Chapter 3

State-space models

3.1 The formulation of the state-space model

State-space models are a flexible family of models which is a generalization models that that used in many scenarios. It is convenient to think of the state-space models in the Gaussian setting, in which the basic derivation has a simple interpretation in terms of likelihoods and conditional likelihoods. However, these models are meaningful in non-Gaussian settings as well. A famous example of a state-space model that is used in non-Gaussian settings is the Hidden Markov Model (HMM). However, we will deal primarily with models that are best described in the Gaussian setting.

The strongest feature of state-space models is the existence of very general algorithms for forecast, filtering (estimating the current hidden state) and smoothing (estimating past hidden states).

The state-space model is a two-layer model. The external layer involves the observed process y. This process is assumed to follow the measurement equation:

$$y_t = X_t \beta_t + \epsilon_t \ . \tag{3.1}$$

For each t, y_t is a n-vector. The $n \times m$ matrix X_t is a matrix of regressors, with β_t the regression coefficients. The vectors ϵ_t are independent multinormals with zero mean and covariance Σ_t.

The internal layer is an unobserved process of regression coefficients β_t. This process is assumed to evolve like a multi-dimensional AR(1) process, in the sense that it follows the transition equation:

$$\beta_t = T_t \beta_t + \eta_t \ . \tag{3.2}$$

Here T_t is an $m \times m$ matrix and the components of white noise η_t have a multi-normal distribution with zero mean and covariance matrix Q. The process is initiated with the random vector β_0, which has a mean of a_0 and a covariance matrix of P_0.

The elements X_t, Σ, T_t and Q are referred to as the *system matrices*. If they do not vary over time the the system is said to be time-invariant or time homogeneous. The system is also stationary for a specific selection of a_0 and P_0.

Let us consider few examples of models for time-series and see how they can be formulated as a state-space model. The first three are examples of Structural Models for time series.

Example 1 (Local Level)**.** *Let* $y_t = \mu_t + \epsilon_t$*, where* $\mu_t = \mu_{t-1} + \eta_t$ *is a random walk. Clearly, this falls into the general formulation by specifying* $X_t = T_t = 1$ *and* $\beta_t = \mu_t$*. Observe that this model can also be written a ARIMA(0,1,1) model, with restriction on the values of the parameters.*

Example 2 (Local Linear Trend)**.** *This model also sets* $y_t = \mu_t + \epsilon_t$*, but now* $\mu_t = \mu_{t-1} + \nu_{t-1} + \eta_{1,t}$ *and* $\nu_t = \nu_{t-1} + \eta_{2,t}$*. This model can be written in the state-space formulation by taking* $X_t = (1, 0)$*,* $\beta_t = (\mu_t, \nu_t)'$*, and*

$$T_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}; \quad Q = \begin{pmatrix} q_{11} & 0 \\ 0 & q_{22} \end{pmatrix}.$$

This model is also a ARIMA(0,2,2) model with the appropriate restriction on the values of the parameters.

Example 3 (Basic Structural Model, BSM)**.** *This model is a local linear trend model with the addition of a seasonal component. Hence* $y_t = \mu_t + s_t + \epsilon_t$ *with* $\mu_t = \mu_{t-1} + \nu_{t-1} + \eta_{1,t}$ *and* $\nu_t = \nu_{t-1} + \eta_{2,t}$ *defined as before. The additional seasonal component satisfies* $s_t = -s_{t-1} - \cdots - s_{t-c+1} + \eta_{3,t}$*, for* c *the length of a cycle. For example, when* $c = 3$ *this model can be written in the state-space formulation by taking* $X_t = (1, 0, 1, 0, 0)$*,* $\beta_t = (\mu_t, \nu_t, s_t, s_{t-1}, s_{t-2})'$*, and*

$$T_t = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}; \quad Q = \begin{pmatrix} q_{11} & 0 & 0 & 0 & 0 \\ 0 & q_{22} & 0 & 0 & 0 \\ 0 & 0 & q_{33} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The three models above my be fit to a one-dimensional time series with the aid of the function `StructTS`. This function operates of `ts` objects and is part of the standard distribution of `R`.

In the next set of examples we return to the ARIMA modeling and present them as state-space models. These examples represent a general rule which states that any ARIMA model possesses has such a representation. In the sequel we will consider general methods for prediction and estimation that can be applied to state-space models. As a corollary we will obtain methods for estimation and prediction in ARIMA models. Such models, unlike methods of moments associated with the Yule-Walker Equations, do not rely on the assumption of stationarity.

Example 4 (The MA(1) process)**.** *Consider the $MA(1)$ process that has the form $y_t = \eta_t - \theta_1\eta_{t-1}$. One can be represent this process in a space-state formulation by taking*

$$X_t = (1,0),\ \beta_t = (y_t, -\theta\eta_t)',\ T = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},\ Q = q_{11}\begin{pmatrix} 1 & -\theta_1 \\ -\theta_1 & \theta_1^2 \end{pmatrix}.$$

Example 5 (The ARMA(1,1) process)**.** *For the ARMA(1,1) process one may reuse the quantities X_t, β_t, and Q as before and alter the transition matrix of the state equation to take the form*

$$T = \begin{pmatrix} \phi_1 & 1 \\ 0 & 0 \end{pmatrix}.$$

Example 6. *Give two different representations of the AR(2) process as a state-space model.*

3.2 The Kalman filter

The Kalman filter is an efficient recursive algorithm for the computation of the optimal estimator $\hat{\beta}_t$ of β_t, given the information up to (and including) t. A by product is the computation of the error in estimation:

$$P_t = \mathbb{E}[(\beta_t - \hat{\beta}_t)(\beta_t - \hat{\beta}_t)'].$$

Suppose that $\hat{\beta}_{t-1}$ and P_{t-1} are given at time $t-1$. The algorithm commences the recursion step by computing the predicted values of y_t, given the information available up to time $t-1$ (including):

$$\hat{y}_t = X_t\hat{\beta}_{t|t-1}\ .$$

The MSE of the innovation $\nu_t = y_t - \hat{y}_t$ is fiven by

$$F_t = X_t P_{t|t-1} X_t' + \Sigma.$$

The terms $\hat{\beta}_{t|t-1}$ and $P_{t|t-1}$ that are used are computed via the *prediction equations*:

$$\begin{aligned}\hat{\beta}_{t|t-1} &= T_t \hat{\beta}_{t-1}\\ P_{t|t-1} &= T_t P_{t-1} T_t' + Q \ .\end{aligned}$$

In the second step the the observation y_t is introduced to produce an estimate of the unobserved state and the error of estimation.This is carried out by the application of the *updating equations*:

$$\begin{aligned}\hat{\beta}_t &= \hat{\beta}_{t|t-1} + P_{t|t-1} X_t' F_t^{-1} (y_t - X_t a_{t|t-1})\\ P_t &= P_{t|t-1} - P_{t|t-1} X_t' F_t^{-1} X_t P_{t|t-1}.\end{aligned}$$

The motivation to this step is the computation of the conditional mean and variance of $\hat{\beta}_{t|t-1}$, given y_t.