## Dynamic Linear Models

Dynamic linear models (DLMs) arise via state-space formulation of standard time series models as natural structures for modeling time series with nonstationary components.

## Basic Normal DLMs

We start with describing the class of normal DLMs for univariate time series of equally spaced observations. Specifically, assume that $y_t$ is modeled over time by the equation

$$y_t = \boldsymbol{F}_t^{'}\boldsymbol{\theta}_t + \nu_t, \qquad (1)$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \quad (2)$$

with the following components assumptions:

1. $\boldsymbol{\theta}_t = \left(\theta_{t,1}, \dots, \theta_{t,p}\right)^{'}$ is the $p \times 1$ state vector at time $t$;
2. $\boldsymbol{F}_t$ is the $p$ dimensional vector of *known* constants or regressors at time $t$;
3. $\nu_t$ is the observation noise, with $N(\nu_t|0, \mathrm{v}_t)$;
4. $\boldsymbol{G}_t$ is a known $p \times p$ matrix, usually referred to as the state evolution matrix at time $t$;
5. $\boldsymbol{w}_t$ is the state evolution noise, or innovation, at time $t$, distributed as $N(\boldsymbol{w}_t|\mathbf{0}, \mathbf{W}_t)$;
6. The noise sequences $\nu_s$ and $\boldsymbol{w}_t$ are independent and mutually independent[1].

A shorthand notation for the structure described above is given by the quadrupe $\{F_t, G_t, \mathrm{v}_t, \mathbf{W}_t\}$.

## Examples of Basic NDLMs

1. *Regressions*. Take $\boldsymbol{G}_t = \boldsymbol{I}_p$, the $p \times p$ identity matrix, to give a linear regression $y_t$ on regressors in $\boldsymbol{F}_t$, for all $t$, but in which the regression parameters are now time-varying according to a random walk, i.e., $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \mathbf{w}_t$. Traditional static regression is the special case in which $\mathbf{w}_t = 0$ for all $t$, arising (with probability one) by specifying $\mathbf{W}_t = \mathbf{0}$ for all $t$.
2. *Autoregressions*. A particular class of static regression is the class of autoregressions (AR) described by

   $$y_t = \phi_1 y_{t-1} + \cdots \phi_p y_{t-p} + \nu_t,$$

   in which $\boldsymbol{F}_t' = (y_{t-1}, \dots, y_{t-p})$, $\boldsymbol{\theta}_t' = (\phi_{t,1}, \dots, \phi_{t,p})$, and $\boldsymbol{G}_t = \boldsymbol{I}_p$, with $\mathbf{W}_t = \mathbf{0}$ for all $t$.
3. *Autoregressive moving average (ARMA) models*. Consider a zero-mean ARMA model described by

---

[1] Generalizations of the basic DLM assumptions to allow for known, nonzero means for innovations and for independencies between $\nu_t$ and $\boldsymbol{w}_t$.

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t,$$

with $N(\epsilon_t|0, w_t)$. Set $m = \max(p, q+1)$, extend the ARMA coefficients to $\phi_j = 0$ for $j > p$ and $\theta_j = 0$ for $j > q$, and write $\boldsymbol{u} = (1, \theta_1, \dots, \theta_{m-1})'$. Then, the DLM form holds with $\boldsymbol{F}' = (1,0, \dots, 0)$, $v_t = 0$, $\mathbf{W}_t = w_t \boldsymbol{u}\boldsymbol{u}'$, and

$$\boldsymbol{G}_p = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ \phi_{p-1} & 0 & 0 & \cdots & 1 \\ \phi_p & 0 & 0 & \cdots & 0 \end{bmatrix}_{p\times p}$$

and

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{G}_p & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \phi_{p+1} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ \phi_m & 0 & \cdots & 0 \end{bmatrix}_{m\times m}.$$

The AR model is a special case when $q = 0$.

## Recursive Estimation Procedure

The term "Kalman filter" or "Kaman filtering" refers to a recursive procedure for inference about the state of nature $\boldsymbol{\theta}_t$. The key notion here is that given the data $\boldsymbol{y}_t = (y_t, \dots, y_1)'$, inference about $\boldsymbol{\theta}_t$ can be carried out through a direct application of Bayes's theorem:

$$Prob\{\text{State of Nature}|\text{Data}\} \propto Prob\{\text{Data}|\text{State of Nature}\} \times Prob\{\text{State of Nature}\}$$

which can also be written as

$$\underbrace{P(\boldsymbol{\theta}_t|\boldsymbol{y}_t)}_{\text{posterior}} \propto \underbrace{P(y_t|\boldsymbol{\theta}_t, \boldsymbol{y}_{t-1})}_{\text{likelihood}} \times \underbrace{P(\boldsymbol{\theta}_t|\boldsymbol{y}_{t-1})}_{\text{prior}}. \quad (3)$$

At time $t-1$, our state of knowledge about $\boldsymbol{\theta}_{t-1}$ is embodied in the posterior distribution of $\boldsymbol{\theta}_{t-1}$

$$(\boldsymbol{\theta}_{t-1}|\boldsymbol{y}_{t-1}) \sim N(\widehat{\boldsymbol{\theta}}_{t-1}, \Sigma_{t-1}), \quad (4)$$

where $\widehat{\boldsymbol{\theta}}_{t-1}$ and $\Sigma_{t-1}$ denote the expectation and the variance of the posterior distribution. We now at time $t$ but in two stages:

1. Prior to observing $y_t$, and
2. after observing $y_t$

<u>Stage 1</u>: Prior to observing $y_t$, our best choice for $\boldsymbol{\theta}_t$ is governed by eqn. (2) and is given as $\boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t$. Since $\boldsymbol{\theta}_{t-1}$ is described by eqn. (4), our knowledge about $\boldsymbol{\theta}_t$ is embodied in

$$(\boldsymbol{\theta}_t|\boldsymbol{y}_{t-1}) \sim N(\boldsymbol{G}_t\widehat{\boldsymbol{\theta}}_{t-1}, R_t = \boldsymbol{G}_t\Sigma_{t-1}\boldsymbol{G}_t' + \mathbf{W}_t). \quad (4)$$

This is our prior distribution.

<u>Stage 2</u>: On observing $y_t$, our goal is to compute the posterior of $\boldsymbol{\theta}_t$ using eqn. (3). However, to do so, we need to know the likelihood $P(y_t|\boldsymbol{\theta}_t, \boldsymbol{y}_{t-1})$, and the determination of which is under taken via the following arguments.

Let $e_t$ denote the one-step ahead forecast error in predicting $y_t$; thus

$$e_t = y_t - \hat{y}_t = y_t - \boldsymbol{F}_t' \boldsymbol{G}_t \widehat{\boldsymbol{\theta}}_{t-1}. \quad (5)$$

Since $\boldsymbol{F}_t, \boldsymbol{G}_t$ and $\widehat{\boldsymbol{\theta}}_{t-1}$ are all known, observing $y_t$ is equivalent to observing $e_t$. Thus, eqn. (3) can be rewritten as

$$\underbrace{P(\boldsymbol{\theta}_t|\boldsymbol{y}_t)}_{\text{posterior}} = P(\boldsymbol{\theta}_t|e_t, \boldsymbol{y}_{t-1}) \propto \underbrace{P(e_t|\boldsymbol{\theta}_t, \boldsymbol{y}_{t-1})}_{\text{likelihood}} \times \underbrace{P(\boldsymbol{\theta}_t|\boldsymbol{y}_{t-1})}_{\text{prior}}. \quad (6)$$

Using eqn. (1), eqn. (5) can be written as $e_t = \boldsymbol{F}_t'\left(\boldsymbol{\theta}_t - \boldsymbol{G}_t \widehat{\boldsymbol{\theta}}_{t-1}\right) + \nu_t$, so that $E(e_t|\boldsymbol{\theta}_t, \boldsymbol{y}_{t-1}) = \boldsymbol{F}_t'\left(\boldsymbol{\theta}_t - \boldsymbol{G}_t \widehat{\boldsymbol{\theta}}_{t-1}\right)$. Since $\nu_t \sim N(0, \mathrm{v}_t)$, it follows that the likelihood is described by

$$(e_t|\boldsymbol{\theta}_t, \boldsymbol{y}_{t-1}) \sim N\left(\boldsymbol{F}_t'\left(\boldsymbol{\theta}_t - \boldsymbol{G}_t \widehat{\boldsymbol{\theta}}_{t-1}\right), \mathrm{v}_\mathrm{t}\right). \quad (7)$$

We can now use Bayes's theorem (eqn. (6)) to obtain the posterior distribution as follows

$$P(\boldsymbol{\theta}_t|y_t, \boldsymbol{y}_{t-1}) = \frac{P(e_t|\boldsymbol{\theta}_t, \boldsymbol{y}_{t-1}) \times P(\boldsymbol{\theta}_t|\boldsymbol{y}_{t-1})}{\int_{\theta_t} P(e_t, \boldsymbol{\theta}_t|\boldsymbol{y}_{t-1}) d\boldsymbol{\theta}_t}. \quad (8)$$

Once eqn. (8) is computed, we can go back to eqn. eqn. (4) for the next cycle of the recursive procedure.