap\mathcal{N}(\theta_0)^{\complement}} Q(\theta).$$

Hence, $\hat{\theta}_T \in \mathcal{N}(\theta_0)$ with probability approaching one, that is, $\mathbb{P}(\hat{\theta}_T \in \mathcal{N}(\theta_0)) \to 1$. Note that this holds for any $\mathcal{N}(\theta_0)$ (being an open subset of Θ containing

θ_0). For any $\epsilon > 0$, there exists an open subset $\mathcal{N}_\epsilon(\theta_0)$ such that the event $\hat{\theta}_T \in \mathcal{N}_\epsilon(\theta_0) \implies \|\hat{\theta}_T - \theta_0\| < \epsilon$. Hence, for any $\epsilon > 0$,

$$\mathbb{P}\left(\|\hat{\theta}_T - \theta_0\| < \epsilon\right) \geq \mathbb{P}(\hat{\theta}_T \in \mathcal{N}_\epsilon(\theta_0)) \rightarrow 1,$$

and we conclude that $\hat{\theta}_T \xrightarrow{p} \theta_0$ as $T \rightarrow \infty$.

Proof of Theorem VI.4.1

The proof follows the arguments given in the proofs of Theorems 9.1 and 9.2 in NM. Recall that.

$$LR_T\,(H) = 2T[Q_T(\hat{\theta}_T) - Q_T(\tilde{\theta}_T)].$$

A second order mean-value expansion of $Q_T(\tilde{\theta}_T)$ around $\hat{\theta}_T$ gives that

$$Q_T(\tilde{\theta}_T) = Q_T(\hat{\theta}_T) + \frac{\partial Q_T(\hat{\theta}_T)}{\partial \theta'}(\tilde{\theta}_T - \hat{\theta}_T) + \frac{1}{2}(\tilde{\theta}_T - \hat{\theta}_T)'\frac{\partial^2 Q_T(\theta_T^*)}{\partial\theta\partial\theta'}(\tilde{\theta}_T - \hat{\theta}_T),$$

with θ_T^* between $\tilde{\theta}_T$ and $\hat{\theta}_T$, and $\theta_T^* \xrightarrow{p} \theta_0$. Recall that with probability approaching one $\partial Q_T(\hat{\theta}_T)/\partial\theta' = 0_{k\times 1}$, so that

$$LR_T\,(H) = T(\hat{\theta}_T - \tilde{\theta}_T)'\left[-\frac{\partial^2 Q_T(\theta_T^*)}{\partial\theta\partial\theta'}\right](\hat{\theta}_T - \tilde{\theta}_T). \tag{VI.23}$$

Note that the right-hand side of (VI.23) is a Wald-type statistic analogous to (VI.15). From Lemma VI.4.1 $\partial^2 Q_T(\theta_T^*)/\partial\theta\partial\theta' \xrightarrow{p} \Sigma_0$, so it remains to find the limiting distribution of

$$\sqrt{T}(\hat{\theta}_T - \tilde{\theta}_T) = \sqrt{T}(\hat{\theta}_T - \theta_0) - \sqrt{T}(\tilde{\theta}_T - \theta_0).$$

To do so, we derive expressions for $\sqrt{T}(\hat{\theta}_T - \theta_0)$ [Step 1] and $\sqrt{T}(\tilde{\theta}_T - \theta_0)$ [Step 2], and combine [Step 3], and lastly plug into (VI.23) [Step 4].

Step 1: From the arguments given in the proof of Theorem VI.3.1, we have that

$$\sqrt{T}(\hat{\theta}_T - \theta_0) = -\Sigma_0^{-1}\sqrt{T}\frac{\partial Q_T(\theta_0)}{\partial\theta} + o_p(1). \tag{VI.24}$$

Step 2: Note that $\tilde{\theta}_T$ solves a maximization problem with Lagrangian

$$\mathcal{L}_T(\theta) = Q_T(\theta) - [R'\theta - r]'\lambda_T,$$

where λ_T is a $(l \times 1)$ vector of Lagrange multipliers. The first-order condition to this maximization problem is given by

$$0_{k\times 1} = \sqrt{T}\frac{\partial Q_T(\tilde{\theta}_T)}{\partial \theta} - \sqrt{T}R\lambda_T. \tag{VI.25}$$

A mean-value expansion of $\partial Q_T(\tilde{\theta}_T)/\partial\theta$ around θ_0 gives

$$\frac{\partial Q_T(\tilde{\theta}_T)}{\partial \theta} = \frac{\partial Q_T(\theta_0)}{\partial \theta} + \frac{\partial^2 Q_T(\theta_T^{**})}{\partial\theta\partial\theta'}(\tilde{\theta}_T - \theta_0), \tag{VI.26}$$

with θ_T^{**} between $\tilde{\theta}_T$ and θ_0, and $\theta_T^{**} \xrightarrow{p} \theta_0$. Note that by Lemma VI.4.1 $\partial^2 Q_T(\theta_T^{**})/\partial\theta\partial\theta' \xrightarrow{p} \Sigma_0$, so $\partial^2 Q_T(\theta_T^{**})/\partial\theta\partial\theta'$ is invertible with probability approaching one. Under the hypothesis H it holds that $r = R'\theta_0$, such that

$$0_{l\times 1} = [R'\tilde{\theta}_T - r] = R'(\tilde{\theta}_T - \theta_0). \tag{VI.27}$$

Combining (VI.25)-(VI.27), gives that

$$\begin{bmatrix} 0_{k\times 1} \\ 0_{l\times 1}\end{bmatrix} = \begin{bmatrix} \frac{\partial Q_T(\theta_0)}{\partial\theta} \\ 0 \end{bmatrix} - \begin{bmatrix} -\frac{\partial^2 Q_T(\theta_T^{**})}{\partial\theta\partial\theta'} & R \\ R' & 0 \end{bmatrix}\begin{bmatrix} \sqrt{T}(\tilde{\theta}_T - \theta_0) \\ \sqrt{T}\lambda_T \end{bmatrix},$$

and we seek to solve for $\sqrt{T}(\tilde{\theta}_T - \theta_0)$. We have the following result from linear algebra: Suppose that the $(l \times k)$ matrix A with $l \leq k$ has rank l, and that the $(k \times k)$ matrix B has full rank. Then

$$\begin{bmatrix} B & A' \\ A & 0 \end{bmatrix}^{-1} = \begin{bmatrix} B^{-1/2}MB^{-1/2} & B^{-1}A'(AB^{-1}A')^{-1} \\ (AB^{-1}A')^{-1}AB^{-1} & -(AB^{-1}A')^{-1}\end{bmatrix},$$

with $M = I_k - B^{-1/2}A'(AB^{-1}A')^{-1}AB^{-1/2}$. Using this result, we find that

$$\sqrt{T}(\tilde{\theta}_T - \theta_0) = \left(-\frac{\partial^2 Q_T(\theta_T^{**})}{\partial\theta\partial\theta'}\right)^{-1/2} M_T \left(-\frac{\partial^2 Q_T(\theta_T^{**})}{\partial\theta\partial\theta'}\right)^{-1/2}\sqrt{T}\frac{\partial Q_T(\theta_0)}{\partial\theta},$$

with

$$\begin{aligned} M_T &:= I_k - \left(-\frac{\partial^2 Q_T(\theta_T^{**})}{\partial\theta\partial\theta'}\right)^{-1/2} R\left(R'\left(-\frac{\partial^2 Q_T(\theta_T^{**})}{\partial\theta\partial\theta'}\right)^{-1} R\right)^{-1} R'\left(-\frac{\partial^2 Q_T(\theta_T^{**})}{\partial\theta\partial\theta'}\right)^{-1/2} \\ &\xrightarrow{p} I_k - (-\Sigma_0)^{-1/2}R\left[R'(-\Sigma_0)^{-1}R\right]^{-1}R'(-\Sigma_0)^{-1/2}. \end{aligned}$$

Since $\sqrt{T}\partial Q_T(\theta_0)/\partial\theta = O_p(1)$,

$$\begin{aligned}
&\sqrt{T}(\tilde{\theta}_T - \theta_0) \\
&= (-\Sigma_0)^{-1} - (-\Sigma_0)^{-1}R\left[R'(-\Sigma_0)^{-1}R\right]^{-1}R'(-\Sigma_0)^{-1}\sqrt{T}\frac{\partial Q_T(\theta_0)}{\partial\theta} + o_p(1).
\end{aligned} \tag{VI.28}$$

Step 3: Combining (VI.24) and (VI.28) gives that

$$\begin{aligned}
\sqrt{T}(\hat{\theta}_T - \tilde{\theta}_T) &= \sqrt{T}(\hat{\theta}_T - \theta_0) - \sqrt{T}(\tilde{\theta}_T - \theta_0) \\
&= \left[(-\Sigma_0)^{-1}R\left[R'(-\Sigma_0)^{-1}R\right]^{-1}R'(-\Sigma_0)^{-1}\right]\sqrt{T}\frac{\partial Q_T(\theta_0)}{\partial\theta} + o_p(1) \\
&\xrightarrow{D} Z,
\end{aligned} \tag{VI.29}$$

with $Z \overset{D}{=} N(0, B)$ and, using that $\Omega_0 = -\Sigma_0$,

$$\begin{aligned}
B &= \left[(-\Sigma_0)^{-1}R\left[R'(-\Sigma_0)^{-1}R\right]^{-1}R'(-\Sigma_0)^{-1}\right]\Omega_0\left[(-\Sigma_0)^{-1}R\left[R'(-\Sigma_0)^{-1}R\right]^{-1}R'(-\Sigma_0)^{-1}\right]' \\
&= \left[(-\Sigma_0)^{-1}R\left[R'(-\Sigma_0)^{-1}R\right]^{-1}R'(-\Sigma_0)^{-1}\right](-\Sigma_0)\left[(-\Sigma_0)^{-1}R\left[R'(-\Sigma_0)^{-1}R\right]^{-1}R'(-\Sigma_0)^{-1}\right]' \\
&= \left[(-\Sigma_0)^{-1}R\left[R'(-\Sigma_0)^{-1}R\right]^{-1}R'(-\Sigma_0)^{-1}\right].
\end{aligned}$$

Step 4: From (VI.23), using that $\sqrt{T}(\hat{\theta}_T - \tilde{\theta}_T) = O_p(1)$,

$$LR_T(H) = T(\hat{\theta}_T - \tilde{\theta}_T)'(-\Sigma_0)(\hat{\theta}_T - \tilde{\theta}_T) + o_p(1).$$

Noting that B has rank l, and that $B(-\Sigma_0)B = B$, it follows from Lemma 9.7 in NM and (VI.29) that

$$LR_T(H) \xrightarrow{D} Z'(-\Sigma_0)Z \overset{D}{=} \chi^2_l.$$