Time Series Models 2025
Assignment Notes

Karim Moussa
Last update: February 3, 2025

Introduction

These notes provide several suggestions for making the assignment of the Time Series Models course. To start with some general advice on group work, an efficient strategy is to proceed in the following order:

1. Read all the related sections from the book Durbin and Koopman (2012, DK hereafter). For Part 1 of the assignment, it is Chapter 2 of DK.
2. Read the assignment and write down all the results that you are required to generate. These are the requirements of your program.
3. Design your program: think and discuss in your group how you can structure your program to generate all the required results. What is the best solution?
4. Allocate parts of the program to the members in your group to implement.
5. Start implementing your part.
6. Combine and check all parts. Next, generate the results, and write the report. (Note that for Part 1 the report should only contain the requested figures; for Part 2 a discussion and interpretation of the results is required.)

Notice that not a single character is coded until step 5; indeed starting without having made explicit your goal (the requirements) or your design (what you are going to create) almost always results in having to redo your work.

Lastly, a question that always arises is whether a particular package may be used. The answer is "**no**" if it concerns functionality directly related to state space methods, and yes otherwise. A rule of thumb: if a formula is given on the slides or in the book, you have to program it yourself. For example, the Kalman filter recursions and the log likelihood have to be implemented, but a package can (and should) be used for the performing the corresponding numerical optimization and for creating plots.

Part 1: Local level model

The first part of the assignment is concerned with applying the Kalman filter and smoother to the Gaussian local level (LL) model,

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t, & \varepsilon_t &\sim \mathrm{N}(0, \sigma_\varepsilon^2), \\ \alpha_{t+1} &= \alpha_t + \eta_t, & \eta_t &\sim \mathrm{N}(0, \sigma_\eta^2), \end{aligned}$$

for the Nile data, where $\alpha_t = \mu_t$ is the state; see also the lecture slides of Week 1.

(a)

- To recreate the figures from DK Chapter 2, note that the quantities related to 1-period *prediction* are typically omitted for $t = 1$ (e.g. $a_1 = \mathbb{E}[\alpha_1], v_1 = y_1 - a_1, P_1 = \mathbb{V}\mathrm{ar}[\alpha_1]$, etc.) because they would otherwise come to dominate the plots due to diffuse initialization. This pertains to Figure 2.1, 2.6, and 2.7.
- When applying the Kalman smoother (Eq. 2.37 and 2.43), note that the quantities r_0 and N_0 have to be computed for the smoothed state $\hat{\alpha}_1$ and variance V_1 at $t = 1$, respectively. The output of the Kalman smoother should therefore be $\hat{\alpha}_t$ and V_t for $t = n, \ldots, 1$, as well as r_t and N_t for $t = n, \ldots, 0$. You may choose to include or omit the values r_0 and N_0 in Figure 2.2.
- For Figure 2.5, it is required to adjust the Kalman filter and smoother to deal with missing data. The neatest way to do this is by setting your missing data to NaN values, and to include an if-clause in the recursions to deal with such values. If y_t is missing, the adjustment to the filtering update step becomes

  $$\begin{aligned} a_{t|t} &= a_t, & P_{t|t} &= P_t, \\ a_{t+1} &= a_t, & P_{t+1} &= P_t + \sigma_\eta^2, \end{aligned}$$

  and the smoothing update step requires the following adjustment for r_t and N_t:

  $$r_{t-1} = r_t, \qquad N_{t-1} = N_t.$$

  See DK Section 4.10 for an explanation for general state space models.

  Remark 1. *The book also mentions a trick to deal with missing data by "simply" setting $K_t := 0$, but this also requires that v_t is set to some arbitrary number (and not to a NaN, which is safer) and that in the smoothing recursions (2.37) and (2.43) the inverse of F_t is computed by $F_t^{-1} = K_t/P_t$. However, experience has shown that it is easier to just use the above adjustments to the update steps instead.*

- For Figure 2.7, the standardised one-step forecast errors are used, which are defined in Eq. (2.65) on p.38 of DK. For Figure 2.8, the standardised smoothed residuals are used, which are defined at the bottom of p.39 in DK.

- For several figures it is required to compute confidence infervals (CIs). For example, to compute the 90%-CI in Figure 2.1 (i), which corresponds to significance level $c = 1 - 90\% = 10\%$, you need to use the formula

$$a_t \pm z_{1-c/2} \cdot \sqrt{P_t},$$

  with $z_{1-c/2} = z_{95\%}$ the 95% quantile of the standard normal distribution. To understand the above CI formula, note that an x% CI is defined as *an interval determined in such a way that if a large number of samples would be drawn, the interval would contain the true parameter approximately x-percent of the time.*

  Remark 2. *The aspect that is probably new compared to your earlier courses is that the CI is for the* ***random*** *parameter α_t instead of a* ***fixed*** *parameter. This means that an x% CI is simply an interval that contains α_t with probability x. The above formula then arises because we know that $\alpha_t|Y_{t-1} \sim \mathrm{N}(a_t, P_t)$, which implies that*

$$P(\alpha_t \in (a_t - z_{95\%} \cdot \sqrt{P_t}, a_t + z_{95\%} \cdot \sqrt{P_t})|Y_{t-1}) = 90\%.$$

(b)

To perform the maximum likelihood estimation, note that you may use any of the variants described in Chapter 2.10 of DK, which should all result in estimates close to those reported in the book, where "close" is meant in a relative sense (percentage differences).

- Most optimization algorithms require a starting value for the parameter vector $\theta = (\sigma_\varepsilon^2, \sigma_\eta^2)'$.[1] It is important to use a reasonable estimate here because starting too far from the optimum may cause the algorithm to diverge.

  A simple strategy that can always be used is to compute the objective function (in this case the log likelihood) for several possible values of the parameter vector θ, and use the best candidate as starting value. You can typically find out what a reasonable range is for θ by looking at related papers in which the model of interest is estimated, or in this case, by considering the estimates from the book.

  Another strategy that often gives reasonable starting values is the method of moments, which works as follows. For the local level model, there are two parameters

[1] Note that if you use the concentrated log likelihood from Section 2.10.2 in DK, the parameter vector is instead a scalar, $\theta = q = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$.

that we would like to estimate. We can also derive expressions for two moments of (the first differences of) the data in terms of the parameters; for details see the slides of Week 1. For example,

$$\begin{aligned} \mathbb{V}\mathrm{ar}\left[\Delta y_t\right] &= \sigma_\eta^2 + 2\sigma_\varepsilon^2, \\ \mathbb{C}\mathrm{ov}\left[\Delta y_t, \Delta y_{t-1}\right] &= -\sigma_\varepsilon^2, \end{aligned}$$

gives us two equations in two unknowns, which we can solve analytically:

$$\begin{aligned} \sigma_\varepsilon^2 &= -\mathbb{C}\mathrm{ov}\left[\Delta y_t, \Delta y_{t-1}\right], \\ \sigma_\eta^2 &= \mathbb{V}\mathrm{ar}\left[\Delta y_t\right] + 2\mathbb{C}\mathrm{ov}\left[\Delta y_t, \Delta y_{t-1}\right]. \end{aligned}$$

By substituting the sample variance and sample autocovariance for $\mathbb{V}\mathrm{ar}\left[\Delta y_t\right]$ and $\mathbb{C}\mathrm{ov}\left[\Delta y_t, \Delta y_{t-1}\right]$, respectively, in the expressions above, we get an estimate of σ_η^2 and σ_ε^2. These are called the method of moments estimates.

- If your maximum likelihood estimates are not close to the ones from the book (despite using decent starting values), the problem may be either in the log likelihood, or in the optimization thereof. It is therefore helpful to check whether your log likelihood is correct, which can be done by comparing the likelihood based on (2.60) or (2.63) with Table 2.1 from DK. Note that the log likelihood values there omit the constant terms from expression (2.63), so

$$LL = LL_{DK} + \frac{n}{2}\log 2\pi + \frac{n-1}{2}$$

should be used to compare your values with the table, with LL your computed log likelihood value and LL_{DK} the values in Table 2.1. Note also that if you are using expression (2.60) to compute the log likelihood, only the last row in the table can be used for comparison because the parameter values there correspond to the reported estimates $\widehat{\sigma}_\varepsilon^2 = 15099$ and $\widehat{\sigma}_\eta^2 = 1469.1$ (for the other rows the corresponding values for $\widehat{\sigma}_\varepsilon^2$ and $\widehat{\sigma}_\eta^2$ are not given).

References

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