Time Series Models 2023
Assignment

Gabriele Mingoli   Karim Moussa   Ilka van de Werve
Last update: March 17, 2023

Introduction

- The assignment consists of two parts, which are both collected in this document.
- You work in groups of 4 students (make sure to enroll **by Friday February 10** via `https://forms.gle/6r72Q4YoA5jmjGCSA`).
- Students who did the assignments last year, should contact Ilka van de Werve **by Friday February 10** (`i.vande.werve@vu.nl`).
- <u>**Deadlines:**</u> Monday Febuary 27 at 23h59 (Part 1) and Friday March 17 at 23h59 (Part 2).
- As a group, hand in the solutions (.pdf and code) via Canvas Assignments.
- You can use any programming language (e.g. `Python`, `R`, `Matlab`, `Ox`), but packages related to state space methods are **not** allowed.
- Data from the DK-book and from other sources can be found at Canvas Files.
- Support for the assignments is given by Karim Moussa. The Canvas Discussions board can be used for all questions regarding the assignment that do not require an inspection of your code. The latter sort of questions can be asked during the weekly (assignment) office hours as announced via Canvas.
- Part 1 is graded as either a "pass" or a "fail," where the latter means that the work must be amended in order to earn a pass.
- Given a pass for Part 1, the assignment grade is determined as the grade for Part 2 (a number between 1 and 10), with `FeedbackFruits`.

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. Implement it according to the set of recursions for the local level model.[1] 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.

[1] Please note that these equations are a result of declaring $Z_t = T_t = R_t = 1$, $d_t = c_t = 0$, $H_t = \sigma_\varepsilon^2$ and $Q_t = \sigma_\eta^2$ in the general state space model. It may be worthwhile to implement the general version of the univariate Kalman filter, so that you can re-use this part of your code in Part 2.

Part 2: Stochastic volatility model

Denote closing price at trading day t by P_t, with its return

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

We consider the following stochastic volatility (SV) model for the daily log returns y_t:[2]

$$\begin{aligned} y_t &= \mu + \exp\left(\frac{\xi + \widetilde{h}_t}{2}\right)\varepsilon_t, \qquad && \varepsilon_t \sim N(0, 1), \\ \widetilde{h}_{t+1} &= \phi \widetilde{h}_t + \eta_t, && \eta_t \sim N(0, \sigma_\eta^2), \end{aligned} \tag{1}$$

with $\sigma_\eta > 0$ and $0 < \phi < 1$. Since both the volatility $\sigma_t = \exp((\xi + \widetilde{h}_t)/2)$ and the observation error ε_t are stochastic processes, we have a nonlinear time series model. Some background information on this SV model can be found in the appendix.

A common simplification that allows for approximate analysis is the quasi maximum likelihood (QML) approach of Harvey, Ruiz, and Shephard (1994). This method starts by applying the data transformation

$$x_t = \log(y_t - \mu)^2. \tag{2}$$

Note that μ, the mean of the log returns y_t, is typically estimated by the sample mean and that subtracting the latter from y_t is important because it generally prevents taking the logs of zeros. Then, after some redefinitions we obtain the linear state space model

$$\begin{aligned} x_t &= h_t + u_t, \qquad && u_t \sim p(u_t), \\ h_{t+1} &= \omega + \phi h_t + \eta_t, && \eta_t \sim N(0, \sigma_\eta^2), \end{aligned} \tag{3}$$

where $u_t = \log \varepsilon_t^2$, $\omega = (1 - \phi)\xi$ and $h_t = \widetilde{h}_t + \xi$. The disturbance u_t is not Gaussian: when we assume ε_t is Gaussian, u_t is generated from a $\log \chi^2$ distribution, which has mean -1.27 and variance $\pi^2/2 = 4.93$ (mean adjustment and fixed variance). By assuming that u_t is Gaussian with the mean and variance set to those from the $\log \chi^2$ distribution, we can apply the Kalman filter and related methods to perform approximate analysis and parameter estimation.

- Use the SV-data of the DK-book
  - (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).

[2]**Remark**: as of March 3, 2023, the notation in model (1) has been slightly adjusted: $H_t \to \widetilde{h}_t$, to prevent confusion with the variance matrix H_t in the state space model from DK. Please keep this in mind when considering the related lecture slides (e.g., week 3) or earlier versions of the assignment.

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

(c) Use the QML approach based on the linearized model in (3) with observations x_t to estimate the corresponding parameters ω, ϕ, and σ_η. This means you must assume that u_t is Gaussian with appropriate mean and variance, so that the Kalman filter can be used for (approximate) maximum likelihood estimation. Present the estimates in a table.

(d) Take the QML-estimates as your final estimates. Compute the smoothed values of h_t based on the approximate model for x_t in equation (3) by using the Kalman filter and smoother, and present them in a graph along with the transformed data x_t. In addition, present both the *filtered* $\left(\mathbb{E}[\widetilde{h}_t|x_1, \ldots, x_t]\right)$ and *smoothed* $\left(\mathbb{E}[\widetilde{h}_t|x_1, \ldots, x_n]\right)$ estimates of $\widetilde{h}_t$ in a graph.

(e) Extension 1: For a period of at least five years, consider the daily returns for S&P500 index (or another stock index) that you can obtain from Yahoo Finance, and re-visit the analysis from questions (a) - (d).[3]

Next, to improve the performance of the SV model, you can extend your analysis with a Realized Volatility measure of your choice that can be be found on Canvas. For this purpose, consider the extended model

$$\begin{aligned} x_t &= \beta \log \mathrm{RV}_t + h_t + u_t, \\ h_{t+1} &= \omega + \phi h_t + \eta_t, \end{aligned} \tag{4}$$

where β is the regression coefficient and RV_t is the realized volatility measure of your choice. How does the analysis above change with the inclusion of RV? Implement the procedure and interpret your results.

Remark 1. *The realized volatility indices file contains several columns that correspond to different RV measures, as well as other related data. For example, the column "`rv5`" contains the 5-minute realized variance. The suffix "`_ss`" indicates that the measure was computed using subsampling, more information on which can be found on p. 3 of this paper. Lastly, note that the data set also contains the daily prices of the indices in the column "`close_price`", so you do not have to do any merging of data sets.*

(f) Extension 2: We return to the original SV model in (1) (so **not** the linearized form in (3), and **without** RV). Compute the filtered estimates of $\widetilde{h}_t$ in equation (1) using

[3]This means that you have to re-generate all corresponding results and provide a corresponding discussion.

the bootstrap filter (e.g., Durbin & Koopman, 2012, Ch.12.4) and compare it with the earlier *filtered* QML estimates of $\widetilde{h}_t$ in a graph. Do this for the original data set and repeat it for the stock index of Extension 1.

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]."*

References

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (Vol. 38). OUP Oxford.

Harvey, A., Ruiz, E., & Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, *61*(2), 247–264.

Appendix: background information on the SV model

The log price $\log P_t$ can be regarded as a discretized realisation from a continuous-time process given by

$$\mathrm{d}\log P(t) = \mu\,\mathrm{d}\,t + \sigma(t)\,\mathrm{d}\,W(t),$$

where μ is the mean-return, $\sigma(t)$ is a continuous volatility process and $W(t)$ is standardised Brownian motion. We concentrate on the volatility process and we let $\log\sigma(t)^2$ follow a so-called Ornstein-Uhlenbeck process

$$\log\sigma(t)^2 = \xi + \widetilde{h}(t), \qquad \mathrm{d}\,\widetilde{h}(t) = -\lambda\widetilde{h}(t)\,\mathrm{d}\,t + \sigma_\eta\,\mathrm{d}\,B(t),$$

where ξ is constant, $0 < \lambda < 1$, σ_η is the "volatility-of-volatility" coefficient (strictly positive) and $B(t)$ is standardised Brownian motion, independent of $W(t)$.

The general framework can lead to a statistical model for the daily returns y_t. By applying the Euler-Maruyama discretisation method, we obtain the SV model in (1).