State-Space Models and the Kalman Filter

Leopoldo Catania
Aarhus University and CREATES

leopoldo.catania@econ.au.dk

A brief introduction

- State-space models are models that use state variables to describe a system by a set of difference equations;
- The object of the methodology is to infer relevant properties of the state variables α_t from knowledge of the observations $y_1, ..., y_t$
- Most time series models can be written in state-space form.
- The state-space form representation allows for a straightforward modelling of additive feature of the data such as missing values, seasonal components, measurement errors and outliers.

State Space form: univariate model

Measurement equation:

$$y_t = Z\alpha_t + D\varepsilon_t, \quad t = 1, 2, \ldots, T, \qquad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2), \tag{1}$$

where Z is a $1 \times m$ matrix and D is a selection matrix.

Transition equation:

$$\alpha_{t+1} = T\alpha_t + H\eta_t, \qquad \eta_t \sim NID(0, Q), \tag{2}$$

where T is $m \times m$ and H is $m \times g$ selection matrix, and η_t is a $g \times 1$ disturbance vector. Finally Q is a $g \times g$ variance covariance matrix.

Examples: local level model

Measurement equation:

$$y_t = \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2),$$

Transition equation:

$$\alpha_{t+1} = \alpha_t + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2),$$

which gives a random-walk plus noise model.

Examples: trend model

The model is

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2) \\
\mu_{t+1} &= \mu_t + \nu_t + \zeta_t \quad \zeta_t \sim NID(0, \sigma_\zeta^2) \\
\nu_{t+1} &= \nu_t + \xi_t \quad \xi_t \sim NID(0, \sigma_\xi^2)
\end{aligned}
$$

if $\sigma_\xi^2 = \sigma_\zeta^2 = 0$, then we get

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t \quad t = 1, 2, \ldots, T, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2) \\
\mu_{t+1} &= \mu_t + \nu \quad t = 1, 2, \ldots, T,
\end{aligned}
$$

which is a deterministic trend plus noise.

Examples: TVP models

Measurement equation:

$$y_t = X_t \alpha_t + \varepsilon_t \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2),$$

Transition equation:

$$\alpha_{t+1} = \alpha_t + \eta_t, \quad \eta_t \sim NID(0, Q),$$

X_t is a $1 \times g$ set of observable regressors at time t, the states in this case are the time-varying parameters. Note the matrix Q is a full $g \times g$ matrix of parameters that need to be estimated.

Examples: AR(2)

- The model is
  $y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \xi_t$ with $\xi_t \sim \mathcal{NID}\left(0, \sigma_\epsilon^2\right)$
- Can be put into state-space form: let $\alpha_t = \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}$

Measurement equation:

$$y_t = (1, 0)\alpha_t \quad t = 1, 2, \dots, T,$$

Transition equation:

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} \varphi_1 & \varphi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \xi_t$$

Examples: AR(2) – alternative formulation

- The model is
  $y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \xi_t$ with $\xi_t \sim \mathcal{NID}\left(0, \sigma_\epsilon^2\right)$
- Can be put into state-space form: let $\alpha_t = \begin{pmatrix} y_t \\ \phi_2 y_{t-1} \end{pmatrix}$

Measurement equation:

$$y_t = (1, 0)\alpha_t \quad t = 1, 2, \dots, T,$$

Transition equation:

$$\begin{pmatrix} y_t \\ \phi_2 y_{t-1} \end{pmatrix} = \begin{pmatrix} \varphi_1 & 1 \\ \varphi_2 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \phi_2 y_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \xi_t$$

Examples: MA(1)

- The model is
  $y_t = \xi_t + \theta\xi_{t-1}$ with $\xi_t \sim \mathcal{NID}\left(0, \sigma_\epsilon^2\right)$
- Can be put into state-space form: let $\alpha_t = \begin{pmatrix} y_t \\ \theta\xi_t \end{pmatrix}$

Measurement equation:

$$y_t = (1, 0)\alpha_t \quad t = 1, 2, \ldots, T,$$

Transition equation:

$$\begin{pmatrix} y_t \\ \theta\xi_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \theta\xi_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ \theta \end{pmatrix} \xi_t$$

Examples: ARMA(1, 1)

- The model is
  $y_t = \varphi y_{t-1} + \xi_t + \theta \xi_{t-1}$ with $\xi_t \sim \mathcal{NID}\left(0, \sigma_\epsilon^2\right)$
- Can be put into state-space form: let $\alpha_t = \begin{pmatrix} y_t \\ \theta \xi_t \end{pmatrix}$

Measurement equation:

$$y_t = (1, 0)\alpha_t \quad t = 1, 2, \dots, T,$$

Transition equation:

$$\begin{pmatrix} y_t \\ \theta \xi_t \end{pmatrix} = \begin{pmatrix} \varphi & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \theta \xi_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ \theta \end{pmatrix} \xi_t$$

Examples: ARMA(p, q)

- The model is
  $y_t = \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \xi_t + \theta_1 \xi_{t-1} + \cdots + \theta_q \xi_{t-q}$ with
  $\xi_t \sim \mathcal{NID}\left(0, \sigma_\epsilon^2\right)$
- Can be put into state-space form. Let $m = \max(p, q+1)$ and re-write the ARMA(p,q) model as:
  $y_t = \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \xi_t + \theta_1 \xi_{t-1} + \cdots + \theta_{m-1} \xi_{t-m+1}$
  where some of the AR of MA coefficients will be zero unless $p = q + 1$.
- Define:

$$\alpha_t = \begin{pmatrix} y_t \\ \phi_2 y_{t-1} + \cdots + \phi_p y_{t-m+1} + \theta_1 \eta_t + \cdots + \theta_{m-1} \eta_{t-m+2} \\ \vdots \\ \phi_m y_{t-1} + \theta_m \eta_t \end{pmatrix}$$

Examples: ARMA(p, q)

Measurement equation:

$$y_t = (1, \mathbf{0}_{m-1})\alpha_t \quad t = 1, 2, \ldots, T,$$

Transition equation:

$$\alpha_t = \begin{pmatrix} \phi_1 & 1 & 0 & 0 & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{m-1} & 0 & 0 & \cdots & 1 \\ \phi_m & 0 & 0 & \cdots & 0 \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{m-2} \\ \theta_{m-1} \end{pmatrix} \eta_t$$

Kalman filter

The Kalman filter is a recursive method, developed by Rudolf Kalman in 1960, to produce conditional expectations of the states variables given a linear Gaussian state-space.

- Kalman filter routine can be used to compute the log-likelihood function of the model when the parameters are unknown.
- Under Gaussianity and linearity the Kalman filter is the *optimal* filter, in the sense it produces the best estimate, (minimum variance), of the states.

Assumptions for standard Kalman Filter

The crucial assumptions for the Kalman filter method are:

- Linearity
- Gaussianity
- Independence between measurement errors and innovations to states. (This one can be removed after a proper reparameterization of the model.)

Deviations from these assumptions generate non-optimal filters and hence biases in the parameter estimates.

Intermezzo: the regression lemmas

Lemma 1

Consider the following two vectors, $\mathbf{x}$, and $\mathbf{y}$ jointly normally distributed

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\Sigma}_{\mathbf{xy}} \\ \boldsymbol{\Sigma}_{\mathbf{yx}} & \boldsymbol{\Sigma}_{\mathbf{yy}} \end{bmatrix}; \qquad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}} \\ \boldsymbol{\mu}_{\mathbf{y}} \end{bmatrix}$$

Then

$$E[\mathbf{x}|\mathbf{y}] = E[\mathbf{x}] + \boldsymbol{\Sigma}_{\mathbf{xy}} \boldsymbol{\Sigma}_{\mathbf{yy}}^{-1} (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})$$
$$Var[\mathbf{x}|\mathbf{y}] = \boldsymbol{\Sigma}_{\mathbf{xx}} - \boldsymbol{\Sigma}_{\mathbf{xy}} \boldsymbol{\Sigma}_{\mathbf{yy}}^{-1} \boldsymbol{\Sigma}_{\mathbf{yx}}$$

Intermezzo: the regression lemma2

Lemma 2

Consider the following three vectors, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ jointly normally distributed

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{bmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\Sigma}_{\mathbf{xy}} & \boldsymbol{\Sigma}_{\mathbf{xz}} \\ \boldsymbol{\Sigma}_{\mathbf{yx}} & \boldsymbol{\Sigma}_{\mathbf{yy}} & \mathbf{0} \\ \boldsymbol{\Sigma}_{\mathbf{zx}} & \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{zz}} \end{bmatrix}; \qquad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}} \\ \boldsymbol{\mu}_{\mathbf{y}} \\ \mathbf{0} \end{bmatrix}$$

In the multivariate normal regression we have that

$$\begin{aligned} E(\mathbf{x}|\mathbf{y},\mathbf{z}) &= E(\mathbf{x}|\mathbf{y}) + \boldsymbol{\Sigma}_{\mathbf{xz}}\boldsymbol{\Sigma}_{\mathbf{zz}}^{-1}\mathbf{z} \\ Var(\mathbf{x}|\mathbf{y},\mathbf{z}) &= Var(\mathbf{x}|\mathbf{y}) - \boldsymbol{\Sigma}_{\mathbf{xz}}\boldsymbol{\Sigma}_{\mathbf{zz}}^{-1}\boldsymbol{\Sigma}_{\mathbf{xz}}' \end{aligned}$$

Derivation of the Kalman filter (local-level model)

The model is

$$y_t = \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$$
$$\alpha_{t+1} = \alpha_t + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2).$$

The object of the filtering is to update the knowledge of the state each time a new observation is brought in. Indeed, conditional on the information set up to $t-1$,

$$\alpha_t | Y_{t-1} \sim N(a_t, P_t) \tag{3}$$

If a_t and P_t are known then we can calculate a_{t+1} and P_{t+1} when y_t is brought in. In the local level model,

$$\begin{aligned} a_{t+1} &= E[\alpha_{t+1}|Y_t] = E[\alpha_t + \eta_t|Y_t] = E[\alpha_t|Y_t] \\ P_{t+1} &= Var[\alpha_{t+1}|Y_t] = Var[\alpha_t + \eta_t|Y_t] = Var[\alpha_t|Y_t] + \sigma_\eta^2 \end{aligned}$$

The starting values a_1 and P_1 must be fixed.

Derivation of the Kalman filter (local-level model)

Define

$$v_t = y_t - a_t = y_t - E[\alpha_t | Y_{t-1}]$$

and $F_t = Var[v_t | Y_{t-1}]$ with $E[v_t | Y_{t-1}] = 0$ and $E[v_t y_{t-j}] = 0$ for $j = 1, ..., t-1$.
Hence,

$$\begin{aligned} E[\alpha_t | Y_t] &= E[\alpha_t | Y_{t-1}, v_t] \\ Var[\alpha_t | Y_t] &= Var[\alpha_t | Y_{t-1}, v_t] \end{aligned}$$

Since all variables are normally distributed, the $E[\alpha_t | Y_t]$ and $Var[\alpha_t | Y_t]$ are given by standard formulae from multivariate normal regression theory.

Derivation of the Kalman filter (local-level model)

It follows that

$$E[\alpha_t|Y_t] = E(\alpha_t|Y_{t-1}) + Cov(\alpha_t, v_t)Var(v_t)^{-1}v_t$$

where

$$\begin{aligned} Cov[\alpha_t, v_t|Y_{t-1}] &= E[\alpha_t v_t|Y_{t-1}] = E[\alpha_t(\alpha_t + \epsilon_t - a_t)|Y_{t-1}] \\ &= E[\alpha_t^2|Y_{t-1}] - E[\alpha_t|Y_{t-1}]a_t \\ &= E[\alpha_t^2|Y_{t-1}] - a_t^2 = Var(\alpha_t|Y_{t-1}) = P_t. \end{aligned}$$

For the variance we have

$$\begin{aligned} F_t &= Var[v_t|Y_{t-1}] \\ &= Var[\alpha_t + \epsilon_t - a_t|Y_{t-1}] \\ &= Var[\alpha_t|Y_{t-1}] + Var(\epsilon_t) = P_t + \sigma_\epsilon^2 \end{aligned}$$

Derivation of the Kalman filter (local-level model)

Thus

$$E[\alpha_t | Y_t] = a_t + \frac{P_t}{F_t} v_t$$

where $K_t := \frac{P_t}{F_t}$ is the *Kalman gain*. Similarly

$$\begin{aligned} Var[\alpha_t | Y_t] &= Var[\alpha_t | Y_{t-1}] - Cov(\alpha_t, v_t | Y_{t-1})^2 Var[v_t | Y_{t-1}]^{-1} \\ &= P_t - \frac{P_t^2}{F_t} \\ &= P_t(1 - K_t) \end{aligned}$$

The Kalman filter for the local level model

Finally, the set of recursions of the Kalman filter for the local level model is

$$
\begin{aligned}
v_t &= y_t - a_t, & F_t &= P_t + \sigma_\epsilon^2 \\
K_t &= \frac{P_t}{F_t} \\
a_{t+1} &= a_t + K_t v_t, & P_{t+1} &= P_t(1 - K_t) + \sigma_\eta^2
\end{aligned}
$$

The Kalman filter (General Formula)

Finally, the set of recursions of the Kalman filter for the state space model

$$y_t = Z\alpha_t + D\varepsilon_t, \qquad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2), \tag{4}$$

$$\alpha_{t+1} = T\alpha_t + H\eta_t, \qquad \eta_t \sim NID(0, Q), \tag{5}$$

is

$$\begin{aligned} v_t &= y_t - Za_t, & F_t &= ZP_tZ' + \sigma_\varepsilon^2 DD', \\ K_t &= TP_tZ'F_t^{-1}, & L_t &= T - K_tZ \\ a_{t+1} &= Ta_t + K_tv_t, & P_{t+1} &= TP_tL_t' + HQH'. \end{aligned}$$

Constructing the log-likelihood

The log-likelihood function of a model represented in state-space form can be computed within the Kalman filter routine. Indeed, at each iteration, after the prediction step, we obtain the so called *one-step ahead prediction errors* and *one-step ahead prediction variance*:

$$\begin{aligned} v_t &= Y_t - Za_t \\ F_t &= ZP_tZ' + \sigma_\varepsilon^2 DD' \end{aligned}$$

We can compute the conditional Gaussian log-likelihood function, $\log \mathcal{L}_t = \sum_{t=1}^{T} \ell_t$, where ℓ_t is the likelihood contribution at time t:

$$\log \mathcal{L}_t = -\frac{Tp}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \left(\log |F_t| + v_t' F_t^{-1} v_t \right)$$

Smoothing

Smoothing means estimation of $\alpha_1, ..., \alpha_N$ based on the entire sample.
The conditional density

$$\alpha_t | Y_N \sim N(\hat{\alpha}_t, V_t) \tag{6}$$

where the smoothed state

$$\hat{\alpha}_t = E[\alpha_t | Y_N]$$

and the smoothed state variance

$$V_t = Var[\alpha_t | Y_N]$$

The operation of calculating $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_N$ is called state smoothing.

Smoothed state

- The one-step ahead errors $v_1, \dots, v_N$ are mutually independent and are a linear transformation of $y_1, \dots, y_N$.
- The errors $(v_t, \dots, v_N)$ are independent of $(y_1, \dots, y_{t-1})$ with zero means.
- When $y_1, \dots, y_N$ are fixed, then Y_{t-1} $(v_t, \dots, v_N)$ are fixed and viceversa.
- By the properties of the multivariate normal

$$E[x|y,z] = E[x|y] + \Sigma_{xz}\Sigma_{zz}^{-1} z \tag{7}$$

- It follows that $\hat{\alpha}_t = E[\alpha_t|Y_N]$ is given by

$$\begin{aligned}
\hat{\alpha}_t &= E[\alpha_t|Y_N] = E[\alpha_t|Y_{t-1}, v_t, \dots, v_N] \\
&= E[\alpha_t|Y_{t-1}] + Cov[\alpha_t, (v_t, \dots, v_N)'] Var[(v_t, \dots, v_N)']^{-1} (v_t, \dots, v_N)' \\
&= a_t + \begin{bmatrix} Cov[\alpha_t, v_t] \\ Cov[\alpha_t, v_{t+1}] \\ \vdots \\ Cov[\alpha_t, v_N] \end{bmatrix} \begin{bmatrix} F_t & & \\ & \ddots & \\ & & F_N \end{bmatrix}^{-1} \begin{bmatrix} v_t \\ v_{t+1} \\ \vdots \\ v_N \end{bmatrix}
\end{aligned}$$

Smoothing

Hence, by the properties of the multivariate normal regression, we get

$$\hat{\alpha}_t = E[\alpha_t | Y_N] = E[\alpha_t | Y_{t-1}] + \sum_{j=t}^{N} Cov(\alpha_t, v_j) F_j^{-1} v_j$$

where

$$\begin{aligned} Cov(\alpha_t, v_t) &= P_t \\ Cov(\alpha_t, v_{t+1}) &= L_t P_t \\ Cov(\alpha_t, v_N) &= L_t L_{t+1} ... L_{N-1} P_t \end{aligned}$$

Smoothing

Therefore, by substituting

$$\begin{aligned}\hat{\alpha}_t &= a_t + P_t \frac{v_t}{F_t} + P_t L_t \frac{v_{t+1}}{F_{t+1}} + \dots \\ &= a_t + P_t r_{t-1}\end{aligned}$$

where

$$r_{t-1} = \frac{v_t}{F_t} + L_t \frac{v_{t+1}}{F_{t+1}} + ... + L_t L_{t+1} ... L_{N-1} \frac{v_N}{F_N}$$

is a weighted sum of innovations after time $t-1$ and needs to be computed by backward recursion (smoothing state recursion)

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

Smoothing States Variance

With a similar argument, we can derive the recursive formula for the smoothed state variance, $V_t = Var[\alpha_t | Y_N]$.

$$\begin{aligned} V_t &= Var[\alpha_t | Y_{t-1}] - Cov[\alpha_t, (v_t, \dots, v_N)'] Var[(v_t, \dots, v_N)']^{-1} Cov[\alpha_t, (v_t, \dots, \\ &= P_t - \sum_{j=t}^{N} Cov(\alpha_t, v_j)^2 F_j^{-1} \end{aligned}$$

so that

$$V_t = P_t - P_t^2 N_{t-1}$$

and N_t is again given by the following backward recursion

$$Var(r_{t-1}) \equiv N_{t-1} = \frac{1}{F_t} - L_t^2 N_t$$

Given V_t, we can construct confidence intervals around the smoothed states $\hat{\alpha}_t$.

Missing Values

The Kalman filter in case of missing values

$$
\begin{aligned}
&\nu_t = y_t - Za_t, && F_t = ZP_tZ' + \sigma_\varepsilon^2 DD', \\
&K_t = TP_tZ'F_t^{-1}, && L_t = T - K_tZ \\
&a_{t+1} = \begin{cases} Ta_t + K_t\eta_t, & \text{No MV} \\ Ta_t & \text{MV} \end{cases} \\
&P_{t+1} = \begin{cases} TP_tL_t' + HQ_tH^{'} & \text{No MV} \\ TP_tT' + HQ_tH^{'} & \text{MV} \end{cases}
\end{aligned}
$$

Forecast

Forecasting is an operation that comes at no-costs from the Kalman filter. We regard forecasting as filtering observations $(y_1, ..., y_t, y_{t+1}, ..., y_{t+J})$ using the Kalman filter and treating the last J observations $y_{t+1}, ..., y_{t+J}$ as missing. The prediction of the states is

$$
\begin{aligned}
\bar{a}_{t+j|n} &= E(\alpha_{t+j}|Y_t) = T^j a_{t|t-1} \\
\bar{P}_{t+j|n} &= TP_{t|t-1}L_t' + HQ_tH^{'}
\end{aligned}
$$

Estimating, Filtering and Smoothing Stochastic Volatility

Following Harvey et al. (1994) we have that the following ARSV model

$$\begin{aligned} r_t &= \sigma \exp(w_t/2) z_t \\ w_t &= \rho w_{t-1} + \eta_t \end{aligned}$$

can be estimated with the Kalman filter by working with $\zeta_t = \log(r_t^2)$

$$\zeta_t = \mu + w_t + \epsilon_t$$

where $\mu = \log(\sigma^2) + E(\log(z_t^2))$, and ϵ_t is treated as $NID(0, \sigma_\epsilon^2)$. If the model is well specified, $\sigma_\epsilon^2 = \pi^2/2$.

The State-Space Form for ARSV

Following Harvey et al. (1994) we have the following ARSV model

$$
\begin{aligned}
log(r_t^2) &= \mu + w_t + \epsilon_t \\
w_t &= \rho w_{t-1} + \eta_t
\end{aligned}
$$

which is an AR(1) plus noise process.

- The estimation can be carried out on $\bar{\zeta}_t = \log(r_t^2) - \hat{\mu}$ where $\hat{\mu}$ is the sample mean of $\log(r_t^2)$, such that we don't have to estimate σ.
- The parameter set is $\theta = [\rho, \sigma_\epsilon^2, \sigma_\eta^2]$.
- The parameter σ_ϵ^2 can also be restricted to $\pi^2/2$