Time Series Analysis

5. State space models and Kalman filtering

Andrew Lesniewski

Baruch College
New York

Fall 2017

Outline

State space models

- A *state space model* is a time series model in which the time series Y_t is interpreted as the result of a noisy observation of a stochastic process X_t.
- The values of the variables X_t and Y_t can be continuous (scalar or vector) or discrete.
- Graphically, a state space model is represented as follows:

$$
\begin{array}{cccccccccc}
X_0 & \to & X_1 & \to & \ldots & \to & X_t & \to & \ldots & \text{state process} \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \ldots & \\
Y_0 & & Y_1 & & \ldots & & Y_t & & \ldots & \text{observation process}
\end{array}
\tag{1}
$$

- State space model have been successfully applied in many fields to solve a broad range of problems.
- Our discussion of state space models follows [1].

State space models

- It is usually assumed that the state process X_t is Markovian, i.e. X_t depends on the history only through X_{t-1}, and Y_t depends only on X_t:

$$\begin{aligned} X_t &\sim p(X_t|X_{t-1}), \\ Y_t &\sim p(Y_t|X_t). \end{aligned} \tag{2}$$

- The most well studied state space model is the *Kalman filter*, for which the processes above are linear and and the sources of randomness are Gaussian.
- Namely, a *linear state space model* has the form:

$$\begin{aligned} X_{t+1} &= GX_t + \varepsilon_t, \\ Y_t &= HX_t + \eta_t. \end{aligned} \tag{3}$$

Here, the *state vector* $X_t \in \mathbb{R}^r$ is possibly unobservable and it can be observed only through the *observation vector* $Y_t \in \mathbb{R}^n$. The matrices $G \in \mathrm{Mat_n}(\mathbb{R})$ and $H \in \mathrm{Mat_{n,r}}(\mathbb{R})$ are assumed to be known. We assume that the distribution of the initial value X_1 is known and Gaussian.

State space models

- The vectors of disturbancies $\varepsilon_t \in \mathbb{R}^r$ and $\eta_t \in \mathbb{R}^n$ satisfy

$$\begin{aligned} \mathsf{E}(\varepsilon_t \varepsilon_s^{\mathrm{T}}) &= \delta_{ts} Q, \\ \mathsf{E}(\eta_t \eta_s^{\mathrm{T}}) &= \delta_{ts} R, \end{aligned} \tag{4}$$

  where δ_{ts} denotes Kronecker's delta, and where Q and R are positive definite (covariance) matrices.
- We also assume that the components of ε_t and η_s are independent of each other for all t and s.
- The first of the equations in (3) is called the *state equation*, while the second one is referred to as the *observation equation*.
- Many time series models, including the models that we have discussed so far, can be represented as state space models.

$AR(2)$ as a state space model

- As a first example, consider the $AR(2)$ model discussed in Lecture Notes #1:

$$X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \varepsilon_t. \tag{5}$$

- We can write the dynamics above in the matrix form:

$$\begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} = \begin{pmatrix} \beta_1 & \beta_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ X_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix} \tag{6}$$

- This is the state equation

$$Z_t = GZ_{t-1} + u_t$$

for the two-dimensional state variable

$$Z_t = \begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix}$$

and transition matrix

$$G = \begin{pmatrix} \beta_1 & \beta_2 \\ 1 & 0 \end{pmatrix}.$$

$AR(2)$ as a state space model

- The two-dimensional disturbance is given by

$$u_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix} .$$

- The observation $Y_t = X_t$ can be written in the form of an observation equation with zero observation noise:

$$Y_t = FZ_t ,$$

where

$$F = \begin{pmatrix} 1 & 0 \end{pmatrix} .$$

$MA(1)$ as a state space model

- The $MA(1)$ process

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1} \tag{7}$$

can be written as a state space model as follows. We set

$$\begin{aligned} Z_t &= \begin{pmatrix} X_t \\ \theta\varepsilon_t \end{pmatrix}, \\ G &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \\ F &= \begin{pmatrix} 1 & 0 \end{pmatrix}, \\ u_t &= \begin{pmatrix} \varepsilon_t \\ \theta\varepsilon_t \end{pmatrix}, \\ Y_t &= X_t. \end{aligned} \tag{8}$$

- Then

$$\begin{aligned} Z_t &= GZ_{t-1} + u_t, \\ Y_t &= FZ_t. \end{aligned}$$

$ARMA(p,q)$ as a state space model

- In general, the $ARMA(p,q)$ model

$$X_t = \beta_1 X_{t-1} + \ldots + \beta_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \tag{9}$$

can be put in the form of a state space model.

- To this end, we define $m = \max(p, q+1)$, and define the state variable

$$Z_t = \begin{pmatrix} X_t \\ \beta_2 X_{t-1} + \ldots + \beta_m X_{t-m+1} + \theta_1 \varepsilon_t + \ldots + \theta_{m-1} \varepsilon_{t-m+2} \\ \beta_3 X_{t-2} + \ldots + \beta_m X_{t-m+2} + \theta_2 \varepsilon_t + \ldots + \theta_{m-1} \varepsilon_{t-m+3} \\ \vdots \\ \beta_m X_{t-1} + \theta_{m-1} \varepsilon_t \end{pmatrix},$$

and the observation $Y_t = X_t$. Then

$$\begin{aligned} Z_t &= G Z_{t-1} + u_t, \\ Y_t &= F Z_t. \end{aligned}$$

$ARMA(p, q)$ as a state space model

- The remaining quantities are defined as follows:

$$G = \begin{pmatrix} \beta_1 & 1 & 0 & \dots & 0 \\ \beta_2 & 0 & 1 & \dots & 0 \\ \beta_3 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \beta_m & 0 & 0 & \dots & 0 \end{pmatrix},$$

$$F = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix},$$

$$u_t = \begin{pmatrix} \varepsilon_t \\ \theta_1 \varepsilon_t \\ \theta_2 \varepsilon_t \\ \vdots \\ \theta_{m-1} \varepsilon_t \end{pmatrix}.$$

- It is also possible to put $ARIMA(p, d, q)$ in the state space form (see [1]).

Inference for state space models

- Our broad goal is to make inference about the states X_t based on a set of observations $Y_1, \ldots, Y_T$.
- Three questions are of particular interest:
  (i) *Filtering:* $t = T$. What can we infer about the current state of the system based on all available observations?
  (ii) *Smoothing:* $t < T$. What can be inferred about the system based on the information contained in the entire data sample? In particular, how can we back fill missing observations?
  (iii) *Forecasting:* $t > T$. What is the optimal prediction of a future observation?

Local level model

- We will analyze in detail the simplest linear state space model, namely the *local level model* specified as follows:

$$\begin{aligned} X_{t+1} &= X_t + \varepsilon_t, \\ Y_t &= X_t + \eta_t, \end{aligned} \tag{10}$$

  for $t = 1, 2, \ldots,$, where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, and $\eta_t \sim N(0, \sigma_\eta^2)$. All the ε_t's and η_t's are assumed mutually independent.
- As mentioned before, we interpret the variables X_t as unobservable internal *states* of the system, while the variables Y_t are the *observations*.
- In principle, any inference for this model can be done using the standard methods of multivariate statistics. However, these methods lead to cumbersome and computationally intractable expressions, and are thus impractical.
- Instead, we will develop a tractable (and widely used) recursive approach to the problem.

Filtering

- The purpose of filtering is to update the knowledge of the system each time a new observation is made.
- We define:

$$\begin{aligned} \mu_t &= \mathrm{E}(X_t | Y_{1:t-1}), \\ P_t &= \mathrm{Var}(X_t | Y_{1:t-1}), \end{aligned} \tag{11}$$

  and

$$\begin{aligned} \mu_{t|t} &= \mathrm{E}(X_t | Y_{1:t}), \\ P_{t|t} &= \mathrm{Var}(X_t | Y_{1:t}), \end{aligned} \tag{12}$$

  Our first objective is to compute the *filtered estimator* $\mu_{t|t}$, its variance $P_{t|t}$, and the variance P_{t+1} of the *one period prediditor* μ_{t+1}, when the observation Y_{t+1} is made.
- Let $v_t = Y_t - \mu_t$ be the one period prediction error.

Filtering

- The prediction errors are often referred to as *innovations*. One can show (Homework Assignment #5) that v_t's are mutually independent.
- Then, for $t = 2, 3, \ldots,$

$$\begin{aligned}
\mathrm{E}(v_t | Y_{1:t-1}) &= \mathrm{E}(X_t + \eta_t - \mu_t | Y_{1:t-1}) \\
&= 0, \\
\mathrm{Var}(v_t | Y_{1:t-1}) &= \mathrm{Var}(X_t + \eta_t - \mu_t | Y_{1:t-1}) \\
&= P_t + \sigma_\eta^2, \\
\mathrm{E}(v_t | X_t, Y_{1:t-1}) &= \mathrm{E}(X_t + \eta_t - \mu_t | X_t, Y_{1:t-1}) \\
&= X_t - \mu_t, \\
\mathrm{Var}(v_t | X_t, Y_{1:t-1}) &= \mathrm{Var}(X_t + \eta_t - \mu_t | X_t, Y_{1:t-1}) \\
&= \sigma_\eta^2.
\end{aligned} \tag{13}$$

Filtering

- Notice that if $Y_{1:t}$ is known, then also $Y_{1:t-1}$ and v_t are known; hence $p(X_t|Y_{1:t}) = p(X_t|Y_{1:t-1}, v_t)$.
- From the definition of conditional probability,

$$
\begin{aligned}
p(X_t|Y_{1:t-1}, v_t) &= \frac{p(X_t, v_t|Y_{1:t-1})}{p(v_t|Y_{1:t-1})} \\
&= \frac{p(X_t|Y_{1:t-1})p(v_t|X_t, Y_{1:t-1})}{p(v_t|Y_{1:t-1})} \\
&\sim \exp\Big(-\frac{1}{2}Q\Big).
\end{aligned}
$$

Filtering

- Here, Q is explicitly given by

$$
\begin{aligned}
Q &= \frac{(X_t-\mu_t)^2}{P_t} + \frac{(v_t - X_t + \mu_t)^2}{\sigma_\varepsilon^2} - \frac{v_t^2}{P_t+\sigma_\varepsilon^2} \\
&= \Big(\frac{1}{P_t} + \frac{1}{\sigma_\varepsilon^2}\Big)(X_t-\mu_t)^2 - 2(X_t-\mu_t)\,\frac{v_t}{\sigma_\varepsilon^2} + \Big(\frac{1}{\sigma_\varepsilon^2} - \frac{1}{P_t+\sigma_\varepsilon^2}\Big)v_t^2 \\
&= \frac{P_t+\sigma_\varepsilon^2}{P_t\sigma_\varepsilon^2}\Big(X_t - \mu_t - \frac{P_t\sigma_\varepsilon^2}{P_t+\sigma_\varepsilon^2}\Big)
\end{aligned}
$$

- This implies that

$$
\begin{aligned}
\mu_{t|t} &= \mu_t + \frac{P_t}{P_t+\sigma_\eta^2}\,v_t, \\
P_{t|t} &= \frac{P_t\sigma_\varepsilon^2}{P_t+\sigma_\eta^2}\,.
\end{aligned}
\tag{14}
$$

Filtering

- Furthermore,

$$
\begin{aligned}
\mu_{t+1} &= \mathrm{E}(X_{t+1} | Y_{1:t}) \\
&= \mathrm{E}(X_t + \varepsilon_t | Y_{1:t}) \\
&= \mathrm{E}(X_t | Y_{1:t}) \\
&= \mu_{t|t}, \\
P_{t+1} &= \mathrm{Var}(X_{t+1} | Y_{1:t}) \\
&= \mathrm{Var}(X_t + \varepsilon_t | Y_{1:t}) \\
&= \mathrm{Var}(X_t | Y_{1:t}) + \sigma_\varepsilon^2 \\
&= P_{t|t} + \sigma_\varepsilon^2 .
\end{aligned}
$$

- This gives

$$
\begin{aligned}
\mu_{t+1} &= \mu_t + \frac{P_t}{P_t + \sigma_\eta^2}\, v_t, \\
P_{t+1} &= \frac{P_t \sigma_\varepsilon^2}{P_t + \sigma_\eta^2} + \sigma_\varepsilon^2 .
\end{aligned}
$$

Filtering

- Following the standard conventions, we introduce the notation

$$\begin{aligned} F_t &= \mathrm{Var}(v_t | Y_{1:t-1}) \\ &= P_t + \sigma_\eta^2, \\ K_t &= \frac{P_t}{F_t} \\ &= \frac{P_t}{P_t + \sigma_\eta^2}. \end{aligned}$$

- The quantity F_t is referred to as the variance of the prediction error, and K_t is known as the *Kalman gain*.

Kalman filter

- Using this notation, we can write the full system of recursive relations for updating from t to $t+1$ in the following form:

$$\begin{aligned} v_t &= Y_t - \mu_t, \\ F_t &= P_t + \sigma_\eta^2, \\ \mu_{t|t} &= \mu_t + K_t v_t, \\ P_{t|t} &= P_t(1 - K_t), \\ \mu_{t+1} &= \mu_t + K_t v_t, \\ P_{t+1} &= P_t(1 - K_t) + \sigma_\varepsilon^2, \end{aligned} \tag{15}$$

  for $t = 1, 2, \ldots$. We assume that the initial values μ_1 and P_1 are known.
- Relations (31) are known as the *Kalman filter* for the local level model.

State smoothing

- We can now estimate the values of the states $X_1, \ldots, X_T$ given the *entire* observation set. This is referred to as *state smoothing*.
- Since all distributions are normal, $X_t | Y_{1:T} \sim N(\widehat{X}_t, V_t)$, where $\widehat{X}_t = \mathrm{E}(X_t | Y_{1:T})$ and $V_t = \mathrm{Var}(X_t | Y_{1:T})$.
- Since the innovations are mutually independnt, we have the following linear regression:

$$\widehat{X}_t = \mu_t + \sum_{j=t}^{T} \alpha_{t,j} v_j. \tag{16}$$

- Since $\mathrm{Cov}(v_i, v_j) = 0$ its coefficients are given by

$$\alpha_{t,j} = \frac{\mathrm{Cov}(X_t, v_j)}{F_j}.$$

- We will rewrite this regression in a more usable form.

State smoothing

- To this end, let us first define

$$\begin{aligned} L_t &= 1 - K_t \\ &= \frac{\sigma_\eta^2}{F_t} . \end{aligned} \tag{17}$$

- Next, define $u_t = X_t - \mu_t$. Then $u_t = v_t - \eta_t$ and using the Kalman filter we obtain the following recursion:

$$\begin{aligned} u_{t+1} &= X_{t+1} - \mu_{t+1} \\ &= X_t + \varepsilon_t - \mu_t - K_t v_t \\ &= u_t + \varepsilon_t - K_t(u_t + \eta_t) \\ &= L_t u_t + \varepsilon_t - K_t \eta_t . \end{aligned}$$

State smoothing

- Using these relations we find that

$$
\begin{aligned}
\mathrm{Cov}(X_t, v_t) &= \mathsf{E}\big(u_t(u_t + \eta_t)\big) \\
&= \mathrm{Var}(X_t) \\
&= P_t \\
\mathrm{Cov}(X_t, v_{t+1}) &= \mathsf{E}\big(u_t(u_{t+1} + \eta_{t+1})\big) \\
&= \mathsf{E}\big(u_t(L_t u_t + \varepsilon_t - K_t \eta_t)\big) \\
&= L_t P_t \\
\mathrm{Cov}(X_t, v_{t+2}) &= L_t L_{t+1} P_t \\
&\dots \\
\mathrm{Cov}(X_t, v_T) &= L_t L_{t+1} \dots L_{T-1} P_t.
\end{aligned}
$$

- This allows us to write:

$$
\alpha_{t,j} = L_t L_{t+1} \dots L_{j-1} \frac{P_t}{F_j}.
$$

State smoothing

- Going back to (16), we can write it as

$$\widehat{X}_t = \mu_t + P_t r_{t-1}, \tag{18}$$

where

$$r_t = \frac{v_{t+1}}{F_{t+1}} + L_{t+1}\frac{v_{t+2}}{F_{t+2}} + L_{t+1}L_{t+2}\frac{v_{t+3}}{F_{t+3}} + \ldots + L_{t+1}L_{t+2}\ldots L_{T-1}\frac{v_T}{F_T} \tag{19}$$

is the weighted average of innovations after t. By definition,

$$r_T = 0. \tag{20}$$

- Notice that r_t satisfies the following recursion:

$$r_{t-1} = \frac{v_t}{F_t} + L_t r_t, \tag{21}$$

along with the terminal condition (20). Equations (21) and (18) are referred to as the *state smoothing recursion*.

Missing data

- The recursive approach to the local level model lends itself well to the situations where some observations are missing.
- Consider a series of observations Y_t, with a segment of data for $\tau \leq t < \tau_1 \leq T$ missing.
- In this case, we denote by

$$\bar{\varepsilon}_t = \sum_{j=\tau}^{t} \varepsilon_j$$

the accumulated observation noise, and notice that

$$\begin{aligned} \mathsf{E}(\bar{\varepsilon}_t | Y_{1:\tau-1}) &= 0, \\ \mathrm{Var}(\bar{\varepsilon}_t | Y_{1:\tau-1}) &= (t-\tau+1)\sigma_\varepsilon^2. \end{aligned}$$

Missing data

- As a consequence,

$$
\begin{aligned}
\mathrm{E}(X_t|Y_t) &= \mathrm{E}(X_t|Y_{1:\tau-1}) \\
&= \mathrm{E}(X_\tau + \bar{\varepsilon}_{t-1}|Y_{1:\tau-1}) \\
&= \mu_\tau, \\
\mathrm{E}(X_{t+1}|Y_t) &= \mathrm{E}(X_{t+1}|Y_{1:\tau-1}) \\
&= \mathrm{E}(X_\tau + \bar{\varepsilon}_t|Y_{1:\tau-1}) \\
&= \mu_\tau, \\
\mathrm{Var}(X_t|Y_t) &= \mathrm{Var}(X_t|Y_{1:\tau-1}) \\
&= \mathrm{Var}(X_\tau + \bar{\varepsilon}_{t-1}|Y_{1:\tau-1}) \\
&= P_\tau + (t-\tau)\sigma_\varepsilon^2, \\
\mathrm{Var}(X_{t+1}|Y_t) &= \mathrm{Var}(X_{t+1}|Y_{1:\tau-1}) \\
&= \mathrm{Var}(X_\tau + \bar{\varepsilon}_t|Y_{1:\tau-1}) \\
&= P_\tau + (t-\tau+1)\sigma_\varepsilon^2.
\end{aligned}
$$

Missing data

- We then proceed recursively:

$$\begin{aligned} \mu_{t|t} &= \mu_t, \\ P_{t|t} &= P_t, \\ \mu_{t+1} &= \mu_t, \\ P_{t+1} &= P_t + \sigma_\varepsilon^2, \end{aligned} \tag{22}$$

  for $t = \tau, \tau + 1, \ldots, \tau_1 - 1$. The recursion for the remaining values of t (less that τ and greater than $\tau_1 - 1$ are given, as before, by the Kalman filter (31).
- This procedure can, of course be extended to the case where several segments of observations are missing.
- We can use these relations to extend the data smoothing algorithm to the case of missing data. Namely, at the missing points $t = \tau, \ldots, \tau_1 - 1$ we have

$$\begin{aligned} v_t &= u_t + \eta_t, \\ u_{t+1} &= u_t + \varepsilon_t, \end{aligned} \tag{23}$$

  since $Kt = 0$, and so $L_t = 1$.

Missing data

- The covariances between the states at the missing time points and the innovations after the missing period are

$$\begin{aligned} \mathrm{Cov}(X_t, v_\tau) &= P_t, \\ \mathrm{Cov}(X_t, v_j) &= P_t L_\tau L_{\tau+1} \ldots L_{j-1}, \end{aligned} \tag{24}$$

  for $j = \tau + 1, \ldots, T$, $t = \tau, \ldots, \tau - 1$.

- The regression (16) for the missing time points is

$$\widehat{X}_t = \mu_t + \sum_{j=\tau}^{T} \alpha_{t,j} v_j, \tag{25}$$

  $j = \tau, \ldots, \tau_1 - 1$.

Missing data

- As a consequence, we find that for all $t = \tau, \ldots, \tau_1 - 1$,

$$\begin{aligned} r_{t-1} &= r_t, \\ \widehat{X}_t &= \mu_t + P_t r_{t-1}. \end{aligned} \tag{26}$$

- We conclude that smoothing data sets with missing data is quite easy: we run the original smoothing recursion, and substitute $K_t = 0$ and $L_t = 1$ at the missing time points.

Forecasting

- The third piece of inference that we wish to discuss is *forecasting*. Following Lecture Notes #1, we focus on MSE forecasts, which are given as conditional expected values.
- Specifically, let $Y^*_{t+k} = \mathrm{E}(Y_{t+k}|Y_{1:t})$ be k-period MSE forecast of the observation, given the past history $Y_{1:t}$, and let F^*_{t+k} denote its variance, $F^*_{t+k} = \mathrm{Var}(Y_{t+k}|Y_{1:t})$.
- Forecasting for the local level model turns out to be straightforward: we apply the filtering recursion to the observations $Y_1, \ldots, Y_T, Y_{T+1}, \ldots, Y_{T+k}$, and treat $Y_{T+1}, \ldots, Y_{T+k}$ as missing data points, i.e. take $K_t = 0$ in (31).
- Denote $\mu^*_{t+k} = \mathrm{E}(X_{t+k}|Y_{1:t})$ and $P^*_{t+k} = \mathrm{Var}(X_{t+k}|Y_{1:t})$. Recursions (22) with $\tau = T + 1$ and $\tau_1 = T + k$ imply that

$$
\begin{aligned}
\mu^*_{T+j+1} &= \begin{cases} \mu_{T+1}, & \text{if } j = 0, \\ \mu^*_{T+j}, & \text{otherwise,} \end{cases} \\
F^*_{t+j+1} &= \begin{cases} P_{T+1}, & \text{if } j = 0, \\ P^*_{T+j} + \sigma^2_{\varepsilon}, & \text{otherwise.} \end{cases}
\end{aligned}
$$

Forecasting

- Finally, we can write

$$
\begin{aligned}
Y^*_{T+j} &= \mathrm{E}(Y_{T+j}|Y_{1:T}) \\
&= \mathrm{E}(X_{T+j}|Y_{1:T}) + \mathrm{E}(\eta_{T+j}|Y_{1:T}) \\
&= \mu^*_{T+j}, \\
P^*_{T+j} &= \mathrm{Var}(Y_{T+j}|Y_{1:T}) \\
&= \mathrm{Var}(X_{T+j}|Y_{1:T}) + \mathrm{Var}(\eta_{T+j}|Y_{1:T}) \\
&= P^*_{T+j} + \sigma^2_\eta .
\end{aligned}
\tag{27}
$$

- As a result, the Kalman filter can be applied for all $t = 1, \ldots, T + k$, where we treat the data points at $t = T + 1, \ldots, T + k$ as missing. The forecasts and their variances are calculated by the Kalman filter with $K_t = 0$, for all t's following T.

Initialization of the filter

- We have left a number of parameters that have not been specified, namely:
  (i) the initial values μ_1 and P_1 that enter the probability distribution $N(\mu_1, P_1)$ of the state X_1,
  (ii) the standard deviations σ_ε and σ_η of the disturbances in the state and observation equations, respectively.
- A way to tackle (i) is to assume that X_1 has a *diffuse prior* density: we choose μ_1 arbitrarily and let $P_1 \to \infty$. Subsituting it to the Kalman filter (31) yields:

$$\begin{aligned} \mu_2 &= Y_1, \\ P_2 &= \sigma_\varepsilon^2. \end{aligned}$$

- An alternative approach to determining μ_1 and P_1 is MLE.

MLE for the local level model

- The parameters $\theta = (\sigma_\varepsilon, \sigma_\eta, \mu_1, P_1)$ can be estimated by means of MLE.
- To this end, we consider the joint probability of the observations:

$$p(Y_{1:T}|\theta) = \prod_{t=1}^{T} p(Y_t|Y_{1:t-1}), \tag{28}$$

  where $p(Y_1|Y_0) = p(Y_1)$.
- Hence,

$$-\log \mathcal{L}(\theta|Y_{1:T}) = \frac{1}{2} \sum_{t=1}^{T} \Big(\log F_t + \frac{v_t^2}{F_t} \Big) + const, \tag{29}$$

  and the value of the log liklihood function can be calculated from the Kalman filter.
- Searching for the minimum of this log liklihood function we find estimates of the parameters $\sigma_\varepsilon, \sigma_\eta, \mu_1$, and P_1.

MLE for the local level model

- It is easy to see the the $P_1 \to \infty$ limit of $\mathcal{L}(\theta | Y_{1:T})$ is explicitly given by

$$-\log \mathcal{L}(\theta | Y_{1:T}) = \frac{1}{2} \sum_{t=2}^{T} \Big(\log F_t + \frac{v_t^2}{F_t} \Big) + const. \tag{30}$$

  This gives the *diffuse log likelihood* and it can be used to estimate σ_ε and σ_η within the diffuse prior approach.

- An alternative approach, which we will discuss later, is Bayesian inference. In this approach, σ_ε and σ_η are *hyperparameters* of the model.

Kalman filter

- We now go back to the general linear state space model (3).
- The analysis of (3) follows the same steps as the calculations we did in detail for the local level model. It is a bit more tedious (which is why we decided to focus on the local level model), and can be found in [1]. Below we just summarize the results.
- As before, we use the following notation:
  (i) $\mu_{t|t} = \mathrm{E}(X_t | Y_{1:t})$
  (ii) $P_{t|t} = \mathrm{Var}(X_t | Y_{1:t})$
  (iii) $\mu_{t+1} = \mathrm{E}(X_{t+1} | Y_{1:t})$
  (iv) $P_{t+1} = \mathrm{Var}(X_{t+1} | Y_{1:t})$
  (v) $v_t = Y_t - \mathrm{E}(Y_t | Y_{1:t-1})$ (innovation).
  (vi) $F_t = \mathrm{Var}(v_t | Y_{1:t-1})$
  (vii) $K_t = HP_tG^{\mathrm{T}}F_t^{-1}$ (Kalman gain)

Kalman filter

- Using this notation, we can write the full system of recursive relations for updating from t to $t+1$ in the following form:

$$\begin{aligned} v_t &= Y_t - G\mu_t, \\ F_t &= GP_tG^{\mathrm{T}} + R, \\ \mu_{t|t} &= \mu_t + P_tG^{\mathrm{T}}F_t^{-1}v_t, \\ P_{t|t} &= P_t - P_tG^{\mathrm{T}}F_t^{-1}GP_t, \\ \mu_{t+1} &= H\mu_t + K_tv_t, \\ P_{t+1} &= HP_t(H - K_tG)^{\mathrm{T}} + Q, \end{aligned} \tag{31}$$

for $t = 1, 2, \ldots$. As before, the initial values μ_1 and P_1 are assumed to be known.

State space models

- The recursion relations for state smoothing, missing data, and forecasting that we derived for the local level model can be extended to the case of the general linear state space model.
- Also, the MLE estimation (including the diffuse prior approach) of the covariance matrices of the disturbances can be extended to the general case.
- There are various extensions of the Kalman filter approach to nonlinear state space models (known as *extended Kalman filter*, *unscented Kalman filter*, etc). Essentially, they rely on approximating a nonlinear model by a linear model, and their usuability depends on the quality of thes approximations.
- The details of these developments, as well as computational issues surrounding Kalman filtering, are presented beautifully in [1].

References

Durbin, J., and Koopman, S. J.: *Time Series Analysis by State Space Methods*, Oxford University Press (2012).