TIME SERIES MODELS
Lectures Week 5

Ilka van de Werve

VU, TI and NSCR

March 3 & 4, 2025

http://www.ilka.nl
i.vande.werve@vu.nl

Introduction

This handout provides a roadmap for implementing importance sampling, including tips and tricks for numerical stability. References to relevant sections of the DK-book are included. These slides supplement the lecture slides mentioned on the front page, but the set in this document is *not* part of the exam material for Time Series Models.

Starting Point

- For parameter estimation or signal extraction in a nonlinear non-Gaussian model, we rely on simulation-based methods because the smoothed density $p(\theta|Y_n)$ is unknown

- Using the concept of importance sampling (see lecture slides), we know that MLE can be based on

$$\ell(\psi) = \log L(\psi) \overset{DK-11.6}{=\cdots=} \log g(Y_n) + \log \mathbb{E}_g[w(\theta, Y_n)],$$

  and that signal extraction can be based on

$$\bar{x} = \mathbb{E}[x(\theta)|Y_n] \overset{DK-11.2}{=\cdots=} \frac{\mathbb{E}_g[x(\theta)w(\theta, Y_n)]}{\mathbb{E}_g[w(\theta, Y_n)]},$$

  where $\mathbb{E}_g[\cdot]$ is with respect to the importance density $g(\theta|Y_n)$, and the importance weights are $w(\theta, Y_n) = p(Y_n|\theta)/g(Y_n|\theta)$

- The estimates are $\hat{\ell}(\psi) = \log g(Y_n) + \log \bar{w}$ and $\hat{x} = \frac{\sum_{i=1}^{M} x_i w_i}{\sum_{i=1}^{M} w_i}$
  - ▷ $x_i = x\big(\tilde{\theta}^{(i)}\big)$, $w_i = w\big(\tilde{\theta}^{(i)}, Y_n\big)$ and $\tilde{\theta}^{(i)} \sim g(\theta|Y_n), \forall i$

Roadmap

Steps to be taken:

1. Construct approximating model "$g(\cdot)$" with SPDK-method
   - Based on the mode

2. Draw paths $\tilde{\theta}^{(i)} \sim g(\theta | Y_n)$ using simulation smoothing, $\forall i$

3. Evaluate $p\big(Y_n | \tilde{\theta}^{(i)}\big)$ and $g\big(Y_n | \tilde{\theta}^{(i)}\big)$

4. Construct importance weights $w_i = w\big(\tilde{\theta}^{(i)}, Y_n\big)$
   - In a slightly different form for numerical stability

5A. If interested in MLE: evaluate $\log g(Y_n)$ and $\log \bar{w}$ to construct $\hat{\ell}(\psi)$

5B. If interested in signal extraction: evaluate $x_i = x\big(\tilde{\theta}^{(i)}\big)$ to construct $\hat{x}$

Step 1a

Construct approximating model "$g(\cdot)$" with SPDK-method (based on the mode, start)

- To obtain the mode $\check{\theta} = \arg\max_\theta p(\theta|Y_n)$ numerically using Newton-Raphson: for current guess g of the mode, introduce $A = -(\ddot{p}(Y_n|\theta)|_{\theta=g})^{-1}$ and $z = g + A\,\dot{p}(Y_n|\theta)|_{\theta=g}$
- Idea: consider the linear Gaussian model with $Y_n = z$ and $H = A$ and obtain new guess g^+ from the KS-output
- While loop: for guess g, evaluate z and A and run KFS on

$$\begin{aligned} z_t &= d_t + Z_t\alpha_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, A_t) \\ \alpha_{t+1} &= c_t + T_t\alpha_t + R_t\eta_t, \quad \eta_t \sim \mathcal{N}(0, Q_t) \end{aligned}$$

  to estimate the smoothed mean $\hat{\alpha}$. Then calculate smoothed mean $\hat{\theta} = d + Z\hat{\alpha} = g^+$ (the updated guess) and check convergence for (if so, mode is $\check{\theta} = \hat{\theta}$; otherwise set $g = g^+$ and repeat)

Step 1b

Construct approximating model "$g(\cdot)$" with SPDK-method (based on the mode, continued)

- Since the mode $\check{\theta}$ is used to store $A = -(\ddot{p}(Y_n|\theta)|_{\theta=\check{\theta}})^{-1}$ and $z = g + A\,\dot{p}(Y_n|\theta)|_{\theta=\check{\theta}}$, they inherently depend on Y_n
  - ▷ Note that final values A and z are evaluated at the mode $\check{\theta}$ (instead of at the guess g in the while-loop)

- The approximating model $g(Y_n, \theta) \equiv g(z, \theta)$ follows from

$$\begin{aligned} z_t &= \theta_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0, A_t) \\ \theta_t &= d_t + Z_t\alpha_t \\ \alpha_{t+1} &= c_t + T_t\alpha_t + R_t\eta_t, & \eta_t &\sim \mathcal{N}(0, Q_t), \end{aligned}$$

- This approximating model with known z and A is linear Gaussian, hence can proceed with KFS-methods

Step 2

Draw $\tilde{\theta}^{(i)} \sim g(\theta|Y_n) \equiv g(\theta|z)$ using simulation smoothing, $\forall i$

- $\hat{\theta} = \mathbb{E}[\theta|z]$ is the smoothed mean from the approximating model, so use KFS methods to obtain it

- $(\theta^+, y^+) \sim g(\theta, z) = g(z|\theta)g(\theta)$ from unconditional simulation
  - Simulate errors $\eta_t^+ \sim \mathcal{N}(0, Q_t)$ and $\varepsilon_t^+ \sim \mathcal{N}(0, A_t)$, $\forall t$
  - Simulate initial state $\alpha_1^+ \sim \mathcal{N}(a_1, P_1)$
  - Recursively compute states $\alpha_{t+1}^+ = c_t + T_t\alpha_t^+ + R_t\eta_t^+$ and, given these states, generate signal $\theta_t^+ = d_t + Z_t\alpha_t^+$ and observations $z_t^+ = \theta_t^+ + \varepsilon_t^+$

- $\hat{\theta}^+ = \mathbb{E}[\theta|z^+]$ is the smoothed mean from the simulated data set, so use KFS methods to obtain it

- Repeat the above $i = 1, \cdots, M$ times to obtain many draws $\tilde{\theta}^{(i)} = \hat{\theta}^{(i)} + \theta^{+,(i)} - \hat{\theta}^{+,(i)}$ from $g(\theta|Y_n) \equiv g(\theta|z)$

Step 3

Evaluate $p\big(Y_n|\tilde{\theta}^{(i)}\big)$ and $g\big(Y_n|\tilde{\theta}^{(i)}\big)$

- We know how $p(Y_n|\theta)$ looks like (it is the model), so we can evaluate $p\big(Y_n|\tilde{\theta}^{(i)}\big) = \prod_{t=1}^{n} p\left(y_t|\tilde{\theta}_t^{(i)}\right)$

- Since the approximating model is linear Gaussian, we can also evaluate $g\big(Y_n|\tilde{\theta}^{(i)}\big) \equiv g\big(z|\tilde{\theta}^{(i)}\big) = \prod_{t=1}^{n} g\left(z_t|\tilde{\theta}_t^{(i)}\right)$ because $\left(z_t|\tilde{\theta}_t^{(i)}\right) \sim \mathcal{N}\left(\tilde{\theta}_t^{(i)}, A_t\right)$

Step 4

Construct importance weights $w_i = w\left(\tilde{\theta}^{(i)}, Y_n\right)$
(in a slightly different form for numerical stability)

- Let $p^{(i)} = \exp\left(\sum_{t=1}^{n} \log p\left(y_t|\tilde{\theta}_t^{(i)}\right)\right)$ and similarly $g^{(i)} = \exp\left(\sum_{t=1}^{n} \log g\left(z_t|\tilde{\theta}_t^{(i)}\right)\right)$

- For numerical stability, use

$$\begin{aligned} w_i = p^{(i)}/g^{(i)} &= \exp\left(\log p^{(i)} - \log g^{(i)}\right) \\ &= \exp(m_i) = \exp(\bar{m})\exp(m_i - \bar{m}), \end{aligned}$$

where $m_i = \log p^{(i)} - \log g^{(i)}$ and $\bar{m} = M^{-1}\sum_{i=1}^{M} m_i$

- So first compute m_i, $\forall i$, and then compute sample average $\bar{m}$
  - ▷ No direct need to compute w_i, remainder will be in terms of m_i

Step 5A

If interested in MLE: Evaluate $\log g(Y_n)$ and $\log \bar{w}$ to construct $\hat{\ell}(\psi)$

- The approximating model is linear Gaussian with A_t is fixed, so use the prediction error decomposition to obtain $\log g(Y_n) \equiv \log g(z)$ with KF-methods
- And, again for numerical stability, use

$$\hat{\ell}(\psi) = \log g(Y_n) + \bar{m} - \log M + \log \left(\sum_{i=1}^{M} \exp(m_i - \bar{m}) \right)$$

- In practice, maximize average log-likelihood function $\hat{\ell}(\psi)/n$
- Use the same random seed (or save the random numbers) and the same value of M for each log-likelihood evaluation

Step 5B

If interested in signal extraction: evaluate $x_i = x\big(\tilde{\theta}^{(i)}\big)$ to construct $\hat{x}$

- For each draw i compute $x_i = x\big(\tilde{\theta}^{(i)}\big)$

- Finally, $\hat{x} = \sum_{i=1}^{M} x_i \exp(m_i - \bar{m}) / \sum_{i=1}^{M} \exp(m_i - \bar{m})$ is the estimate of $\bar{x} = \mathbb{E}[x(\theta)|Y_n]$ given parameter vector ψ

TIME SERIES MODELS
Lectures Week 5

Ilka van de Werve

VU, TI and NSCR

March 3 & 4, 2025

`http://www.ilka.nl`
`i.vande.werve@vu.nl`

VRIJE UNIVERSITEIT AMSTERDAM