SOLUTIONS TO SELECTED EXERCISES

ADVANCED ECONOMETRIC METHODS

A GUIDE TO ESTIMATION AND INFERENCE FOR NONLINEAR DYNAMIC MODELS

Francisco Blasques

2019

ADVANCED TEXTS IN ECONOMETRICS
MACHINE LEARNING AND DATA SCIENCE

This book was typeset by the author using LaTeX.

Published by **A.** Publications' advanced academic textbook series.
Advanced Texts in Econometrics, Machine Learning and Data Science.

Contents

Exercises of Chapter 3

3.1 Recall that a model is a collection of probability measures. Recall further that, by definition, model A is nested by model B if every probability measure p which is an element of model A is also an element of model B. In other words, $p \in A \Rightarrow p \in B$.

Now, just note that every probability measure which is an element of model A is also an element of model A. Indeed, $p \in A \Rightarrow p \in A$ trivially. Therefore, model A is nested by model A. In other words, model A nests itself; i.e. $A \subseteq A$.

3.2 By definition, two models, A and B, are said to be different, and denoted $A \neq B$, if there exists at least one probability measure p satisfying

$$p \in A \text{ but } p \notin B \quad \text{or} \quad p \in B \text{ but } p \notin A. \tag{1}$$

Now, since model A nests model B, we have that $p \in B \Rightarrow p \in A$. Furthermore, since model B nests model A, we have $p \in A \Rightarrow p \in B$. Hence, there does not exist a p satisfying (1) and we have $A = B$.

3.3 This statement is, strictly speaking, false. A Gaussian AR(1) model, is indeed capable of generating many time-series with temporal dependence. For example, the probability measure implicitly defined by

$$x_{t+1} = \alpha + \beta x_{t-1} + \epsilon_t \ , \ \epsilon_t \sim N(0, \sigma^2)$$

with $(\alpha, \beta, \sigma^2) = (0.2, 0.9, 5.1)$ has temporal dependence since each x_t is dependent on its lags and leads. However, the Gaussian AR(1) is also capable of generating many time series with no temporal dependence. For example, the probability measure implicitly defined by the parameter vectors $(\alpha, \beta, \sigma^2) = (0.2, 0, 5.1)$, and $(\alpha, \beta, \sigma^2) = (10, 0, 0.1)$, and $(\alpha, \beta, \sigma^2) = (110, 0, 3.7)$, etc.

3.4 (a) this model is not well specified as there is no parameter vector (a, b) such that $\{x_t\}_{t\in\mathbb{Z}} \sim \text{NID}(0, 1)$.

(d) this model is well specified since we can generate the true probability measure and obtain $\{x_t\}_{t\in\mathbb{Z}} \sim \text{NID}(0, 1)$ for $(\mu, \sigma^2, \delta) = (0, 1, 0)$.

3.4 (f) This model is well specified since we can generate the true probability measure and obtain $\{x_t\}_{t\in\mathbb{Z}} \sim \text{NID}(0, 1)$ for $(\theta_0, \theta_1, \theta_2) = (1/\sqrt{9}, 0, 0)$ or $(\theta_0, \theta_1, \theta_2) = (0, 1/\sqrt{9}, 0)$ or $(\theta_0, \theta_1, \theta_2) = (0, 0, 1/\sqrt{9})$.

3.5 This statement is false. The collection of *all probability measures* is a model. This model is, by construction, correctly specified. Note that talking about 'the collection of all probability measures' should not be disregarded as a theoretical trick. In appropriate settings, very flexible models with non-parametric estimators, or sieve estimators of semi-nonparametric models may attempt to generate all possible measures asymptotically.

3.6 This statement is false. Of course, models can share certain probability measures or nest each other. In general, the fact that $p_0 \in A$ does not imply that $p_0 \notin B$ for some other B.

3.8 (a) The RCAR(1) nests the Gaussian AR(1). The AR(1) is given by

$$x_{t+1} = \rho x_t + \epsilon_t \quad , \quad \{\epsilon_t\} \sim \text{NID}(0, \sigma_\epsilon^2).$$

The RCAR(1) is given by

$$x_{t+1} = \beta_t x_t + e_t \ ,$$

where

$$\{\beta_t\} \sim \text{NID}(b, \sigma_\beta^2) \ \text{ and } \quad \{\epsilon_t\} \sim \text{NID}(0, \sigma_e^2).$$

Note that any probability measure generated by the AR(1) can also be generated by the RCAR(1) by setting $(b, \sigma^2_\beta, \sigma^2_e) = (\rho, 0, \sigma^2_\epsilon)$.

3.8 (d) The Gaussian observation-driven local-level model nests the Gaussian AR(1). The AR(1) model is given by

$$x_{t+1} = \rho x_t + \epsilon_t \quad , \quad \{\epsilon_t\} \sim \text{NID}(0, \sigma^2_\epsilon).$$

The local-level model is given by

$$x_t = \mu_t + e_t \; , \quad \{e_t\} \sim \text{NID}(0, \sigma^2_e) \; , \quad \mu_{t+1} = \omega + \alpha(x_t - \mu_t) + \beta\mu_t \; .$$

Note that any probability measure generated by the AR(1) under a parameter vector $(\rho, \sigma^2_\epsilon)$ can also be generated by the LL model by setting $(\omega, \alpha, \beta, \sigma^2_e) = (0, \rho, \rho, \sigma^2_\epsilon)$.

3.8 (e) The GARCH model, when parameterized by the vector (ω, α, β), is given by

$$x_t = \sigma_t \varepsilon_t \; , \quad \{\varepsilon_t\}_{t \in \mathbb{Z}} \sim \text{NID}(0, 1) \; ,$$

$$\sigma^2_t = \omega + \alpha x^2_{t-1} + \beta \sigma^2_{t-1}.$$

In turn, the QGARCH, parameterised by the vector $(\omega^*, \alpha^*, \gamma^*, \beta^*)$, is given by

$$x_t = \sigma_t \varepsilon_t \; , \quad \{\varepsilon_t\} \sim \text{NID}(0, 1) \; ,$$

$$\sigma^2_t = \omega^* + \alpha^* x^2_{t-1} + \gamma^* x_{t-1} + \beta^* \sigma^2_{t-1}.$$

Clearly, the QGARCH nests the GARCH model since any probability measure implicitly defined by the GARCH under the vector (ω, α, β), can also be obtained by the QGARCH by setting $(\omega^*, \alpha^*, \gamma^*, \beta^*) = (\omega, \alpha, 0, \beta)$.

Exercises of Chapter 4

4.1 (a) False. The answer can actually be found in the previous chapter, in Section 3.5.

4.1 (b) False. Just consider a sequence of *white noise* random variables $\{x_t\}_{t\in\mathbb{Z}}$; revisit p.102 for a definition. Such a sequence is weakly stationary as the mean does not change over time ($\mathbb{E}(x_t) = 0 \ \forall \ t$), the variance does not change over time ($\mathbb{V}\text{ar}(x_t) = \sigma^2 \ \forall \ t$), and the autocovariance of any order does not change in time either ($\mathbb{C}\text{ov}(x_t, x_{t-h}) = 0 \ \forall \ (h,t)$). However, such a sequence may well fail to be strictly stationary if other distributional features change over time. For example, the skewness or the Kurtosis might be changing ($\text{Skew}(x_t) \neq \text{Skew}(x_{t'})$ or $\text{Kurt}(x_t) \neq \text{Kurt}(x_{t'})$ for some $t \neq t'$).

4.1 (c) False. Just think of the moment requirements imposed by weakly stationary processes! Consider for example a sequence of iid Student-t random variables, $\{x_t\}_{t\in\mathbb{Z}} \sim$TID$(\lambda)$ with $\lambda = 1.5$. Note that each random variable in this sequence $x_t \sim t(\lambda)$ does not have two moments since λ is smaller than 2. As a result the variance is not defined, and we cannot state that this sequence is weakly stationary. By its iid nature, it is however strictly stationary.

4.2 (a) The solution is immediate by application of Krengel's Theorem 4.4, p.84. Show it!

4.2 (b) Again, the solution is immediate by application of Krengel's Theorem 4.4, p.84. Show it!

4.2 (c) A solution to the similar Gaussian AR(1) can be found in p.87 using the unfolding method, or in p.93 using Bougerol's theorem.

4.2 (d) We cannot show that data generated by this model is SE because the white noise property of the innovations is not sufficient for our purposes. In particular, it is not clear if the innovations are SE. In principle, they might fail to be SE as they are only assumed to be white noise.

4.2 (e) The solution is immediate by application of Krengel's Theorem 4.4, p.84.

4.2 (f) A solution to a similar RCAR(1) can be found in p.89 using the unfolding method, or in p.95 using Bougerol's theorem.

4.2 (g) The solution is immediate by application of Krengel's Theorem 4.4, p.84.

4.2 (j) The sequence $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ generated by the logistic SESTAR:

$$x_t = g(x_{t-1};\gamma)x_{t-1} + \varepsilon_t \ , \quad \{\varepsilon_t\}_{t\in\mathbb{Z}} \sim NID(0,\sigma_\varepsilon^2) \ ,$$

$$g(x_{t-1};\gamma) \;=\; -0.2 + \frac{\gamma}{1+\exp(2+1.4x_{t-1})} \qquad \forall\, t\in\mathbb{Z} \ ,$$

is SE if the parameter γ is such that

$$\sup_{x\in\mathbb{R}} \left| \frac{\partial g(x;\gamma)}{\partial x} x + g(x;\gamma) \right| < 1.$$

Indeed, $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is SE because, by Bougerol's Theorem, the sequence $\{x_t(\boldsymbol{\theta},x_1)\}_{t\in\mathbb{N}}$ initialized at x_1 converges e.a.s. to an SE limit for any $x_1\in\mathbb{R}$. Bougerol's theorem holds because conditions A1, A2 and A3 are satisfied.

Condition A1 holds because the innovations $\{\epsilon_t\}$ are iid.

Condition A2 holds for any x_1 because

$$\begin{aligned}
\mathbb{E}\log^{+}|\phi(x_1,\epsilon_t,\boldsymbol{\theta})| &= \mathbb{E}\log^{+}|g(x_1;\gamma)x_1+\varepsilon_t| \\
(\log^{+}(z)\leq z\ \forall\ z>0) \quad &\leq \mathbb{E}|g(x_1;\gamma)x_1+\varepsilon_t| \\
(\text{sub-additivity of } |\cdot|) \quad &\leq \mathbb{E}\Big[\ |g(x_1;\gamma)x_1|+|\varepsilon_t|\ \Big] \\
(\text{linearity of expectation}) \quad &\leq \underbrace{|g(x_1;\gamma)|}_{<\infty}\underbrace{|x_1|}_{<\infty}+\underbrace{\mathbb{E}|\varepsilon_t|}_{<\infty}<\infty
\end{aligned}$$

Note that $\mathbb{E}|\varepsilon_t|<\infty$ because ε_t is Gaussian and hence has bounded moments of any order.

Condition A3 holds because

$$\begin{aligned}
&\mathbb{E}\log\sup_x\left|\frac{\partial\phi(x,\epsilon_t,\boldsymbol{\theta})}{\partial x}\right| \\
&\quad = \mathbb{E}\log\sup_x\left|\frac{\partial g(x,\gamma)}{\partial x}x+g(x,\gamma)\right|<0 \\
&\Leftrightarrow \log\sup_x\left|\frac{\partial g(x,\gamma)}{\partial x}x+g(x,\gamma)\right|<0 \\
&\Leftrightarrow \sup_x\left|\frac{\partial g(x,\gamma)}{\partial x}x+g(x,\gamma)\right|<1.
\end{aligned}$$

In some sense, the proof can end here. Of course, if we were asked to give values of γ that satisfy the contraction above, then we could also try to find a more explicit sufficient condition. In particular, note that

$$\begin{aligned}
&\mathbb{E}\log\sup_x\left|\frac{\partial g(x,\gamma)}{\partial x}x+g(x,\gamma)\right|<0 \\
&\quad \Leftarrow \quad \sup_x\left|\frac{\partial g(x,\gamma)}{\partial x}x+g(x,\gamma)\right|<1 \\
&\Leftrightarrow \quad \sup_x\Bigg|-\frac{\gamma 1.4\exp(2+1.4x)}{(1+\exp(2+1.4x))^2}x \\
&\qquad\qquad -0.2+\frac{\gamma}{1+\exp(2+1.4x)}\Bigg|<1
\end{aligned}$$

Finally, defining $w(x) = \exp(2 + 1.4x)$ we have

$$\begin{aligned}
&\sup_x \Big| -\frac{\gamma 1.4 w(x)}{(1+w(x))^2} x - 0.2 + \frac{\gamma}{1+w(x)} \Big| \\
&\quad \leq \sup_x \Big| \frac{\gamma 1.4 w(x)}{(1+w(x))^2} x \Big| + 0.2 + \sup_x \Big| \frac{\gamma}{1+w(x)} \Big| \\
&\qquad \text{(supremum sub-additivity)} \\
&\quad \leq 1.4|\gamma| \sup_x \Big| \frac{w(x)}{(1+w(x))^2} x \Big| + 0.2 + |\gamma| \sup_x \Big| \frac{1}{1+w(x)} \Big| \\
&\text{(positive homogeneity of supremum)} \\
&\quad \leq |\gamma| 1.4k + 0.2 + |\gamma| \\
&\sup_x |w(x)x/(1+w(x))^2| < k < \infty \\
&\qquad \text{and } \sup_x |1/(1+w(x))| < 1
\end{aligned}$$

and hence, we have that

$$\begin{aligned}
\sup_x \Big| \frac{\partial g(x,\gamma)}{\partial x} x + g(x,\gamma) \Big| < 1 \quad &\Leftarrow \quad |\gamma| 1.4k + 0.2 + |\gamma| < 1 \\
&\Leftrightarrow \quad |\gamma| < 0.8/(1 + 1.4k).
\end{aligned}$$

Note that, by supremum sub-additibity, we have that $\sup_x |f(x) + g(x)| \leq \sup_x |f(x)| + \sup_x |g(x)|$, and by Positive homogeneity of the supremum, we have that $\sup_x |af(x) + g(x)| \leq |a| \sup_x |f(x)| + \sup_x |g(x)|$.

4.2 (l) An answer can be obtained by extension of the quadratic AR(1) example in p.99.

4.2 (m) An answer is available as a special case of Example 4.4 in p.98.

4.2 (n) An answer is given in p.98.

4.2 (p) The sequence $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ generated by the Gaussian local-level model

$$x_t = \mu_t + \varepsilon_t \ , \quad \{\varepsilon_t\} \sim \text{NID}(0, 3) \ ,$$

$$\mu_t = 5 + 0.9\mu_{t-1} + v_t \ , \quad \{v_t\} \sim \text{NID}(0, 0.6)$$

is indeed SE. We will first verify that, by Bougerol's theorem, the mean sequence $\{\mu_t(\boldsymbol{\theta}, \mu_1)\}_{t\in\mathbb{N}}$ initialised at μ_1 converges e.a.s. to an SE limit $\{\mu_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ for any $\mu_1 \in \mathbb{R}$. Second, we will show that this implies that the sequence $\{x_t(\boldsymbol{\theta},\mu_1)\}_{t\in\mathbb{N}}$ with mean $\{\mu_t(\boldsymbol{\theta},\mu_1)\}_{t\in\mathbb{N}}$ initialised at μ_1 converges e.a.s. to an SE limit $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ with mean $\{\mu_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ for any $\mu_1 \in \mathbb{R}$.

Let us apply Bougerol's theorem to the conditional mean equation

$$\mu_t = 5 + 0.9\mu_{t-1} + v_t \ , \quad \{v_t\} \sim \mathrm{NID}(0, 0.6).$$

Condition A1 of Bougerol's theorem requires that $\{v_t\}_{t\in\mathbb{Z}}$ be SE. This condition is trivially satisfied because the innovations $\{v_t\}$ are iid.

Condition A2 of Bougerol's theorem requires that $\mathbb{E}\log^+|\phi(\mu_1, v_t, \boldsymbol{\theta})| < \infty$ for some μ_1. This condition actually holds for any μ_1 because

$$\begin{aligned}
\mathbb{E}\log^+|\phi(\mu_1, v_t, \boldsymbol{\theta})| &= \mathbb{E}\log^+|5 + 0.9\mu_1 + v_t| \\
\text{(}\log^+(z) \le z \ \forall\ z > 0\text{)} \qquad &\le \mathbb{E}|5 + 0.9\mu_1 + v_t| \\
\text{(sub-additivity of } |\cdot|\text{)} \qquad &\le \mathbb{E}\Big[\ 5 + 0.9|\mu_1| + |v_t|\ \Big] \\
\text{(linearity of expectation)} \qquad &\le \underbrace{5}_{<\infty} + \underbrace{0.9}_{<\infty}\underbrace{|\mu_1|}_{<\infty} + \underbrace{\mathbb{E}|v_t|}_{<\infty} < \infty
\end{aligned}$$

Note: $\mathbb{E}|v_t| < \infty$ because v_t is Gaussian and hence has bounded moments of any order!

Condition A3 of Bougerol's theorem requires the contraction $\mathbb{E}\log\sup_\mu|\partial\phi(\mu, v_t, \boldsymbol{\theta})/\partial\mu| < 0$ to hold. This condition is clearly satisfied because

$$\mathbb{E}\log\sup_\mu\left|\frac{\partial\phi(\mu, v_t, \boldsymbol{\theta})}{\partial\mu}\right| = \mathbb{E}\log\sup_\mu|0.9| = \log 0.9 < 0$$

We conclude by Bourgerol's theorem that the sequence $\{\mu_t(\boldsymbol{\theta},\mu_1)\}_{t\in\mathbb{N}}$ initialised at μ_1 converges e.a.s. to an SE limit $\{\mu_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ for any $\mu_1 \in \mathbb{R}$.

Finally, we conclude that

$$\left\{ x_t(\boldsymbol{\theta}, \mu_1) \right\}_{t\in\mathbb{N}} = \left\{ \mu_t(\boldsymbol{\theta}, \mu_1) + \varepsilon_t \right\}_{t\in\mathbb{N}}$$

converges e.a.s. to the SE limit

$$\left\{ x_t(\boldsymbol{\theta}) \right\}_{t\in\mathbb{Z}} = \left\{ \mu_t(\boldsymbol{\theta}) + \varepsilon_t \right\}_{t\in\mathbb{Z}}$$

for any $\mu_1 \in \mathbb{R}$ because

$$\begin{aligned} |x_t(\boldsymbol{\theta}, \mu_1) - x_t(\boldsymbol{\theta})| &= |\mu_t(\boldsymbol{\theta}, \mu_1) + \varepsilon_t - \mu_t(\boldsymbol{\theta}) - \varepsilon_t| \\ &= |\mu_t(\boldsymbol{\theta}, \mu_1) - \mu_t(\boldsymbol{\theta})| \overset{e.a.s.}{\to} 0. \quad \square \end{aligned}$$

4.2 (r) An answer can actually be found in Chapter 5, p.138-139.

4.2 (s) An answer can be found in p.121 and p.123.

4.2 (u) We cannot conclude that the sequence $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ generated by the NGARCH:

$$x_t = \sigma_t \varepsilon_t \ , \quad \{\varepsilon_t\} \sim \text{NID}(0,1) \ ,$$

$$\sigma_t^2 = 0.5 + 0.1(x_{t-1} - \sigma_{t-1})^2 + 0.9\sigma_{t-1}^2$$

is SE because the contraction condition of Bougerol's theorem does not hold for the volatility update equation. In particular, by Bougerol's theorem $\{\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)\}_{t\in\mathbb{N}}$ converges to a limit SE sequence $\{\sigma_t^2(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ for every initialization σ_1^2 if conditions A1, A2 and A3 hold.

Condition A1 holds because $\{\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)\}_{t\in\mathbb{N}}$ is generated by a Markov system

$$\begin{aligned} \sigma_t^2 &= 0.5 + 0.1(x_{t-1} - \sigma_{t-1})^2 + 0.9\sigma_{t-1}^2 \\ &= 0.5 + 0.1(\sigma_{t-1}\epsilon_{t-1} - \sigma_{t-1})^2 + 0.9\sigma_{t-1}^2 \\ &= 0.5 + 0.1\epsilon_{t-1}^2\sigma_{t-1}^2 - 0.2\epsilon_{t-1}\sigma_{t-1}^2 + \sigma_{t-1}^2 \end{aligned}$$

with iid innovations $\{\epsilon_t\}$.

Condition A2 holds for any σ_1^2 because

$$\begin{aligned}
&\mathbb{E}\log^+|\phi(\sigma_1^2,\epsilon_t,\boldsymbol{\theta})| \\
&\quad = \mathbb{E}\log^+\left|0.5+0.1\epsilon_t^2\sigma_1^2-0.2\epsilon_t\sigma_1^2+\sigma_1^2\right| \\
&\quad \leq \mathbb{E}\left|0.5+0.1\epsilon_t^2\sigma_1^2-0.2\epsilon_t\sigma_1^2+\sigma_1^2\right| \\
&\quad \text{(}\log^+(z)\leq z\ \forall\ z>0\text{)} \\
&\quad \leq \mathbb{E}\Big[\ 0.5+0.1|\epsilon_t^2\sigma_1^2|+0.2|\epsilon_t||\sigma_1^2|+|\sigma_1^2|\ \Big] \\
&\quad \text{(sub-additivity of } |\cdot|\text{)} \\
&\quad \leq \underbrace{0.5}_{<\infty}+\underbrace{0.1|\sigma_1^2|}_{<\infty}\underbrace{\mathbb{E}|\epsilon_t^2|}_{<\infty}+\underbrace{0.2|\sigma_1^2|}_{<\infty}\underbrace{\mathbb{E}|\epsilon_t|}_{<\infty}+\underbrace{|\sigma_1^2|}_{<\infty}<\infty \\
&\quad \text{(linearity of expectation)}
\end{aligned}$$

Note that $\mathbb{E}|\epsilon_t|<\infty$ and $\mathbb{E}|\epsilon_t|^2<\infty$ because ϵ_t is Gaussian and hence has bounded moments of any order.

However, condition A3 does not hold because

$$\begin{aligned}
\mathbb{E}\log\sup_{\sigma^2}\left|\frac{\partial\phi(\sigma^2,v_t,\boldsymbol{\theta})}{\partial\sigma^2}\right| &= \mathbb{E}\log\sup_{\sigma^2}|0.1\epsilon_t^2-0.2\epsilon_t+1| \\
&\approx 0.0746>0. \quad \square
\end{aligned}$$

4.3 In addition to finding which models generate SE data (which we have done in question 4.2), we should find which models generate data with one bounded moment (for an LLN), or two bounded moments. This can be easily achieved by application of the Power-n Theorem or the Uniform Contraction Theorem.

4.3 (a) A solution can be found by application of the c_n inequality on p.85.

4.3 (b) A solution can be found by application of the c_n inequality on p.85.

4.3 (c) A solution is given in p.116.

4.3 (g) A solution can be found by application of the c_n inequality on p.85.

4.3 (m) The solution is a special case of that given in p.117.

4.3 (n) A solution is given in p.117.

4.3 (p)

The sequence $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ generated by the Gaussian local-level model

$$x_t = \mu_t + \varepsilon_t \ , \quad \{\varepsilon_t\} \sim \text{NID}(0,3) \ ,$$

$$\mu_t = 5 + 0.9\mu_{t-1} + v_t \ , \quad \{v_t\} \sim \text{NID}(0,0.6)$$

satisfies an LLN and can be shown to have two bounded moments.

First, we will show that, by the power-n theorem, the conditional mean sequence $\{\mu_t(\boldsymbol{\theta},\mu_1)\}_{t\in\mathbb{N}}$ initialized at μ_1 converges e.a.s. to an SE limit sequence $\{\mu_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ that satisfies $\mathbb{E}|\mu_t(\boldsymbol{\theta})|^2 < \infty$ for any $\mu_1 \in \mathbb{R}$. Second, we show that this implies that the sequence $\{x_t(\boldsymbol{\theta},\mu_1)\}_{t\in\mathbb{N}}$ with mean $\{\mu_t(\boldsymbol{\theta},\mu_1)\}_{t\in\mathbb{N}}$ initialized at μ_1 converges e.a.s. to an SE limit $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ that satisfies $\mathbb{E}|x_t(\boldsymbol{\theta})|^2 < \infty$ for any $\mu_1 \in \mathbb{R}$. We conclude that the sequence satisfies an LLN.

Let us apply the Power-n theorem with $n > 2$. In particular, we can take $n = 3$ and apply the theorem to the conditional mean equation

$$\mu_t = 5 + 0.9\mu_{t-1} + v_t \ , \quad \{v_t\} \sim \text{NID}(0,0.6).$$

Condition A1 of the power-n theorem requires that $\{v_t\}_{t\in\mathbb{Z}}$ be an SE sequence. This condition holds trivially as the innovations $\{v_t\}$ are iid.

Condition A2, which requires the first iteration of the Markov process to have 3 moments, holds easily for any

μ_1 and $n = 3$ since

$$\begin{aligned}\mathbb{E}|\phi(\mu_1, v_t, \boldsymbol{\theta})|^3 &= \mathbb{E}|5 + 0.9\mu_1 + v_t|^3 \\ &\leq c \cdot \mathbb{E}5^3 + c \cdot \mathbb{E}(0.9|\mu_1|)^3 + c \cdot \mathbb{E}|v_t|^3 \\ &\quad (c_n\text{-inequality}) \\ &\leq c \cdot \underbrace{5^3}_{<\infty} + c \cdot \underbrace{0.9^3}_{<\infty} \underbrace{|\mu_1|^3}_{<\infty} + c \cdot \underbrace{\mathbb{E}|v_t|^3}_{<\infty} < \infty. \\ &\quad (\text{linearity of expectation})\end{aligned}$$

Note that $\mathbb{E}|v_t|^3 < \infty$ because v_t is Gaussian and hence has bounded moments of any order!

Condition A3 holds because $\frac{\partial\phi(\mu, v_t, \boldsymbol{\theta})}{\partial\mu} = 0.9$, and the contraction condition is satisfied

$$\mathbb{E}\sup_{\mu} |0.9|^3 = 0.9^3 = 0.81 < 1.$$

Since conditions A1-A3 are satisfied with $n = 3$, we can conclude by the power-n theorem that $\{\mu_t(\boldsymbol{\theta}, \mu_1)\}_{t\in\mathbb{N}}$ converges e.a.s. to a limit SE sequence $\{\mu_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ that satisfies $\mathbb{E}|\mu_t(\boldsymbol{\theta})|^2 < \infty$ for every initialization μ_1.[1]

Finally, we note that

$$\Big\{\ x_t(\boldsymbol{\theta}, \mu_1)\Big\}_{t\in\mathbb{N}} = \Big\{\mu_t(\boldsymbol{\theta}, \mu_1) + \varepsilon_t\ \Big\}_{t\in\mathbb{N}}$$

converges e.a.s. to the SE limit

$$\Big\{\ x_t(\boldsymbol{\theta})\ \Big\}_{t\in\mathbb{Z}} = \Big\{\ \mu_t(\boldsymbol{\theta}) + \varepsilon_t\ \Big\}_{t\in\mathbb{Z}} \text{ satisfying } \mathbb{E}|x_t(\boldsymbol{\theta})|^2 < \infty$$

for any $\mu_1 \in \mathbb{R}$ because

$$\begin{aligned}|x_t(\boldsymbol{\theta}, \mu_1) - x_t(\boldsymbol{\theta})| &= |\mu_t(\boldsymbol{\theta}, \mu_1) + \varepsilon_t - \mu_t(\boldsymbol{\theta}) - \varepsilon_t| \\ &= |\mu_t(\boldsymbol{\theta}, \mu_1) - \mu_t(\boldsymbol{\theta})| \overset{e.a.s.}{\to} 0\end{aligned}$$

[1] Actually, we have proven that $\mathbb{E}|\mu_t(\boldsymbol{\theta})|^n < \infty$ for any $n \leq 3$.

Furthermore, we conclude that $\mathbb{E}|x_t(\boldsymbol{\theta})|^2 < \infty$ since

$$\begin{aligned}\mathbb{E}|x_t(\boldsymbol{\theta})|^2 &= \mathbb{E}|\mu_t(\boldsymbol{\theta}) + \varepsilon_t|^2 \\ &\underbrace{\leq}_{c_n\text{-inequality}} c \cdot \mathbb{E}|\mu_t(\boldsymbol{\theta})|^2 + c \cdot \mathbb{E}|\varepsilon_t|^2 < \infty. \qquad \square\end{aligned}$$

the sequence $\{x_t(\boldsymbol{\theta}, \mu_1)\}_{t\in\mathbb{N}}$ with mean $\{\mu_t(\boldsymbol{\theta}, \mu_1)\}_{t\in\mathbb{N}}$ initialized at μ_1 converges e.a.s. to an SE limit $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ that satisfies $\mathbb{E}|x_t(\boldsymbol{\theta})|^2 < \infty$ for any $\mu_1 \in \mathbb{R}$.

4.3 (u) We cannot conclude that the sequence $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ generated by the NGARCH:

$$x_t = \sigma_t \varepsilon_t \ , \quad \{\varepsilon_t\} \sim \text{NID}(0,1) \ ,$$

$$\sigma_t^2 = 0.5 + 0.1(x_{t-1} - \sigma_{t-1})^2 + 0.9\sigma_{t-1}^2$$

is SE or has bounded first or second moments because the contraction conditions of the Bougerol and Power-n theorems do not hold for the volatility updating equation. In exercise 4.2 (u), we have already noted that the contraction of Bougerol does not hold, and since

$$\begin{aligned}\mathbb{E}\log\sup_x\left|\frac{\partial\phi(x,\epsilon_t,\boldsymbol{\theta})}{\partial x}\right| < 0 \quad &\Leftarrow \quad \mathbb{E}\sup_x\left|\frac{\partial\phi(x,\epsilon_t,\boldsymbol{\theta})}{\partial x}\right| < 1 \\ &\Leftarrow \quad \mathbb{E}\sup_x\left|\frac{\partial\phi(x,\epsilon_t,\boldsymbol{\theta})}{\partial x}\right|^2 < 1\end{aligned}$$

follows that

$$\begin{aligned}\mathbb{E}\log\sup_x\left|\frac{\partial\phi(x,\epsilon_t,\boldsymbol{\theta})}{\partial x}\right| \geq 0 \quad &\Rightarrow \quad \mathbb{E}\sup_x\left|\frac{\partial\phi(x,\epsilon_t,\boldsymbol{\theta})}{\partial x}\right| \geq 1 \\ &\Rightarrow \quad \mathbb{E}\sup_x\left|\frac{\partial\phi(x,\epsilon_t,\boldsymbol{\theta})}{\partial x}\right|^2 \geq 1.\end{aligned}$$

4.4 Those DGPs for which we can show that the Power-n Theorem holds for $n = 2$, are DGPs that satisfy this representation. The remaining DGPs may or may not satisfy that representation.

If the conditions of the Power-n Theorem hold for $n = 2$, then we are sure that the sequence $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is SE and has bounded moments of second order. As a result, we now that $\{x_t\}_{t\in\mathbb{Z}}$ is weakly stationary. Finally, if $\{x_t\}_{t\in\mathbb{Z}}$ is weakly then it satisfies Wold's representation.

4.6 Note first that the word 'degenerate' typically refers to a parameter space Θ which has Lebesgue measure i.e. a parameter space which consists of a single point in at least one dimension. The quadratic AR(1)

$$x_t = \beta x_{t-1}^2 + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma^2)$$

satisfies Bougerol's contraction only for values of $\boldsymbol{\theta} = (\beta, \sigma^2)$ on a degenerate $\Theta \subseteq \mathbb{R}$ characterized by $\beta = 0$.

$$\mathbb{E} \log \sup_x \left| \frac{\partial \phi(x, \varepsilon_t, \boldsymbol{\theta})}{\partial x} \right| < 0 \quad \Leftrightarrow \quad \log \sup_x |2\beta x| < 0.$$

This can only hold true for $\beta = 0$ (assuming that $\log(0) = -\infty$).

4.7 The quadratic AR(1)

$$x_t = \alpha + x_{t-1}^2 + \varepsilon_t \quad \varepsilon_t \sim \text{NID}(0, \sigma^2)$$

never satisfies Bougerol's contraction because

$$\mathbb{E} \log \sup_x \left| \frac{\partial \phi(x, \varepsilon_t, \boldsymbol{\theta})}{\partial x} \right| < 0 \quad \Leftrightarrow \quad \log \sup_x |2x| < 0.$$

does not hold true for any $(\alpha, \sigma^2) = \boldsymbol{\theta} \in \Theta$.

4.8 The AR(1)

$$x_t = \beta x_{t-1} + \epsilon_t \quad \epsilon_t \sim \text{Log-Cauchy}(\mu, \sigma)$$

never satisfies Bougerol's initial moment condition because

$$\mathbb{E} \log^+ \left| \phi(x_1, \epsilon_t, \boldsymbol{\theta}) \right| = \mathbb{E} \log^+ \left| \beta x_1 + \epsilon_t \right| < \infty$$

does not hold true for any (β, μ, σ).

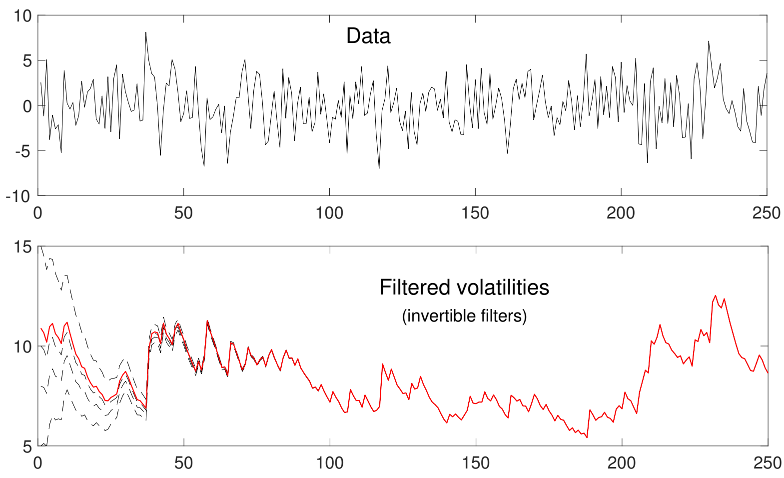
4.10 (b) We can show that $\{x_t\}_{t\in\mathbb{Z}}$ is SE, by first showing that $\{\theta_t\}_{t\in\mathbb{Z}}$ is SE, using Bougerol's Theorem, and then showing that $\{x_t\}_{t\in\mathbb{Z}}$ is SE by application of Krengel's Theorem. Try it yourself!

Exercises of Chapter 5

5.1 You can verify filter invertibility for any of the models mentioned in this question by applying Bougerol's theorem. The question on uniform invertibility is only sensible for the model in (b), since *uniform invertibility* is about obtaining invertibility uniformly over the parameter space, and the model in (b) is the only that allows for multiple parameter values.

(b) This filter is invertible as it satisfies the conditions in p.140 and p.141. You can verify this by applying Bougerol's Theorem yourself! Uniform invertibility fails on the parameter space as it is defined here. But it holds on any compact parameter space Θ contained inside the specified interval $\omega > 0$, $\alpha > 0$, $|\beta| < 1$. Verify this!

5.2 Yes, we find evidence of filter invertibility. This is evident by the fact that the filtered volatilities converge to a unique path regardless of their initialization; see the figure below.


Data
Filtered volatilities
(invertible filters)
10
5
0
-5
-10
0
50
100
150
200
250
15
10
5

Exercises of Chapter 6

6.1 (a) The conditions of Theorem 6.3 fail to hold because Θ is not compact. Hence, we have no reason to suppose that a maximum exists. For this reason, we cannot ensure that the estimator exists; i.e. the argmax set may be empty.

6.1 (c) The least squares criterion is naturally given by

$$Q_T(\mathbf{X_T}, \boldsymbol{\theta}) = -\frac{1}{T}\sum_{t=2}^{T}(x_t - g(x_{t-1}; \boldsymbol{\theta})x_{t-1})^2$$

6.1 (d) Conditions of Theorem 6.3 hold because the criterion is continuous in ϕ and the data x_t, x_{t-1}, and the parameter space is compact. The estimator exists and is a random variable (measurable).

6.2 (a) The least squares criterion is given by

$$Q_T(\mathbf{X_T}, \boldsymbol{\theta}) = -\frac{1}{T}\sum_{t=2}^{T}(x_t - \alpha - \beta\cos(x_{t-1}))^2.$$

6.3 (a) Since the density $f(\epsilon_t, \lambda)$ of a Student's-t distribution with λ degrees of freedom is given by

$$f(\epsilon_t) = \frac{\Gamma(\frac{\lambda+1}{2})}{\sqrt{\lambda\pi}\Gamma(\frac{\lambda}{2})}\Big(1 + \frac{\epsilon_t^2}{\lambda}\Big)^{-\frac{\lambda+1}{2}} ,$$

we have that the density of x_t <u>conditional</u> on x_{t-1} is given by

$$f\big(x_t - \alpha - \beta\cos(x_{t-1})\big)$$
$$= \frac{\Gamma(\frac{\lambda+1}{2})}{\sqrt{\lambda\pi}\Gamma(\frac{\lambda}{2})}\left[1 + \frac{(x_t - \alpha - \beta\cos(x_{t-1}))^2}{\lambda}\right]^{-\frac{\lambda+1}{2}}$$

As a result, the log likelihood function is given by

$$L_T(\mathbf{X_T}, \boldsymbol{\theta}) =$$

$$\frac{1}{T}\sum_{t=2}^{T} A(\lambda) - \frac{\lambda+1}{2}\log\left[1 + \frac{(x_t - \alpha - \beta\cos(x_{t-1}))^2}{\lambda}\right]$$

$$\text{where} \quad A(\lambda) = \log\Gamma(\frac{\lambda+1}{2}) - \log\sqrt{\lambda\pi} - \log\Gamma(\frac{\lambda}{2})$$

This is a natural result of the theorem known by the curious name *transformation of a random variable*,

Theorem 0.1 *Let z_t be a random variable with pdf f_z and x_t be a random variable given by $x_t = g(z_t)$ for some strictly increasing and differentiable function g. Then the pdf f_x of x_t is given by*

$$f_x(x_t) = f_z\big(g^{-1}(x_t)\big)\frac{\partial g^{-1}(x_t)}{\partial x}.$$

6.3 (c) It is clear that

$$x_t|x_{t-1} \;\sim\; N\Big(g(x_{t-1};\boldsymbol{\theta})x_{t-1}\;,\;2\Big)$$

and hence that x_t has conditional density given by

$$f(x_t|x_{t-1}, \boldsymbol{\theta}) = \frac{1}{\sqrt{4\pi}}\exp\left[\frac{-\big(x_t - g(x_{t-1};\boldsymbol{\theta})x_{t-1}\big)^2}{4}\right].$$

As a result, since we can factorize the joint density of the data $x_1, ..., x_T$ as

$$f(x_1, ..., x_T, \boldsymbol{\theta}) = f(x_1, \boldsymbol{\theta}) \cdot \prod_{t=2}^{T} f(x_t|x_{t-1}, \boldsymbol{\theta})$$

we obtain a log likelihood function given by

$$L(\mathbf{x_T}, \boldsymbol{\theta}) = \frac{1}{T}\sum_{t=2}^{T} -\frac{1}{2}\log 4\pi - \frac{(x_t - g(x_{t-1};\boldsymbol{\theta})x_{t-1})^2}{4}.$$

Note that an ML estimator with exact initialization assigns some distribution to x_1, while a conditional ML estimator starts at $t = 2$ and conditions on value of x_1.

6.3 (d) Since $\mu_t(\boldsymbol{\theta}, \mu_1)$ is given <u>conditional</u> on the past data $D_{t-1} = x_{t-1}, x_{t-2}, ...$ we have that $x_t|D_{t-1} \sim N(\mu_t, \sigma_\varepsilon^2)$

$$f(x_t|D_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left[\frac{-\big(x_t - \mu_t(\boldsymbol{\theta}, \mu_1)\big)^2}{2\sigma_\varepsilon^2}\right].$$

and hence

$$L(\mathbf{x_T}, \boldsymbol{\theta}) = \frac{1}{T}\sum_{t=2}^{T} -\frac{1}{2}\log 2\pi\sigma_\varepsilon^2 - \frac{\big(x_t - \mu_t(\boldsymbol{\theta}, \mu_1)\big)^2}{2\sigma_\varepsilon^2}.$$

6.3 (g) Since $\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)$ is given <u>conditional</u> on the past data $D_{t-1} = x_{t-1}, x_{t-2}, ...$ we have that $x_t|D_{t-1} \sim N(0, \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2))$

$$f(x_t|D_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)}} \exp\left[\frac{-x_t^2}{2\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)}\right].$$

and hence

$$L(\mathbf{x_T}, \boldsymbol{\theta}) = \frac{1}{T}\sum_{t=2}^{T} -\frac{1}{2}\log 2\pi\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2) - \frac{x_t^2}{2\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)}.$$

Exercises of Chapter 7

7.5 (a) Since (i) Θ is compact; (ii) $\{\beta x_t\}$ is SE (by Krengel's theorem); and (iii) $\mathbb{E}|\beta x_t| = |\beta|\mathbb{E}|x_t| < \infty$ for every $\beta \in [0, 100]$; it follows by an LLN for SE sequences that the criterion converges pointwise as $T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^{T} \beta x_t \xrightarrow{p} \mathbb{E}\beta x_t \quad \text{for all } \beta \in [0, 100].$$

Furthermore, since $\{Q_T(\beta)\}$ is stochastically equicontinuous

$$\mathbb{E} \sup_{\beta \in [0,100]} \left| \frac{\partial \beta x_t}{\partial \beta} \right| = \mathbb{E}|x_t| < \infty,$$

we have, by Theorems 7.5 and 7.6, that Q_T converges uniformly in probably over $[0, 100]$ to the limit criterion $Q_\infty(\beta) = \mathbb{E}\beta x_t$,

$$\sup_{\beta \in [0,100]} \left| \frac{1}{T}\sum_{t=1}^{T} \beta x_t - \mathbb{E}\beta x_t \right| \xrightarrow{p} 0 \quad \text{as} \quad T \to \infty.$$

Note that the limit criterion $Q_\infty(\beta) = \mathbb{E}\beta x_t$ is a deterministic function of β. It is not a function of x_t because the expectation 'integrates x_t out'.

7.5 (b) We cannot show that Q_T converges uniformly over Θ to the limit Q_∞ using a pointwise LLN for SE sequences and stochastic equicontinuity, because $\mathbb{E}|x_t| < \infty$ is not enough to ensure that $\mathbb{E}|x_t - \beta x_{t-1}|^2 < \infty$. In particular, $\mathbb{E}|x_t - \beta x_{t-1}|^2 < \infty$ is needed for the pointwise LLN to hold

$$\frac{1}{T}\sum_{t=1}^{T}(x_t - \beta x_{t-1})^2 \xrightarrow{p} \mathbb{E}(x_t - \beta x_{t-1})^2 \quad \forall\, \boldsymbol{\theta} \in \Theta.$$

$\mathbb{E}|x_t| < \infty$ is not enough to ensure that $\mathbb{E}|x_t - \beta x_{t-1}|^2 < \infty$. $\mathbb{E}|x_t|^2 < \infty$ would be sufficient since by the c_n-inequality:

$$\mathbb{E}|x_t - \beta x_{t-1}|^2 \leq c\mathbb{E}|x_t|^2 + c\beta^2\mathbb{E}|x_{t-1}|^2 < \infty$$

and by SE we have $\mathbb{E}|x_t|^2 = \mathbb{E}|x_{t-1}|^2$.

7.5 (g) We cannot show that Q_T converges uniformly over Θ to the limit Q_∞ because the parameter space is not closed, and hence also not compact.

7.6 (a) We can show that the least squares estimator $\hat{\boldsymbol{\theta}}_T$ is consistent for $\boldsymbol{\theta}_0 \in \Theta$ as long as Θ is compact and $\boldsymbol{\theta}_0$ is the unique maximizer of

$$Q_\infty(\boldsymbol{\theta}) = \mathbb{E} - (x_t - \alpha - \beta\cos(x_{t-1}))^2.$$

First, we note that the least squares criterion function takes the form

$$Q_T(\mathbf{X_T}, \boldsymbol{\theta}) \;=\; -\frac{1}{T}\sum_{t=2}^{T}(x_t - \alpha - \beta\cos(x_{t-1}))^2.$$

Now, since the data satisfies $\mathbb{E}|x_t|^2 < \infty$, we have by the c_n-inequality that,

$$\begin{aligned}&\mathbb{E}|x_t - \alpha - \beta\cos(x_{t-1})|^2 \\ &\quad\leq\; c\underbrace{\mathbb{E}|x_t|^2}_{<\infty} + c\underbrace{|\alpha|^2}_{<\infty} + c\underbrace{|\beta|^2}_{<\infty}\underbrace{\mathbb{E}|\cos(x_{t-1})|^2}_{\leq 1 < \infty} < \infty.\end{aligned}$$

Additionally, since the data $\{x_t\}$ is SE, then $\{(x_t - \alpha - \beta\cos(x_{t-1}))^2\}$ is also SE by Krengel's theorem. Hence, we obtain the pointwise LLN $\forall\, \boldsymbol{\theta} \in \Theta$

$$\frac{1}{T}\sum_{t=2}^{T}\Big(x_t - \alpha - \beta\cos(x_{t-1})\Big)^2 \;\overset{p}{\to}\; \mathbb{E}\Big(x_t - \alpha - \beta\cos(x_{t-1})\Big)^2.$$

Furthermore, we note that $\{Q_T\}$ is stochastically equicontinuous because

$$
\begin{aligned}
\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{\partial q(x_t,x_{t-1},\boldsymbol{\theta})}{\partial\alpha}\Big| &= \\
&\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|-2x_t+2\alpha+2\beta\cos(x_{t-1})| \\
&\le\ 2\mathbb{E}|x_t|+2\sup_{\boldsymbol{\theta}\in\Theta}|\alpha|+2\sup_{\boldsymbol{\theta}\in\Theta}|\beta|\mathbb{E}|\cos(x_{t-1})| \\
&\text{(norm sub-additivity)} \\
&\le\ 2\mathbb{E}|x_t|+2\bar{\alpha}+2\bar{\beta}\mathbb{E}|\cos(x_{t-1})| \\
&\text{(compact } \Theta) \\
&\le\ 2\mathbb{E}|x_t|+2\bar{\alpha}+2\bar{\beta}\ <\ \infty \\
&(\cos(z)<1\ \forall\ z)
\end{aligned}
$$

Note that if Θ is compact (bounded and closed), then there must exists some maximum value $\bar{\alpha}$ and $\bar{\beta}$ such that

$$
\sup_{\boldsymbol{\theta}\in\Theta}|\alpha|\le\bar{\alpha}<\infty \quad \text{and} \quad \sup_{\boldsymbol{\theta}\in\Theta}|\beta|\le\bar{\beta}<\infty.
$$

The derivative with respect to β is also bounded in expectation,

$$
\begin{aligned}
\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{\partial q(x_t,x_{t-1},\boldsymbol{\theta})}{\partial\beta}\Big| &= \\
&\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\big|-2x_t\cos(x_{t-1})+2\alpha\cos(x_{t-1})+2\beta\cos(x_{t-1})^2\big| \\
&\le\ 2\mathbb{E}|x_t\cos(x_{t-1})|+2\sup_{\boldsymbol{\theta}\in\Theta}|\alpha|\mathbb{E}|\cos(x_{t-1})| \\
&\qquad +2\sup_{\boldsymbol{\theta}\in\Theta}|\beta|\mathbb{E}|\cos(x_{t-1})|^2 \\
&\text{(norm sub-additivity)} \\
&\le\ 2\mathbb{E}|x_t\cos(x_{t-1})|+2\bar{\alpha}\mathbb{E}|\cos(x_{t-1})|+2\bar{\beta}\mathbb{E}|\cos(x_{t-1})|^2 \\
&\text{(compact } \Theta) \\
&\le\ 2\mathbb{E}|x_t|+2\bar{\alpha}+2\bar{\beta}\ <\ \infty \\
&(\cos(z)<1\ \forall\ z)
\end{aligned}
$$

Note that when $\boldsymbol{\theta}$ is a vector $\boldsymbol{\theta}=(\theta_1,...,\theta_n)$ then the vector condition for stochastic equicontinuity

$$
\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\Big\|\frac{\partial q(x_t,x_{t-1},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big\|<\infty
$$

holds if, the condition also holds for each individual partial derivative

$$\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{\partial q(x_t, x_{t-1}, \boldsymbol{\theta})}{\partial \theta_i} \right| < \infty \quad \text{for} \quad i = 1, ..., n.$$

and hence, the criterion converges uniformly to its limit

$$\sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{T} \sum_{t=2}^{T} \Big(x_t - \alpha - \beta \cos(x_{t-1}) \Big)^2 \right.$$
$$\left. - \quad \mathbb{E}\Big(x_t - \alpha - \beta \cos(x_{t-1}) \Big)^2 \right| \xrightarrow{p} 0.$$

Finally, if $\boldsymbol{\theta}_0$ is the unique maximizer of

$$Q_\infty(\boldsymbol{\theta}) = \mathbb{E} - (x_t - \alpha - \beta \cos(x_{t-1}))^2.$$

then it is also *identifiably unique* because Q_∞ is continuous on the compact Θ. In particular, continuity of Q_∞ follows by continuity of Q_T on Θ for all T, and uniform convergence to Q_∞; see Lemma 7.2 in Chapter 7. Note that we can either *argue for* or *assume* uniqueness (depending on model specification!). If the model is correctly specified, then given the LS criterion, uniqueness is implied by identification of $\boldsymbol{\theta}_0$ in the sense that each $\boldsymbol{\theta}$ defines a unique regression function; see Theorem 7.12. If the model is misspecified, then uniqueness may hold by assumption. When it fails, then it is easy to show convergence to the set of optimizers Θ_0 of the limit criterion; see Theorem 7.15.

As a result, we conclude by the classical consistency theorem for M-estimators, that

$$\hat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0 \quad \text{as} \quad T \to \infty. \quad \square$$

7.6 (b) We can show that the LS estimator $\hat{\boldsymbol{\theta}}_T$ is consistent for $\boldsymbol{\theta}_0 \in \Theta$ as long as Θ is compact and $\boldsymbol{\theta}_0$ is the unique maximizer of

$$Q_\infty(\boldsymbol{\theta}) = \mathbb{E} - (x_t - g(x_{t-1}; \boldsymbol{\theta}) x_{t-1})^2.$$

Indeed, first we note the least squares criterion function takes the form

$$Q_T(\mathbf{X_T}, \boldsymbol{\theta}) = -\frac{1}{T}\sum_{t=2}^{T}(x_t - g(x_{t-1}; \boldsymbol{\theta})x_{t-1})^2$$

Now, since the data is SE, $\{(x_t - g(x_{t-1}; \boldsymbol{\theta})x_{t-1})^2\}$ is also SE by Krengel's Theorem. Furthermore, since the data satisfies $\mathbb{E}|x_t|^2 < \infty$ we have by the c_n-inequality that

$$\begin{aligned}
\mathbb{E}|x_t - g(x_{t-1}; \boldsymbol{\theta})x_{t-1}|^2 &\leq c\mathbb{E}|x_t|^2 + c\mathbb{E}|g(x_{t-1}; \boldsymbol{\theta})x_{t-1}|^2 \\
\text{(because } |g(x_{t-1}, \boldsymbol{\theta})| < |\gamma|) \quad &\leq c\underbrace{\mathbb{E}|x_t|^2}_{<\infty} + c|\gamma|^2 \underbrace{\mathbb{E}|x_{t-1}|^2}_{<\infty} < \infty
\end{aligned}$$

As a result, we obtain the following pointwise LLN $\forall\, \boldsymbol{\theta} \in \Theta$ as $T \to \infty$

$$\frac{1}{T}\sum_{t=2}^{T}\Big(x_t - g(x_{t-1}; \boldsymbol{\theta})x_{t-1}\Big)^2 \xrightarrow{p} \mathbb{E}\Big(x_t - g(x_{t-1}; \boldsymbol{\theta})x_{t-1}\Big)^2 .$$

We note further that $\{Q_T\}$ is stochastically equicontinuous because

$$\mathbb{E}|q(x_t, x_{t-1}, \boldsymbol{\theta})| = \mathbb{E}|x_t - g(x_{t-1}; \boldsymbol{\theta})x_{t-1}|^2 < \infty ,$$

and since q is continuously differentiable and well behaved of order 1, and Θ is compact, we have[2]

$$\mathbb{E}|q(x_t, x_{t-1}, \boldsymbol{\theta})| < \infty \Rightarrow \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial q(x_t, x_{t-1}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\| < \infty.$$

Hence the criterion converges uniformly to its limit

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\Theta}\Bigg|\frac{1}{T}\sum_{t=2}^{T}&\Big(x_t - g(x_{t-1}; \boldsymbol{\theta})x_{t-1}\Big)^2 \\
&- \mathbb{E}\Big(x_t - g(x_{t-1}; \boldsymbol{\theta})x_{t-1}\Big)^2\Bigg| \xrightarrow{p} 0.
\end{aligned}$$

Finally, $\boldsymbol{\theta}_0$ is the identifiably unique maximizer of the limit criterion

$$Q_\infty(\boldsymbol{\theta}) = \mathbb{E} - (x_t - g(x_{t-1}; \boldsymbol{\theta})x_{t-1})^2$$

since

[2] We assume that q is well behaved!

(a) each $\boldsymbol{\theta}$ defines a unique function $g(x;\boldsymbol{\theta})x$, and hence $\boldsymbol{\theta}_0$ is the unique maximizer (see Theorem 7.10 on the uniqueness for least squares);

(b) Θ is compact (assumed);

(c) Q_∞ is continuous since Q_T is continuous on Θ for every T and it converges uniformly over Θ to Q_∞.

As a result, we conclude by the classical consistency theorem for M-estimators, that

$$\hat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0 \quad \text{as} \;\; T \to \infty. \quad \square$$

Note that we could have shown directly that Q_T is stochastically equicontinuous by taking the derivatives explicitly. For convenience, let us define $w(x_{t-1},\beta) := 1/(1+\exp(\beta x_{t-1}))$. We thus have $g(x_{t-1},\boldsymbol{\theta}) = \gamma w(x_{t-1},\beta)$. The stochastic equicontinuity condition is then obtained by noting that

$$\begin{aligned}
&\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial q(x_t,x_{t-1},\boldsymbol{\theta})}{\partial\gamma}\right| \\
&\quad = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|(x_t - g(x_{t-1};\boldsymbol{\theta})x_{t-1})w(x_{t-1};\beta)x_{t-1}| \\
&\quad \le \mathbb{E}|x_t x_{t-1}|\sup_{\boldsymbol{\theta}\in\Theta}|w(x_{t-1};\beta)| \\
&\qquad + \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|g(x_{t-1};\boldsymbol{\theta})w(x_{t-1};\beta)||x_{t-1}|^2 \\
&\quad (\sup_{X,\beta}|w(x;\beta)| < 1) \\
&\quad \le \mathbb{E}|x_t x_{t-1}| + \bar{\gamma}\mathbb{E}|x_{t-1}|^2 \\
&\quad (\sup_{x,\boldsymbol{\theta}}|g(x,\boldsymbol{\theta})| \le \sup_{\gamma\in\Theta}|\gamma| \equiv \bar{\gamma}) \\
&\quad \le \mathbb{E}|x_{t-1}|^2 \; < \; \infty \\
&\quad (\text{because } \mathbb{E}|x_t x_{t-1}| \le \mathbb{E}|x_{t-1}|^2)
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{\partial q(x_t,x_{t-1},\boldsymbol{\theta})}{\partial\beta}\Big| \\
&\qquad = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\Big|\big(x_t - g(x_{t-1};\boldsymbol{\theta})x_{t-1}\big)\gamma\frac{\partial w(x_{t-1};\beta)}{\partial\beta}x_{t-1}\Big| \\
&\qquad \leq \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\gamma|\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{\partial w(x_{t-1};\beta)}{\partial\beta}x_{t-1}\Big||x_t| \\
&\qquad\qquad + \sup_{\boldsymbol{\theta}\in\Theta}|\gamma|\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{\partial w(x_{t-1};)}{\partial\beta}x_{t-1}\Big| \\
&\qquad \leq \bar{\gamma}k\mathbb{E}|x_t| + \bar{\gamma}k \;<\; \infty \;,
\end{aligned}$$

because there exists a finite k such that

$$\begin{aligned}
\sup_{x,\beta}\frac{\partial w(x;\beta)}{\partial\beta}x &= \frac{\exp(\beta x)x}{(1+\exp(\beta x))^2}x \\
&= \frac{\exp(\beta x)x^2}{(1+\exp(\beta x))^2} \leq k < \infty.
\end{aligned}$$

7.7 (a) If the model is well specified, then the observed data was generated by the model under some $\boldsymbol{\theta}_0 \in \Theta$. Hence, we could, for example, use the Power-n Theorem to ensure that the model generates SE paths with bounded second moment for every $\boldsymbol{\theta} \in \Theta$. For our model,

$$x_t = \alpha + \beta\cos(x_{t-1}) + \varepsilon_t \;, \quad \{\varepsilon_t\}_{t\in\mathbb{Z}} \sim TID(7).$$

we know that $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is SE and satisfies $\mathbb{E}|x_t(\boldsymbol{\theta})|^2 < \infty$ for every $\boldsymbol{\theta} \in \Theta$ as long as

A1. $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an n_ε-variate iid sequence (trivially satisfied!)

A2. $\mathbb{E}|\phi(x_1,\boldsymbol{\varepsilon}_t)|^2 \;\leq\; c|\alpha|^2 + c|\beta|^2|\cos(x_1)|^2 + c\mathbb{E}|\epsilon_t|^2 < \infty$

- So A2 requires no restriction on (α,β), any $\Theta \subseteq \mathbb{R}^2$.

A3. $\mathbb{E}\sup_{x\in\mathcal{X}}\big|\partial\phi(x,\boldsymbol{\varepsilon}_t)/\partial x\big|^2 \;=\; \mathbb{E}\sup_{x\in\mathcal{X}}\big|-\beta\sin(x)\big|^2$ $= \; |\beta|^2\sup_{x\in\mathcal{X}}|\sin(x)|^2 \;=\; |\beta|^2 < 1 \;\Leftrightarrow\; |\beta| < 1.$

- So A3 requires that Θ be subset of $\mathbb{R}\times(-1,1)$.

7.8 (a) By the classical consistency theorem for M-estimators, the ML estimator $\hat{\boldsymbol{\theta}}_T$ is consistent for $\boldsymbol{\theta}_0 \in \Theta$ if Θ is compact with $\lambda > 0$ and $\boldsymbol{\theta}_0$ is the unique maximizer of

$$Q_\infty(\boldsymbol{\theta}) = A(\lambda) + B(\lambda)\mathbb{E}\Big[\log w(\boldsymbol{\theta}, x_t, x_{t-1})\Big]$$

$$\text{where} \quad A(\lambda) = \log\Gamma(\frac{\lambda+1}{2}) - \log\sqrt{\lambda\pi} - \log\Gamma(\frac{\lambda}{2})$$

$$B(\lambda) = -\frac{\lambda+1}{2} \quad \text{and}$$

$$w(\boldsymbol{\theta}, x_t, x_{t-1}) = \Big(1 + \frac{(x_t - \alpha - \beta\cos(x_{t-1}))^2}{\lambda}\Big)$$

Indeed, we have already seen in the previous exercise that

$$L_T(\mathbf{x_T}, \boldsymbol{\theta}) = \frac{1}{T}\sum_{t=2}^{T} A(\lambda) + B(\lambda)\log w(\boldsymbol{\theta}, x_t, x_{t-1})$$

Now, it is clear that just need to focus on the uniform convergence of the term $h_t(\boldsymbol{\theta}) := \log w(\boldsymbol{\theta}, x_t, x_{t-1})$ because

$$\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{1}{T}\sum_{t=2}^{T} A(\lambda) + B(\lambda)h_t(\boldsymbol{\theta}) \ - \ A(\lambda) - B(\lambda)\mathbb{E}h_t(\boldsymbol{\theta})\Big| \xrightarrow{p} 0$$

$$\Leftrightarrow \quad \sup_{\boldsymbol{\theta}\in\Theta}|B(\lambda)| \sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{1}{T}\sum_{t=2}^{T} h_t(\boldsymbol{\theta}) \ - \ \mathbb{E}h_t(\boldsymbol{\theta})\Big| \xrightarrow{p} 0$$

$$\Leftrightarrow \quad \bar{B}\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{1}{T}\sum_{t=2}^{T} h_t(\boldsymbol{\theta}) \ - \ \mathbb{E}h_t(\boldsymbol{\theta})\Big| \xrightarrow{p} 0$$

$$\text{for any } \bar{B} = \frac{\bar{\lambda}+1}{2} < \infty.$$

Recall that if z_t is a random variable satisfying $0 < a \leq z_t$ then $\mathbb{E}|\log z_t| < \infty$ holds as long as $\mathbb{E}|z_t|^n < \infty$ for some $n > 0$.

The pointwise convergence of $h_t(\boldsymbol{\theta}) := \log w(\boldsymbol{\theta}, x_t, x_{t-1})$

$$\frac{1}{T}\sum_{t=2}^{T} h_t(\boldsymbol{\theta}) \xrightarrow{p} \mathbb{E}h_t(\boldsymbol{\theta}) \quad \text{holds for all} \ \ \boldsymbol{\theta} \in \Theta \ \ \text{because}$$

(a) The data is SE and hence $\{\log w(\boldsymbol{\theta}, x_t, x_{t-1})\}$ is also SE for every $\boldsymbol{\theta} \in \Theta$ (Krengel's theorem)

(b) $\mathbb{E}|\log w(\boldsymbol{\theta}, x_t, x_{t-1})| < \infty$ for every $\boldsymbol{\theta} \in \Theta$ because $w(\boldsymbol{\theta}, x_t, x_{t-1}) > 1$ and $|\alpha + \beta\cos(x_{t-1})| < |\alpha| + |\beta| < k < \infty$

$$\begin{aligned}\mathbb{E}|w(\boldsymbol{\theta},x_t, x_{t-1})| &= \mathbb{E}\Big|1 + (x_t - \alpha - \beta\cos(x_{t-1}))^2/\lambda\Big| \\ &\leq 1 + \lambda^{-1}\mathbb{E}|x_t|^2 + \lambda^{-1}k^2 + \lambda^{-1}k\mathbb{E}|x_t| \ < \ \infty\end{aligned}$$

Hence, assuming that the criterion is well behaved of order 1, it is also stochastically equicontinuous and converges uniformly to its limit

$$\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{1}{T}\sum_{t=2}^{T} A(\lambda) + B(\lambda)h_t(\boldsymbol{\theta}) \ - \ A(\lambda) - B(\lambda)\mathbb{E}h_t(\boldsymbol{\theta})\Big| \xrightarrow{p} 0$$

Finally, if $\boldsymbol{\theta}_0$ is the unique maximizer of

$$Q_\infty(\boldsymbol{\theta}) = A(\lambda) + B(\lambda)\mathbb{E}\Big[\log w(\boldsymbol{\theta}, x_t, x_{t-1})\Big]$$

then it is also *identifiably unique* because Q_∞ is continuous on the compact Θ; continuity of Q_∞ follows by continuity of Q_T on Θ for all T, and uniform convergence to Q_∞. Naturally, if the model is correctly specified then uniqueness of $\boldsymbol{\theta}_0$ is implied by identification of $\boldsymbol{\theta}_0$ and the information inequality theorem for ML estimators (see information inequality in Chapter 7). In contrast, if the model is misspecified, then uniqueness holds by assumption. As a result, by the classical consistency theorem, we have

$$\hat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0 \quad \text{as } \ T \to \infty. \quad \square$$

7.8 (b) By the classical consistency theorem for M-estimators, the ML estimator $\hat{\boldsymbol{\theta}}_T$ is consistent for $\boldsymbol{\theta}_0 \in \Theta$ as $T \to \infty$ if

(i) $\hat{\boldsymbol{\theta}}_T$ exists and is measurable;

(ii) The log likelihood L_T converges uniformly to the limit L_∞

$$\sup_{\boldsymbol{\theta}\in\Theta}\big|L_T(\mathbf{x_T}, \boldsymbol{\theta}) - L_\infty(\boldsymbol{\theta})\big| \xrightarrow{p} 0 \quad \text{as} \quad T \to \infty \ ;$$

(iii) $\boldsymbol{\theta}_0 \in \Theta$ is the identifiably unique maximizer of L_∞

$$\sup_{\boldsymbol{\theta} \in S^c(\boldsymbol{\theta}_0, \delta)} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0).$$

Note first that, in this case, $x_t | x_{t-1} \sim N\Big(g(x_{t-1}; \boldsymbol{\theta}) x_{t-1}, \sigma_\varepsilon^2\Big)$, the likelihood function is given by,

$$\begin{aligned} L(\mathbf{x_T}, \boldsymbol{\theta}) &= \frac{1}{T} \sum_{t=2}^{T} \ell(x_t, x_{t-1}, \boldsymbol{\theta}) \\ &= \frac{1}{T} - \sum_{t=2}^{T} \frac{1}{2} \log 2\pi\sigma_\varepsilon^2 - \sum_{t=2}^{T} \frac{\big(x_t - g(x_{t-1}; \boldsymbol{\theta}) x_{t-1}\big)^2}{2\sigma_\varepsilon^2}. \end{aligned}$$

We obtain the existence and measurability of $\hat{\boldsymbol{\theta}}_T$ immediately from the continuity of ℓ and the compactness of Θ (which is assumed).

We obtain the uniform convergence in (ii) from a pointwise LLN, a stochastic equicontinuity condition, and the compactness of Θ. These conditions which are stated below as conditions (ii.i), (ii.ii) and (ii.iii) respectively:

(ii.i) *Pointwise convergence of the log likelihood*

$$\frac{1}{T} \sum_{t=2}^{T} \ell(x_t, x_{t-1}, \boldsymbol{\theta}) \xrightarrow{p} \mathbb{E}\ell(x_t, x_{t-1}, \boldsymbol{\theta}) \quad \text{as}\, T \to \infty\, \forall\, \boldsymbol{\theta} \in \Theta.$$

The pointwise LLN holds by application of a LLN for every $\boldsymbol{\theta} \in \Theta$. The LLN holds since:

(a) $\{\ell(x_t, x_{t-1}, \boldsymbol{\theta})\}$ is SE for every $\boldsymbol{\theta} \in \Theta$. This follows by Krengel's theorem because $\{x_t\}_{t \in \mathbb{Z}}$ is SE and ℓ is continuous (hence Borel measurable).

(b) $\ell(x_t, x_{t-1}, \boldsymbol{\theta})$ has one bounded moment $\mathbb{E}|\ell(x_t, x_{t-1}, \boldsymbol{\theta})| < \infty$ for every $\boldsymbol{\theta} \in \Theta$, Indeed,

$$
\begin{aligned}
\mathbb{E}|\ell(x_t, x_{t-1}, \boldsymbol{\theta})| &= \mathbb{E}\bigg| -\frac{1}{2}\log 2\pi\sigma_\varepsilon^2 - \frac{\big(x_t - g(x_{t-1};\boldsymbol{\theta})x_{t-1}\big)^2}{2\sigma_\varepsilon^2}\bigg| \\
&\quad \text{(by definition of } \ell(x_t, x_{t-1}, \boldsymbol{\theta})\text{)} \\
&\leq \mathbb{E}\bigg|\frac{1}{2}\log 2\pi\sigma_\varepsilon^2\bigg| + \mathbb{E}\bigg|\frac{\big(x_t - g(x_{t-1};\boldsymbol{\theta})x_{t-1}\big)^2}{2\sigma_\varepsilon^2}\bigg| \\
&\quad \text{(norm-subadditivity)} \\
&\leq |\frac{1}{2}\log 2\pi\sigma_\varepsilon^2| + \frac{1}{2\sigma_\varepsilon^2}\mathbb{E}\big|x_t - g(x_{t-1};\boldsymbol{\theta})x_{t-1}\big|^2 \\
&\leq |\frac{1}{2}\log 2\pi\sigma_\varepsilon^2| + \frac{c}{2\sigma_\varepsilon^2}\mathbb{E}|x_t|^2 \\
&\qquad + \frac{c}{2\sigma_\varepsilon^2}\sup_x |g(x;\boldsymbol{\theta})|\mathbb{E}|x_{t-1}|^2 \\
&\quad \text{(} c_n\text{-inequality)} \\
&< \infty
\end{aligned}
$$

where $|\frac{1}{2}\log 2\pi\sigma_\varepsilon^2| < \infty$ because $0 < \sigma_\varepsilon^2 < \infty$, furthermore $\frac{c}{2\sigma_\varepsilon^2} < \infty$ because $\sigma_\varepsilon^2 > 0$ and the c_n-inequality holds for some $c < \infty$, finally, $\mathbb{E}|x_t|^2 = \mathbb{E}|x_{t-1}|^2 < \infty$ because $\{x_t\}$ is SE with two bounded moments, and $\sup_x |g(x;\boldsymbol{\theta})| < \infty$ because g is uniformly bounded.

(ii.ii) *Stochastic equicontinuity of log likelihood*

$$
\sup_{\boldsymbol{\theta}\in\Theta}\bigg\|\frac{\partial\ell(x_t, x_{t-1}, \boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg\| < \infty,
$$

This stochastic equicontinuity condition is implied by the moment condition $\mathbb{E}|\ell(x_t, x_{t-1}, \boldsymbol{\theta})| < \infty$ established above and the fact that ℓ is well behaved of order 1 (assumed).

(ii.iii) *Compactness of* Θ. This condition holds by assumption and is easy to impose in practice.

Finally, we obtain the identifiable uniqueness of $\boldsymbol{\theta}_0 \in \Theta$ from conditions (iii.i)-(iii.iii) below:

(iii.i) the continuity of the limit criterion function L_∞ given by

$$L_\infty(\boldsymbol{\theta}) = \mathbb{E}\ell(x_t, x_{t-1}, \boldsymbol{\theta})$$

which follows by continuity of ℓ and the uniform convergence in (i.i);

(iii.ii) the compactness of Θ, which holds again by assumption;

(iii.iii) and the uniqueness of $\boldsymbol{\theta}_0$ which holds either by assumption, if the model is misspecified, or it holds by the information inequality, if the model is correctly specified, $\boldsymbol{\theta}_0$ is identified (which we assume here without proof).

7.8 (e) The ML estimator $\hat{\boldsymbol{\theta}}_T$ is consistent for $\boldsymbol{\theta}_0 \in \Theta$ if

(i) The log likelihood L_T converges uniformly to the limit L_∞

$$\sup_{\boldsymbol{\theta}\in\Theta} \big|L_T(\mathbf{x_T}, \boldsymbol{\theta}) - L_\infty(\boldsymbol{\theta})\big| \xrightarrow{p} 0 \quad \text{as} \quad T \to \infty.$$

(ii) $\boldsymbol{\theta}_0 \in \Theta$ is the identifiably unique maximizer of L_∞

$$\sup_{\boldsymbol{\theta}\in S^c(\boldsymbol{\theta}_0,\delta)} L_\infty(\boldsymbol{\theta}) < L_\infty(\boldsymbol{\theta}_0).$$

We have already seen that

$$L(\mathbf{x_T}, \boldsymbol{\theta}) = \frac{1}{T}\sum_{t=2}^{T} \ell\big(x_t, \hat{\mu}_t(\boldsymbol{\theta}, \hat{\mu}_1), \boldsymbol{\theta}\big)$$

where

$$\ell\big(x_t, \hat{\mu}_t(\boldsymbol{\theta}, \hat{\mu}_1), \boldsymbol{\theta}\big) = -\frac{1}{2}\log 2\pi\sigma_\varepsilon^2 - \frac{\big(x_t - \hat{\mu}_t(\boldsymbol{\theta}, \hat{\mu}_1)\big)^2}{2\sigma_\varepsilon^2}.$$

Hence, we can now obtain a solution, simply by applying the usual steps. Note however, two crucial details in time-varying parameter models:

(1) $\{\ell(x_t, \hat{\mu}_t(\boldsymbol{\theta}, \hat{\mu}_1), \boldsymbol{\theta})\}_{t\in\mathbb{N}}$ cannot be SE, but we can apply the LLN to the SE limit $\{\ell(x_t, \mu_t(\boldsymbol{\theta}), \boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ instead!

(2) to show that $\ell(x_t, \mu_t(\boldsymbol{\theta}), \boldsymbol{\theta})$ is stochastically equicontinuous we can follow the simplification of Theorem 7.9 and check for two moments for the criterion,

$$\mathbb{E}|\ell(x_t, \mu_t(\boldsymbol{\theta}), \boldsymbol{\theta})|^2 < \infty \quad \text{for some } \boldsymbol{\theta} \in \Theta.$$

and two moments for the filter at the initialization $t = 1$, and two moments for the contraction condition of the filter.

If we know that $\{x_t\}_{t\in\mathbb{Z}}$ is SE, then the filter $\{\hat{\mu}_t(\boldsymbol{\theta}, \hat{\mu}_1)\}_{t\in\mathbb{N}}$

$$\hat{\mu}_t = \omega + \alpha(x_{t-1} - \hat{\mu}_{t-1}) + \beta\hat{\mu}_{t-1}$$

converges e.a.s. to a limit SE process $\{\mu_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ as long as

A1: $\{x_t\}_{t\in\mathbb{Z}}$ is SE

A2: $\mathbb{E}\log^+|\omega + \alpha(x_1 - \hat{\mu}_1) + \beta\hat{\mu}_1| < \infty$

A3: $\mathbb{E}\log\sup_\mu|\beta - \alpha| < 0 \quad \Leftrightarrow \quad |\beta - \alpha| < 1.$

If we know that $\{x_t\}_{t\in\mathbb{Z}}$ is SE, then the filter $\{\hat{\mu}_t(\boldsymbol{\theta}, \hat{\mu}_1)\}_{t\in\mathbb{N}}$

$$\hat{\mu}_t = \omega + \alpha(x_{t-1} - \hat{\mu}_{t-1}) + \beta\hat{\mu}_{t-1}$$

has a limit that satisfies $\mathbb{E}|\mu_t(\boldsymbol{\theta})|^n < \infty$ as long as

A1: $\{x_t\}_{t\in\mathbb{Z}}$ is SE

A2: $\mathbb{E}|\omega + \alpha(x_1 - \hat{\mu}_1) + \beta\hat{\mu}_1|^n < \infty$

A3: $\sup_{\hat{\mu},x}|\beta - \alpha| < 1 \quad \Leftrightarrow \quad |\beta - \alpha| < 1.$

We thus have that, if $|\beta-\alpha| < 1$, then $\{\mu_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is SE and $\mathbb{E}|\mu_t(\boldsymbol{\theta})|^8 < \infty$. Indeed, verification of conditions A1-A3 of the uniform contraction theorem delivers:

A1: $\{x_t\}_{t\in\mathbb{Z}}$ is SE

A2: $\mathbb{E}|\omega + \alpha(x_{t-1} - \hat{\mu}_1) + \beta\hat{\mu}_1|^8$
(c_n-inequality) $\leq c|\omega|^8 + c|\alpha|^8\mathbb{E}|x_{t-1}|^8 + |\beta-\alpha|^8|\hat{\mu}_1|^8 < \infty$

A3: $\mathbb{E}\sup_{\hat{\mu},x}|\beta - \alpha| < 1 \quad \Leftrightarrow \quad |\beta - \alpha| < 1.$

Additionally, the log likelihood converges pointwise since:

(a) $\{\ell\big(x_t, \mu_t(\boldsymbol{\theta}), \boldsymbol{\theta}\big)\}$ is SE by Krengel's theorem

(b) $\mathbb{E}|\ell\big(x_t, \mu_t(\boldsymbol{\theta}), \boldsymbol{\theta}\big)| < \infty$ (using usual inequalities)

Furthermore, the log likelihood is stochastically equicontinuous because:

(a) Θ is compact;

(b) ℓ is continuously differentiable and well behaved of order two in both $\boldsymbol{\theta}$ and $\hat{\mu}_t$;

(c) $\phi(x_t, \hat{\mu}_t) = \omega + \alpha(x_t - \hat{\mu}_t) + \beta\hat{\mu}_t$ is well behaved of order two in both $\boldsymbol{\theta}$ and $\hat{\mu}_t$;

(d) The filter contraction holds uniformly $\sup_{\hat{\mu},x} |\beta - \alpha| < 1$;

(e) $\mathbb{E}|\ell\big(x_t, \hat{\mu}_t(\boldsymbol{\theta}), \boldsymbol{\theta}\big)|^2 < \infty$ (using usual inequalities)

Note also that the criterion converges uniformly in probability

$$\frac{1}{T}\sum_{t=2}^{T} \ell\big(x_t, \hat{\mu}_t(\boldsymbol{\theta}, \hat{\mu}_1), \boldsymbol{\theta}\big) \xrightarrow{p} \mathbb{E}\ell\big(x_t, \mu_t(\boldsymbol{\theta}), \boldsymbol{\theta}\big)$$

since $\frac{1}{T}\sum_{t=2}^{T} \ell\big(x_t, \mu_t(\boldsymbol{\theta}), \boldsymbol{\theta}\big)$ satisfies an LLN, ℓ is continuously differentiable, and $|\mu_t(\boldsymbol{\theta}) - \hat{\mu}_t(\boldsymbol{\theta}, \hat{\mu}_1)| \xrightarrow{p} 0$ as $t \to \infty$; see Theorem 7.13.

Note that for a.s. convergence we would further have to show that $\mathbb{E}\log^+ \sup_\mu |\partial\ell\big(x_t, \mu_t(\boldsymbol{\theta}), \boldsymbol{\theta}\big)|/\partial\mu < \infty$.

Finally, $\boldsymbol{\theta}_0$ is identifiably unique if:

- Θ is compact: which holds by assumption;
- $Q_\infty \in \mathbb{C}(\Theta)$: holds by Lemma 7.2 since $Q_T \in \mathbb{C}(\Theta)\ \forall\ T$ and $Q_T \xrightarrow{p} Q_\infty$ uniformly
- $\boldsymbol{\theta}_0$ is unique: which holds by identification of $\boldsymbol{\theta}_0$ and the information inequality (for well specified model), or by assumption (for a misspecified model). If the uniqueness assumption is too restrictive, then set consistency is also available under the same conditions.

Hence, $\hat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0$ by classical consistency theorem!

Note that the criterion

$$\ell\big(x_t, \mu_t(\boldsymbol{\theta}), \boldsymbol{\theta}\big) = -\frac{1}{2}\log 2\pi\sigma_\varepsilon^2 - \frac{\big(x_t - \mu_t(\boldsymbol{\theta})\big)^2}{2\sigma_\varepsilon^2}$$

is not differentiable at σ_ε^2, hence Θ must be a compact parameter space with $\sigma_\varepsilon^2 > 0$. For example

$$\Theta = \Big\{(\omega, \alpha, \beta, \sigma_\varepsilon^2) \in [0,5] \times [0,1] \times [0,1] \times [0.1, 10]\Big\}$$

7.8 (f) The ML estimator $\hat{\boldsymbol{\theta}}_T$ is consistent for $\boldsymbol{\theta}_0$ if Θ is a compact parameter space satisfying

(a) $|\beta| < 1$ for every $\boldsymbol{\theta} \in \Theta$,

(b) $\omega \geq a > 0, \quad \alpha \geq a > 0, \quad \beta \geq a > 0, \quad$ for every $\boldsymbol{\theta} \in \Theta$

and $\boldsymbol{\theta}_0$ is the unique maximizer of

$$L_\infty(\boldsymbol{\theta}) = \mathbb{E}\Big[-\log\sigma_t^2(\boldsymbol{\theta}) - x_t^2/\sigma_t^2(\boldsymbol{\theta})\Big]$$

where $\sigma_t^2(\boldsymbol{\theta})$ denotes the limit filtered conditional volatility. Indeed, we first note that the log likelihood function is given by

$$L_T(\mathbf{x_T}, \boldsymbol{\theta}) = \frac{1}{T}\sum_{t=2}^{T} -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\hat{\sigma}_t^2(\boldsymbol{\theta}, \hat{\sigma}_1^2) - \frac{1}{2}\frac{x_t^2}{\hat{\sigma}_t^2(\boldsymbol{\theta}, \hat{\sigma}_1^2)}.$$

Hence, we can define $\hat{\boldsymbol{\theta}}_T$ as the maximizer of

$$\begin{aligned} L_T(\mathbf{x_T}, \boldsymbol{\theta}) &= \frac{1}{T}\sum_{t=2}^{T} \ell(x_t, \hat{\sigma}_t^2(\boldsymbol{\theta}, \hat{\sigma}_1^2)) \\ &= \frac{1}{T}\sum_{t=2}^{T} -\log\hat{\sigma}_t^2(\boldsymbol{\theta}, \hat{\sigma}_1^2) - \frac{x_t^2}{\hat{\sigma}_t^2(\boldsymbol{\theta}, \hat{\sigma}_1^2)}. \end{aligned}$$

Now, we can obtain consistency through the uniform convergence of the criterion function and the identifiable uniqueness of $\boldsymbol{\theta}_0$. Note that,

(1) $\{\hat{\sigma}_t^2(\boldsymbol{\theta}, \hat{\sigma}_1^2)\}$ converges to a limit SE sequence satisfying $\mathbb{E}|\sigma_t^2(\boldsymbol{\theta})|^2 = \mathbb{E}|\sigma_t(\boldsymbol{\theta})|^4 < \infty$ by the Uniform Contraction theorem:

A1: $\{x_t\}_{t\in\mathbb{Z}}$ is SE

A2: $\mathbb{E}|\omega+\alpha x_t^2+\beta\hat{\sigma}_1^2|^2 \;\leq\; c|\omega|^2+c|\alpha|^2\mathbb{E}|x_t^2|^2+|\beta|^2|\hat{\sigma}_1^2|^2$

$$= \; c|\omega|^2 + c|\alpha|^2\mathbb{E}|x_t|^4 + |\beta|^2|\hat{\sigma}_1^2|^2 \; < \infty$$

A3: $\sup_{\hat{\sigma}^2,x} |\beta| < 1 \quad \Leftrightarrow \quad |\beta| < 1.$

and $\{\ell(x_t, \sigma_t^2(\boldsymbol{\theta}))\}_{t\in\mathbb{Z}}$ is SE by Krengel's theorem

(2) Since $\omega \geq a > 0, \quad \alpha \geq a > 0, \quad \beta \geq a > 0$, then it follows that $\sigma_t^2(\boldsymbol{\theta}) > a$. Hence,

$$\begin{aligned}
\mathbb{E}|\ell(x_t, \sigma_t^2(\boldsymbol{\theta}))|^2 \;&\leq\; \mathbb{E}| - \log\sigma_t^2(\boldsymbol{\theta}) - x_t^2/\sigma_t^2(\boldsymbol{\theta})|^2 \\
\text{(c_n-inequality)} \qquad &\leq\; c\mathbb{E}|\log\sigma_t^2(\boldsymbol{\theta})|^2 + c\mathbb{E}|x_t^2/\sigma_t^2(\boldsymbol{\theta})|^2 \\
&\leq\; c\mathbb{E}|\log\sigma_t^2(\boldsymbol{\theta})|^2 + \frac{c}{a^2}\mathbb{E}|x_t|^4 \; < \infty
\end{aligned}$$

Note that if $z_t > a > 0$ then $\mathbb{E}|\log(z_t)|^k < \infty$ if $\mathbb{E}|z_t|^n < \infty$. Also note that if $z_t > a$, then $\mathbb{E}|x_t^2/z_t|^n \;\leq\; \mathbb{E}|x_t^2/a|^n \;\leq\; \frac{1}{a^n}\mathbb{E}|x_t^2|^n$. As a result, we have that $\{\ell(x_t, \sigma_t^2(\boldsymbol{\theta}))\}$ satisfies a LLN

$$\frac{1}{T}\sum_{t=2}^{T} \ell(x_t, \sigma_t^2(\boldsymbol{\theta})) \;\xrightarrow{p}\; \mathbb{E}\ell(x_t, \sigma_t^2(\boldsymbol{\theta})).$$

Now, $\{\ell(x_t, \sigma_t^2(\boldsymbol{\theta}))\}$ is stochastically equicontinuous if the criterion function is well behaved and we have bounded moments of appropriate order; see Theorem 7.9 and check these yourself! Hence, we obtain uniform convergence of the SE criterion

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{T}\sum_{t=2}^{T} \ell(x_t, \sigma_t^2(\boldsymbol{\theta})) \;-\; \mathbb{E}\ell(x_t, \sigma_t^2(\boldsymbol{\theta}))\right| \xrightarrow{p} 0.$$

Furthermore, we obtain convergence of the filter-based criterion

$$\frac{1}{T}\sum_{t=2}^{T} \ell(x_t, \hat{\sigma}_t^2(\boldsymbol{\theta}, \hat{\sigma}_1^2)) \;\xrightarrow{p}\; \mathbb{E}\ell(x_t, \sigma_t^2(\boldsymbol{\theta}))$$

by application of Theorem 7.13. Finally, we obtain identifiable uniqueness of $\boldsymbol{\theta}_0$ using the same argument as in the previous exercise. We thus conclude that the ML estimator $\hat{\boldsymbol{\theta}}_T$ is consistent for $\boldsymbol{\theta}_0$ by the classical consistency theorem for M-estimators.

$$\hat{\boldsymbol{\theta}}_T \xrightarrow{p} \boldsymbol{\theta}_0 \quad \text{as} \quad T \to \infty.$$

Note also that for strong consistency $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0$ we would need to further check a logarithmic moment for the score w.r.t. σ_t^2. You can try this at home!

7.9 If the model is well specified, then the observed data was generated by the model under some $\boldsymbol{\theta}_0 \in \Theta$. Hence, we can use the Power-n Theorem, or the uniform contraction theorem, to ensure that the model generates SE paths with enough bounded moments for every $\boldsymbol{\theta} \in \Theta$. For example, for the local-level model we have,

$$x_t = \mu_t + \varepsilon_t \qquad \text{where} \quad \mu_t = \omega + \alpha(x_{t-1} - \mu_{t-1}) + \beta\mu_{t-1}.$$

Now, if Θ is such that $|\beta| < 1$ for any $\boldsymbol{\theta} \in \Theta$, then $\{x_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is SE and satisfies $\mathbb{E}|x_t(\boldsymbol{\theta}_0)|^n < \infty$ for any $n > 0$ and any $\boldsymbol{\theta}_0 \in \Theta$ because

$$\mu_t = \omega + \alpha\epsilon_{t-1} + \beta\mu_{t-1}.$$

and hence, by the uniform contraction theorem $\{\mu_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is SE and satisfies $\mathbb{E}|\mu_t(\boldsymbol{\theta}_0)|^n < \infty$ for any $n > 0$ and any $\boldsymbol{\theta}_0 \in \Theta$. As a result, $\{x_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is SE (Krengel's theorem) and $\mathbb{E}|x_t(\boldsymbol{\theta}_0)|^n < \infty$ (c_n-inequality) $\mathbb{E}|x_t(\boldsymbol{\theta}_0)|^n \leq c\mathbb{E}|\mu_t(\boldsymbol{\theta})|^n + c\mathbb{E}|\epsilon_t|^n < \infty$.

7.11 The data is composed of Stock prices of IBM and Microsoft, from 01-01-2000 to 24-09-2017, (`stocks_data.mat`). We optimize the portfolio by selecting weights $k_{1,t}$ and $k_{2,t} = 1 - k_{1,t}$ which optimize a Sharpe Ratio; i.e. which minimize risk and maximize expected returns,

$$\max_{k_{1t}} \frac{k_{1t}\mu_{1t} + (1-k_{1t})\mu_{2t}}{\sqrt{k_{1t}^2\sigma_{1t}^2 + (1-k_{1t})^2\sigma_{2t}^2 + 2k_{1t}(1-k_{1t})\sigma_{12t}}},$$

$$\text{s.t. } 0 \leq k_{1t} \leq 1.$$

$$k_{1t} = \frac{\mu_{1t}\sigma_{2t}^2 - \mu_{2t}\sigma_{12t}}{\mu_{1t}\sigma_{2t}^2 + \mu_{2t}\sigma_{1t}^2 - (\mu_{1t} + \mu_{2t})\sigma_{12t}} \quad , \quad k_{2t} = 1 - k_{1t}.$$

We consider independent stocks (so that $\sigma_{12t} = 0$), and calculate optimal weights k_{1t} and k_{2t} for every period t. Figure 1 shows the resulting weights.

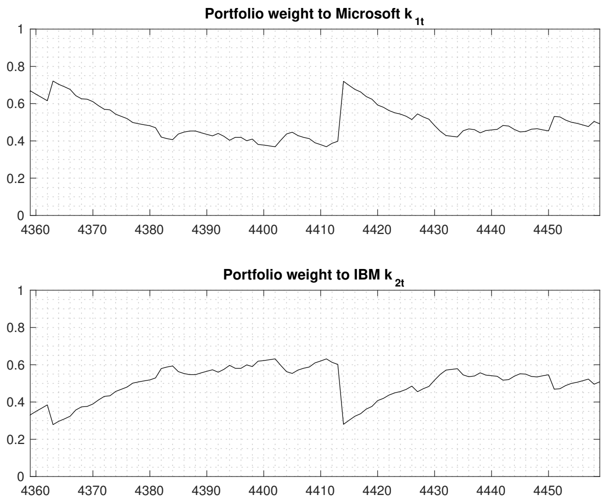


Figure 1: Optimal portfolio weights obtained using GARCH model for Microsoft and IBM returns and CCC covariance

7.16 With a CCC model the portfolio optimization delivers the following results detailed in Figure 2.

7.17 We estimate a GARCH(1,1) on the *heart rate variability* (HRV) data depicted in Figure 3. The parameter estimates obtained from series 1 are $(\hat{\alpha}, \hat{\beta}) = (0.018, 0.001)$ and the estimates for series 2 are $(\hat{\alpha}, \hat{\beta}) = (0.205, 0.758)$. Series 2 reveals considerably larger conditional volatility than Series 1. The prognostic seems correct. A proper statistical test should be preformed.

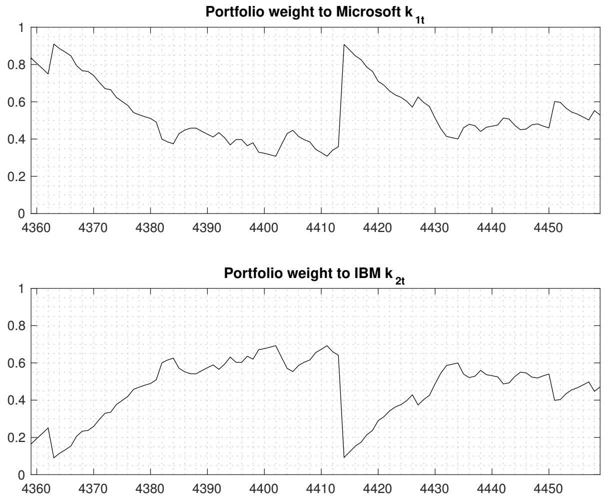


Figure 2: Optimal portfolio weights obtained using GARCH model for Microsoft and IBM returns `Optim_Portfolio.m`

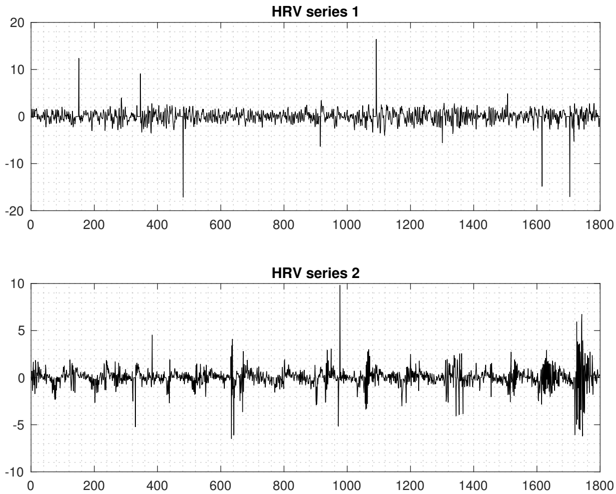


Figure 3: Heart rate variability data for two individuals.

Exercises of Chapter 8

8.1 (a) The least-squares estimator $\hat{\boldsymbol{\theta}}_T$ is given by

$$\hat{\boldsymbol{\theta}}_T \in \arg\max_{\boldsymbol{\theta}\in\Theta} -\frac{1}{T}\sum_{t=2}^{T}(x_t - \alpha - \beta\cos(x_{t-1}))^2.$$

Since $\hat{\boldsymbol{\theta}}_T$ is consistent for $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$, then we obtain asymptotic normality if

(a) Θ is compact (just assume it!)
(b) $\sqrt{T}\nabla Q_T(\mathbf{x}_T, \boldsymbol{\theta}_0)$ converges to a normal
(c) $\nabla^2 Q_T$ converges uniformly to $\nabla^2 Q_\infty$
(d) $\nabla^2 Q_\infty(\boldsymbol{\theta}_0)$ is invertible.

Since $q(x_t, x_{t-1}, \boldsymbol{\theta}) = (x_t - \alpha - \beta\cos(x_{t-1}))^2$ is continuously differentiable of any order, we know by Krengel's theorem that the n-th derivative $\{\nabla^n q(x_t, x_{t-1}, \boldsymbol{\theta})\}$ is SE for any n.

Additionally, we note that $\mathbb{E}|q(x_t, x_{t-1}, \boldsymbol{\theta})|^2 < \infty$

$$\begin{aligned}\mathbb{E}|x_t - \alpha - \beta\cos(x_{t-1})|^4 \\ \leq\ c\underbrace{\mathbb{E}|x_t|^4}_{<\infty} + c\underbrace{|\alpha|^4}_{<\infty} + c\underbrace{|\beta|^4}_{<\infty}\underbrace{\mathbb{E}|\cos(x_{t-1})|^4}_{\leq 1<\infty} \ <\ \infty\end{aligned}$$

Furthermore, since $q \in \mathbb{C}^3$, and assuming that q well behaved of order 2, and ∇q and $\nabla^2 q$ are both well behaved of order 1, we have that,

(i) $\mathbb{E}\|\nabla q(x_t, x_{t-1}, \boldsymbol{\theta}_0)\|^2 < \infty$;
(ii) $\mathbb{E}\|\nabla^2 q(x_t, x_{t-1}, \boldsymbol{\theta})\| < \infty \ \forall\ \boldsymbol{\theta} \in \Theta$;

(iii) $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\nabla^3 q(x_t, x_{t-1}, \boldsymbol{\theta})\| < \infty$.

Now, when the model is well specified, we have that $\{\nabla q(x_t, x_{t-1}, \boldsymbol{\theta}_0)\}$ is a martingale difference sequence and hence satisfies a CLT,

$$\sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\nabla q(x_t, x_{t-1}, \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \Sigma) \ \text{ as } \ T \to \infty$$

In contrast, if the model is mis-specified a CLT may hold if $\{x_t\}_{t\in\mathbb{Z}}$ is NED on a mixing sequence of appropriate size and ∇q is Lipschitz continuous in (x_t, x_{t-1}),

$$\sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\nabla q(x_t, x_{t-1}, \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \Sigma) \ \text{ as } \ T \to \infty.$$

Naturally, since $\{\nabla^2 q(x_t, x_{t-1}, \boldsymbol{\theta})\}$ is SE and $\mathbb{E}\|\nabla^2 q(x_t, x_{t-1}, \boldsymbol{\theta})\| < \infty$ for every $\boldsymbol{\theta} \in \Theta$ we have that,

$$\frac{1}{T}\sum_{t=2}^{T}\nabla^2 q(x_t, x_{t-1}, \boldsymbol{\theta}) \xrightarrow{p} \mathbb{E}\nabla^2 q(x_t, x_{t-1}, \boldsymbol{\theta})$$

as $T \to \infty \ \forall \ \boldsymbol{\theta} \in \Theta$. Furthermore, since $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\nabla^3 q(x_t, x_{t-1}, \boldsymbol{\theta})\| < \infty$ it follows that,

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{T}\sum_{t=2}^{T}\nabla^2 q(x_t, x_{t-1}, \boldsymbol{\theta}) - \mathbb{E}\nabla^2 q(x_t, x_{t-1}, \boldsymbol{\theta})\right\| \xrightarrow{p} 0$$

as $T \to \infty$. Finally, if $\boldsymbol{\theta}_0$ is the unique maximizer of Q_∞ and $\nabla^2 Q_\infty$ is regular, then $\nabla^2 Q_\infty(\boldsymbol{\theta}_0)$ is invertible.

We conclude that all conditions of the classical asymptotic normality theorem are satisfied, and hence, we have

$$\sqrt{T}\big(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\big) \ \xrightarrow{d} \ N\big(0 \ , \ \Omega\Sigma\Omega^\top\big)$$

as $T \to \infty$ where

$$\Omega = \Big(\mathbb{E}\nabla^2 q(x_t, x_{t-1}, \boldsymbol{\theta}_0)\Big)^{-1}.$$

8.1 (b)-(d) A solution to these models is found as a special case of the NLAR(1) example in p.227.

8.2 (a) The solution is similar to that of the MLE for the Gaussian AR(1) model in p.225.

8.2 (b)-(c) A solution to these models is found as a special case of the MLE for the NLAR(1) model in p.223.

8.3 When a model is correctly specified, we can use the Theory of Chapter 4 to establish SE properties and bounded moments for the data generated by the model. This allows us to apply LLNs to the criterion function and its second derivative. Additionally, we know that the score evaluated at the true parameter $\boldsymbol{\theta}_0$ is a martingale difference sequence (mds) when the model is correctly specified, which opens the door to obtaining a CLT for the scaled score. Finally, we can easily ensure the uniqueness of $\boldsymbol{\theta}_0$ and the invertibility of the limit Hessian when $\boldsymbol{\theta}_0$ is identifiable.

In contrast, then the model is misspecified, then we cannot use the theory of Chapter 4 to verify the properties of the data, because the DGP is unknown. In practice, instead of assuming that we know the DGP, we must assume properties for the data such as stationarity, ergodicity, bounded moments, etc.[3] Additionally, since the score cannot be ensured to be an mds, we must make use of other CLTs, such as the CLTs for NED sequences studied in Chapter 4. Recall that the score is NED if it is Lipschitz continuous on a NED sequence. Finally, either we assume the uniqueness of the pseudo-true parameter $\boldsymbol{\theta}_0 \in \Theta$, which may be quite restrictive, or the asymptotic normality may fail.

[3]These are properties which we may attempt to test empirically using statistical tests for stationarity, bounded moments, etc.

Exercises of Chapter 9

9.7 Since the innovations are additive and Gaussian the conditional probabilities are easy to calculate and there is no need for simulations. In particular, we have that the probability that output will fall in the first quarter of 2016 is given by

$$\text{AR: } P(x_{T+1} < 0) \approx 0.04$$
$$\text{SESTAR: } P(x_{T+1} < 0) \approx 0.187.$$

In contrast, the probability that output will rise above 5% is given by

$$\text{AR: } P(x_{T+1} > 0.05) \approx 0.30$$
$$\text{SESTAR: } P(x_{T+1} > 0.05) \approx 0.083.$$

9.11 The forecasts and respective bounds are shown in Figure 4.

9.12 Figure 5 shows how the forecast of the AR(1) and SESTAR models compare against the realized data. The SESTAR model performs better in this case.

9.13 The IRF is shown in Figure 6. In the case of nonlinear models, the impact of the shock depends on the origin of the IRF.

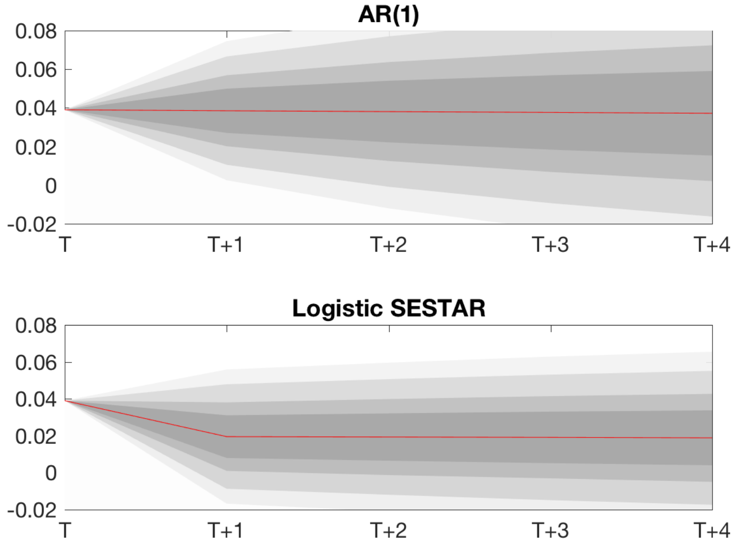


Figure 4: Forecasts from AR(1) and SESTAR.

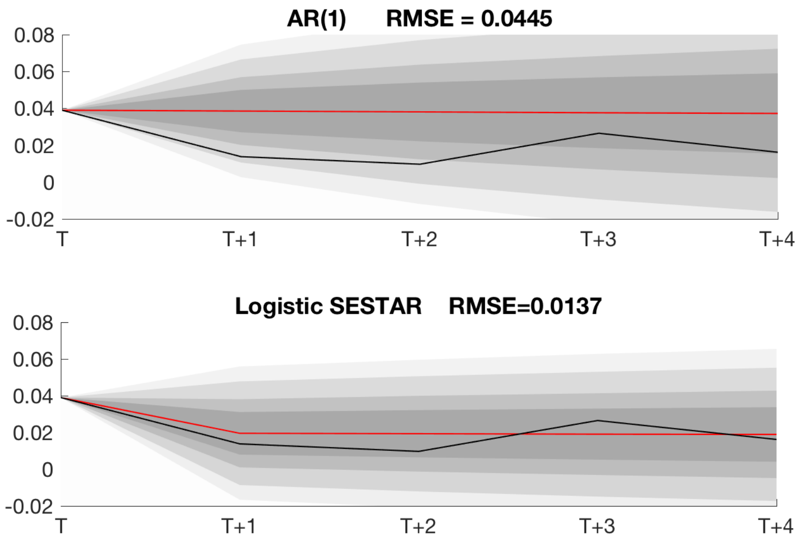


Figure 5: Forecasts against realized data.

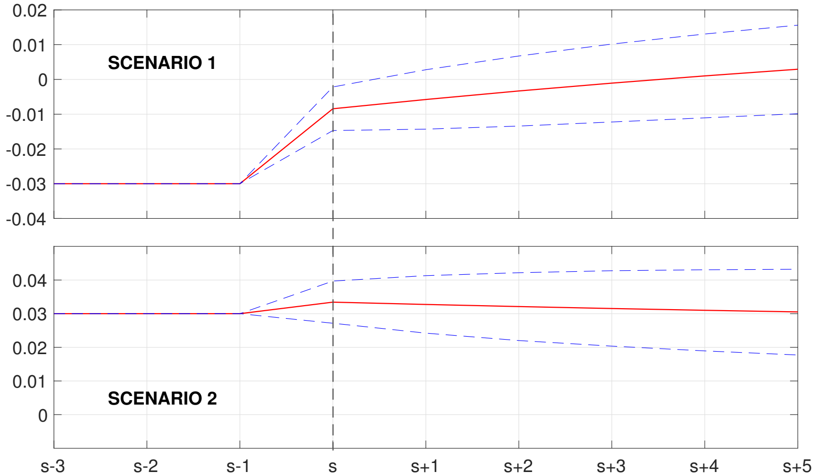


Figure 6: IRFs with different origins

Exercises of Chapter 10

10.1 (a) Not necessarily.

$\boldsymbol{\theta}_0^{\text{LS}}$ minimizes a transformation of an L^2-norm distance between the true regression function and the model's regression functions.

$\boldsymbol{\theta}_0^{\text{ML}}$ minimizes the Kullback-Leibler distance between the true conditional density of the data and the model's conditional densities.

These 'best approximation' points are not necessarily the same.

10.1 (b) In general yes.

If the model parameters are identified
(i.e. each $\boldsymbol{\theta} \in \Theta$ defines a unique regression function and a unique conditional density for the data)

Then $\boldsymbol{\theta}_0^{\text{LS}} = \boldsymbol{\theta}_0^{ML} = \boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ is the *true parameter*.

10.2 $\hat{\boldsymbol{\theta}}_T$ minimizes the *sum of squared residuals.* If $\mathbb{E}|g(x_{t-1}, \boldsymbol{\theta})|^4 < \infty$, Then, by the c_n-inequality, we have

$$\begin{aligned} &\mathbb{E}|(x_t - g(x_{t-1}, \boldsymbol{\theta})x_{t-1})^2| \\ &\qquad \leq \; c\underbrace{\mathbb{E}|x_t|^2}_{<\infty} + c\underbrace{\mathbb{E}|g(x_{t-1}, \boldsymbol{\theta})|^4}_{<\infty}\underbrace{\mathbb{E}|x_{t-1}|^4}_{<\infty} \; < \; \infty \end{aligned}$$

Hence, we know by application of a LLN that the limit criterion Q_∞ takes the form

$$\begin{aligned} Q_\infty(\boldsymbol{\theta}) &= -\mathbb{E}\Big(x_t - g(x_{t-1}, \boldsymbol{\theta})x_{t-1}\Big)^2 \\ &= -\mathbb{E}\Big(\phi_0(x_{t-1}) + \epsilon_t - g(x_{t-1}, \boldsymbol{\theta})x_{t-1}\Big)^2 \end{aligned}$$

Note that we eliminate ϵ_t above because this does not change the maximizer $\boldsymbol{\theta}_0$.[4] We thus have that

$$\boldsymbol{\theta}_0 = \arg\max_{\boldsymbol{\theta}\in\Theta} - \int \Big(\phi_0(x) - g(x, \boldsymbol{\theta})x\Big)^2 dP_0(x)$$

where P_0 denotes the true distribution of the data. As a result, $\boldsymbol{\theta}_0$ minimizes also an L^2-norm between ϕ_0 and $\phi(\cdot, \boldsymbol{\theta})$ where $\phi(x, \boldsymbol{\theta}) := g(x, \boldsymbol{\theta})x \ \ \forall\ x$

$$\boldsymbol{\theta}_0 \ = \ \arg\min_{\boldsymbol{\theta}\in\Theta} \left(\int \Big|\phi_0(x) - g(x, \boldsymbol{\theta})x\Big|^2 dP_0(x) \right)^{\frac{1}{2}} ,$$

since the square-root function $(\cdot)^{1/2}$ is strictly increasing.

10.3 If model C nests model D, then the R^2 of model C must be larger equal than the R^2 of model D. Furthermore, the adjusted R^2 of model B can not be larger than the R^2.

10.4 It depends... Selecting model A (with lower likelihood than model B) or selecting model D (with lower likelihood than model C) seems unreasonable. Now, between B and C one may find multiple answers depending on the criterion that one chooses to use, but according to the AIC, model C is the best.

[4] This is true in particular, when ϵ_t is and independent of x_{t-1}. Define $\Delta_{t-1} = \phi_0(x_{t-1}) - g(x_{t-1}, \boldsymbol{\theta})x_{t-1}$, then

$$\begin{aligned} \mathbb{E}(\Delta_{t-1} + \epsilon_t)^2 &= \mathbb{E}\Delta_{t-1}^2 + \mathbb{E}\epsilon_t^2 + 2\mathbb{E}\epsilon_t\Delta_{t-1} \\ &= \mathbb{E}\Delta_{t-1}^2 + \mathbb{E}\epsilon_t^2 + 0 \ , \end{aligned}$$

since by independence $\mathbb{E}\epsilon_t\Delta_{t-1} = \mathbb{E}\epsilon_t\mathbb{E}\Delta_{t-1} = 0$. Finally, we note that the argmax of $-\mathbb{E}\Delta_{t-1}^2 - \mathbb{E}\epsilon_t^2$ is the same as the argmax of $-\mathbb{E}\Delta_{t-1}^2$ because $\mathbb{E}\epsilon_t^2$ is just a constant.

10.5 If we use the AIC as a guiding principle, then the answer does not change with the additional information about sample size. But this would not be the case when using other criteria such as the corrected AIC (AICc), or the Hannan–Quinn information criterion (HQC), among others.

Exercises of Chapter 11

11.1 (a) All of these variables are simultaneously determined. On the one hand, interest rates set by the central bank affect aggregate investment and economic performance. On the other hand, central banks shift interest rates in reaction to aggregate investment and economic performance (either explicitly like the US Federal Reserve, or implicitly like the ECB, whose mandate dictates inflation as being the sole relevant criterion).

11.5 That's true! Many great structural models have a negative R^2. Wait, is that even possible? Sure! A model has a negative R^2 when its predictive accuracy is worse than using the sample average. This just highlights further the difference between predictive modeling and structural modeling!