Time Series Models 2025
Assignment

Karim Moussa   Ilka van de Werve
Last update: January 23, 2025

Introduction

Assessment

- This assignment consists of two parts, which are collected in this document
- Deadline Part 1: Friday February 21 at 23h59
- Deadline Part 2: Friday March 14 at 23h59
- The course criteria are explained on Canvas here
- You work in groups of four students, please make sure to enroll your group by Friday February 7, via `this form`)
- Students who completed the assignment last year and wish to retain their grade must contact Ilka van de Werve at `i.vande.werve@vu.nl` by the end of the first week (Friday February 7)

Submission guidelines

- Submit your solutions (both the .pdf file and the code) as a group via Canvas Assignments
- You may use any programming language (e.g., Python, R, Matlab, Ox), however, state space method packages are not allowed.
- Generative AI is not allowed for generating text, developing code and conducting data analysis. It may be used for correcting English and debugging code, but ensure you fully understand your code, as you may be asked to explain it

Additional resources

- Data from the DK-book and coding tips are available on Canvas
- Subquestions from the assignment that relate to recent study material are discussed weekly during the lectures by Ilka van de Werve
- Support for the assignments is provided by Karim Moussa. Use the Canvas Discussions boards for questions that do not require code inspection. For code-specific questions, visit the weekly (assignment) office hours, as announced on Canvas

Good luck!

Part 1: Local level model

(a) Consider Chapter 2 of the DK-book, there are 8 figures for the Nile data. Write computer code that can reproduce all these figures, the data is available on Canvas. Implement it according to the set of recursions for the local level model.

Note that for this part, your report only needs to contain the replicated figures; a corresponding discussion is not required.

Some remarks for clarification:

- In this subquestion, you can use the estimates of σ_ε^2 and σ_η^2 from the DK-book.
- To clarify whether the *predicted* (a_t, P_t) or *filtered* ($a_{t|t}$, $P_{t|t}$) estimates are used: Figure 2.1 (i) and (ii) are predicted estimates, whereas Figure 2.5 (i) and (ii) are filtered estimates.
- Figures 2.3 (ii),(iv) plot standard deviations instead of variances.
- It is not possible to replicate Figure 2.4 exactly because that would require the variates that were used to create the figure. However, the results from simulation smoothing should be close. To start the simulation, you can set α_1^+ to any reasonable value, such as $\alpha_1^+ := y_1$ or $\alpha_1^+ := \mathbb{E}[\alpha_1|Y_n]$.
- In Figure 2.6 (i) the confidence interval is for $\alpha_{n+j}|Y_n$, so the variance to be used is $\bar{P}_{n+j|n}$ defined at the bottom of p.30. Note that you should **not** use $\bar{F}_{n+j|n}$; this is stated at the top of p.32, but is not consistent with Figure 2.6.
- You do not need to perfectly replicate the histograms in Figure 2.7 (ii) and Figure 2.8 (ii) and (iv) as these are dependent on the chosen bin widths; they only need to be roughly similar. Including a kernel density estimate is optional.

(b) Implement the maximum likelihood estimator for the local level model, and use it to estimate the parameters σ_ε^2 and σ_η^2 for the Nile data. To validate your implementation, check whether your estimates are close to those from the DK-book (those used for the first subquestion).

Part 2: Stochastic volatility model

For a financial asset, such as a stock or an exchange rate, denote the closing price at trading day t by p_t, with its return

$$y_t = \log(p_t \,/\, p_{t-1}) = \Delta \log p_t, \qquad t = 1, \ldots, n.$$

We consider the following stochastic volatility (SV) model for the daily log returns y_t:

$$\begin{aligned} y_t &= \mu + \sigma \exp\left(\frac{\alpha_t}{2}\right)\varepsilon_t, \qquad && \varepsilon_t \sim \mathrm{N}(0,1), \\ \alpha_{t+1} &= \phi\alpha_t + \eta_t, && \eta_t \sim \mathrm{N}(0,\sigma_\eta^2), \end{aligned} \tag{1}$$

with $\sigma, \sigma_\eta > 0$ and $0 < \phi < 1$. Since both the volatility, $\sigma \exp(\alpha_t/2)$, and the observation error, ε_t, are stochastic processes, we have a nonlinear time series model.

A common simplification that allows for approximate analysis is the quasi-maximum likelihood (QML) approach of Harvey, Ruiz and Shephard (1994, RES). The idea is to transform the observations to obtain a linear model, such that we can apply the Kalman filter and related methods to perform approximate analysis and parameter estimation. The QML-method starts by applying the data transformation

$$x_t \coloneqq \log(y_t - \mu)^2 = \log(\sigma^2) + \alpha_t + \log(\varepsilon_t^2). \tag{2}$$

In practice, μ, the mean of the log returns y_t, is typically estimated by the sample mean, and subtracting the latter from y_t is important because it generally prevents taking the logs of zeros.

The above observation equation is linear in the state, α_t, but the disturbance term, $\log(\varepsilon_t^2)$, is non-Gaussian. In particular, when we assume $\varepsilon_t \sim \mathrm{N}(0,1)$ as in (1), $\log(\varepsilon_t^2)$ follows the $\log \chi^2$ distribution, which has mean $\mathbb{E}[\log(\varepsilon_t^2)] = -1.27$ and variance $\mathbb{V}\mathrm{ar}[\log(\varepsilon_t^2)] = \pi^2/2 = 4.93$. However, to use the Kalman filter, the disturbance terms must have mean zero. Define the transformed disturbance term $\xi_t \coloneqq \log(\varepsilon_t^2) + 1.27$ and intercept term $\kappa \coloneqq \log(\sigma^2) - 1.27$. By assuming that the transformed disturbance terms are Gaussian, $\xi_t \sim \mathrm{N}(0, 4.93)$, we obtain the following approximate state space model:

$$\begin{aligned} x_t &= \kappa + \alpha_t + \xi_t, \qquad && \xi_t \sim \mathrm{N}(0, 4.93), \\ \alpha_{t+1} &= \phi\alpha_t + \eta_t, && \eta_t \sim \mathrm{N}(0,\sigma_\eta^2), \end{aligned} \tag{3}$$

with parameter vector $\psi = (\kappa, \phi, \sigma_\eta^2)'$. As the above model is linear Gaussian, the Kalman filter can be used for approximate analysis and parameter estimation.

The SV-model, its the approximating state space model and the QML-method are also discussed in the third lecture week. The data is available on Canvas.

(a) Make sure that the financial series is in returns (transform if needed, see Figure 14.5). Present graphs and descriptives (e.g., sample moments).

(b) The SV-model can be made linear by transforming the returns data to x_t as given in (2). Compute x_t and display the time series in a graph.

(c) Use the QML approach based on the linearized model in (3) with observations x_t to estimate the corresponding parameters κ, ϕ, and σ_η. Present the estimates in a table.

(d) Based on the QML parameter estimates, use the Kalman filter and smoother to create the following two plots:

Plot 1: Contains both the *filtered* $\left(\mathbb{E}[\alpha_t|x_1, \ldots, x_t]\right)$ and *smoothed* $\left(\mathbb{E}[\alpha_t|x_1, \ldots, x_n]\right)$ estimates of α_t.

Plot 2: Contains the smoothed signals $\mathbb{E}[\theta_t|x_1, \ldots, x_n]$, with $\theta_t = \kappa + \alpha_t$, together with the transformed data x_t for $t = 1, \ldots, n$.

(e) We return to the original SV model in (1) (so **not** the linearized form in (3)). Compute the mode of the smoothing density $p(\alpha|Y_n)$, where $\alpha = (\alpha_1, \ldots, \alpha_1)$, using mode estimation (e.g., DK-book Section 10.6.5, but be careful with the typo mentioned in this list of errata). Compare the mode with the earlier *smoothed* QML estimates of α_t in a plot.

(f) We stay with the original SV model in (1) (so **not** the linearized form in (3)). Compute the filtered estimates of α_t in equation (1) using the bootstrap filter (e.g., DK-book Section 12.4), and compare it with the earlier *filtered* QML estimates of α_t in a graph.

Remark 1. *For subquestions e and f, note that the estimate of σ in the original SV model in* (1) *can be obtained from the estimate of κ.*

Remark 2. *In Part 2, you are expected to **discuss** and **interpret** your results:*

1. ***Not sufficient**: "Fig 2. contains the filtered and smoothed states [end of discussion]."*

2. ***Better**: "Fig 2. contains the filtered and smoothed states. It can be seen that [comment on salient aspect of results]."*

3. ***Excellent**: "Fig 2. contains the filtered and smoothed states. It can be seen that [comment on salient aspect of results]. This was expected/unexpected because [insert sound argument]."*