State-space Model

Eduardo Rossi
University of Pavia

November 2014

Outline

Example

Basic linear regression model

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + \epsilon_t \quad t = 1, \ldots, T$$

$\mathbf{x}_t$ is observed.
Latent factor model:

$$y_t = \mathbf{s}_t' \boldsymbol{\lambda} + u_t$$

where $\mathbf{s}_t$ is a $(K \times 1)$ vector of unobserved or latent factors, and $\boldsymbol{\lambda}$ is the vector of factor loadings.

Latent factor models

- The latent factor class of model is encountered frequently in economics and finance and prominent examples include
  1. multi-factor models of the term structure of interest rates,
  2. dating of business cycles,
  3. the identification of ex ante real interest rates,
  4. real business cycle models with technology shocks,
  5. the APT model
  6. models of stochastic volatility.
- A number of advantages stem from being able to identify the existence of a latent factor structure underlying the behaviour of economic and financial time series.
  1. it provides a parsimonious way of modelling dynamic multivariate systems. (the curse of dimensionality arises even in relatively small systems of unrestricted VARs).
  2. it avoids the need to use ad hoc proxy variables for the unobservable variables which can result in biased and inconsistent parameters estimates.

Latent factor models

- Despite the fact that it is latent, the factor $\mathbf{s}_t$ can nonetheless be characterized via the time series properties of the observed dependent variables.
- The system of equations capturing this structure is commonly referred to as state-space (S-S) form and the technique that enables this identification and extraction of latent factors is known as the Kalman filter.
- An important assumption underlying the Kalman filter is that of normality of the disturbance terms impacting on the system. The assumption of normality makes it possible to summarize the entire distribution of the latent factors using conditional means and variances alone.
- Consequently, these moments feature prominently in the derivations of the recursions that define the Kalman filter. For the case in which the assumption of normality of the disturbance terms is inappropriate, a quasi-maximum likelihood estimator may be available.

Term Structure of Interest Rates

In the case of a single latent factor, $K = 1$, the term structure model for interest rates is

$$r_{i,t} = \lambda_i s_t + u_{i,t} \quad i = 1, \ldots, N$$

N maturities (e.g. three months, one year, three years, five years, seven years, and ten years). s_t is a latent factor and $u_{i,t} \sim i.i.d.N(0, \sigma_i^2)$. The disturbance term, $u_{i,t}$, allows for idiosyncratic movements in the i-th yield which are not explained by movements in the factor.
This set of equations can be written as a single equation in matrix form as

Introduction

Basic model for representing a time series is the additive model:

$$y_t = \mu_t + \gamma_t + \epsilon_t \quad t = 1, \ldots, T$$

- μ_t is a slowly varying component called the *trend*;
- γ_t is a periodic component of fixed period called *seasonal*;
- ϵ_t is the *error*

Local level model

Consider $\mu_t = \alpha_t$

$$\begin{aligned} y_t &= \alpha_t + \epsilon_t \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \\ \alpha_t &= \alpha_{t-1} + \eta_t \quad \eta_t \sim N(0, \sigma_\eta^2) \end{aligned}$$

where ϵ_t and η_t are all mutually independent and are independent of α_1.

The purpose

The object of the methodology is to infer *relevant properties* of the α_t's from a knowledge of the observations $y_1, \ldots, y_T$.

Local level model

We assume initially that

$$\alpha_1 \sim \mathcal{N}(a_1, P_1)$$

where a_1, P_1, σ_ϵ^2 and σ_η^2 are known.
The model for y_t is non-stationary.

- The local level model is an example of a *linear Gaussian state space model*
- The variable α_t is called the *state* and is unobserved.

The Kalman filter

The *object of filtering* is to update the knowledge of the system each time a new observation y_t is brought in. Let

$$Y_{t-1} = \{y_1, \ldots, y_{t-1}\}$$

and assume that

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

If a_t and P_t are known we can calculate a_{t+1} and P_{t+1} when y_t is brought in.

Since

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

The Kalman filter

Define

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

and

$$F_t = \mathrm{Var}(v_t).$$

Then

$$E[v_t | Y_{t-1}] = E[\alpha_t + \epsilon_t - a_t | Y_{t-1}] = a_t - a_t = 0$$

thus

$$E[v_t] = E[E[v_t | Y_{t-1}]] = 0$$

and

$$\mathrm{Cov}[v_t, y_j] = E[v_t y_j] = E[E[v_t | Y_{t-1}] y_j] = 0$$

so v_t and y_j are independent for $j = 1, 2, \ldots, t-1$.

When Y_t is fixed $\Rightarrow Y_{t-1}, y_t$ are fixed $\Rightarrow v_t$ and Y_{t-1} are fixed, and viceversa.

The Kalman filter

Consequently,

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

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

The Kalman filter

Suppose that $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ are random vectors of arbitrary orders that are jointly normally distributed

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

where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yz} \\ \boldsymbol{\Sigma}_{zx} & \boldsymbol{\Sigma}_{zy} & \boldsymbol{\Sigma}_{zz} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} & \mathbf{0} \\ \boldsymbol{\Sigma}_{zx} & \mathbf{0} & \boldsymbol{\Sigma}_{zz} \end{bmatrix}$$

with $\boldsymbol{\Sigma}_{yz} = \boldsymbol{\Sigma}'_{zy} = \mathbf{0}$

The Kalman filter

Multivariate normal regression

$$
\begin{aligned}
E[\mathbf{x}|\mathbf{y},\mathbf{z}] &= E[\mathbf{x}|\mathbf{y}] + \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{z} \\
\mathrm{Var}[\mathbf{x}|\mathbf{y},\mathbf{z}] &= \mathrm{Var}[\mathbf{x}|\mathbf{y}] - \mathbf{\Sigma}_{xy}\mathbf{\Sigma}_{yy}^{-1}\mathbf{\Sigma}_{xy}'
\end{aligned}
$$

The Kalman filter

Proof.
By standard multivariate regression theory

$$\begin{aligned} E[\mathbf{x}|\mathbf{w}] &= \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xw}\boldsymbol{\Sigma}_{ww}^{-1}(\mathbf{w}-\boldsymbol{\mu}_w) \\ \mathrm{Var}[\mathbf{x}|\mathbf{w}] &= \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xw}\boldsymbol{\Sigma}_{ww}^{-1}\boldsymbol{\Sigma}_{xw}' \end{aligned}$$

But with $\mathbf{w} = (\mathbf{y}', \mathbf{z}')'$:

$$\boldsymbol{\Sigma}_{ww} = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{zz} \end{bmatrix}$$

because $\boldsymbol{\Sigma}_{yz} = \mathbf{0}$, and $\boldsymbol{\Sigma}_{xw} = [\boldsymbol{\Sigma}_{xy} \quad \boldsymbol{\Sigma}_{xz}]$

$$\begin{aligned} E[\mathbf{x}|\mathbf{y},\mathbf{z}] &= \boldsymbol{\mu}_x + [\boldsymbol{\Sigma}_{xy} \quad \boldsymbol{\Sigma}_{xz}] \begin{bmatrix} \boldsymbol{\Sigma}_{yy}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{zz}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{y}-\boldsymbol{\mu}_y \\ \mathbf{z} \end{bmatrix} \\ &= \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y}-\boldsymbol{\mu}_y) + \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{z} \end{aligned}$$

since $\boldsymbol{\mu}_z = 0$.

The Kalman filter

Since

$$\text{Var}[\mathbf{x}|\mathbf{w}] = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xw}\boldsymbol{\Sigma}_{ww}^{-1}\boldsymbol{\Sigma}_{xw}'$$

$$\begin{aligned}
\text{Var}[\mathbf{x}|\mathbf{y},\mathbf{z}] &= \boldsymbol{\Sigma}_{xx} - \begin{bmatrix} \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xz} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{yy}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{zz}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{xy}' \\ \boldsymbol{\Sigma}_{xz}' \end{bmatrix} \\
&= \left(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{xy}'\right) - \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{xz}' \\
&= \text{Var}[\mathbf{x}|\mathbf{y}] - \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{xz}'
\end{aligned}$$

The Kalman filter

It follows that

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

Now,

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

and

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

The Kalman filter

Thus, since $a_t = E[\alpha_t | Y_{t-1}]$

$$E[\alpha_t | Y_t] = E[\alpha_t | Y_{t-1}, v_t] = a_t + \frac{P_t}{F_t} v_t$$

or

$$E[\alpha_t | Y_t] = E[\alpha_t | Y_{t-1}, v_t] = a_t + K_t v_t$$

where

$$K_t \equiv \frac{P_t}{F_t}$$

is the regression coefficient of α_t on v_t.
We have

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

The Kalman filter

The full set of relations for updating from time t to time $t+1$:

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

a_1 and P_1 are assumed to be known.

Forecast errors

The Kalman Filter (KF) residual

$$v_t = Y_t - a_t$$

and its variance F_t are the *one-step-ahead forecast error* and the *one-step-ahead forecast variance* of $y_t|Y_{t-1}$.

The joint density of $\{y_1, \dots, y_T\}$ is

$$p(y_1, \dots, y_T) = p(y_1) \prod_{t=2}^{T} p(y_t|Y_{t-1})$$

the density of $\{v_1, \dots, v_T\}$, provided that $v_t = y_t - a_t$ is given by

$$\begin{aligned} p(v_1, \dots, v_T) &= p(y_1, \dots, y_T) \left| \frac{\partial y_t}{\partial v_t} \right| \\ &= \prod_{t=1}^{T} p(v_t) \end{aligned}$$

Consequently, $\{v_1, \dots, v_T\}$ are mutually independent.

Error recursions

Define the *state estimation error* as

$$x_t = \alpha_t - a_t$$

with

$$\mathrm{Var}(x_t) = P_t$$

the state estimation errors and the forecast errors are linear functions of the initial state error x_1 and the disturbances ϵ_t and η_t.

$$v_t = x_t + \epsilon_t$$

$$\begin{aligned} x_{t+1} &= \alpha_{t+1} - a_{t+1} \\ &= \alpha_t + \eta_t - a_t - K_t v_t \\ &= x_t + \eta_t - K_t(x_t + \epsilon_t) \\ &= L_t x_t + \eta_t - K_t \epsilon_t \end{aligned}$$

where

$$L_t = 1 - K_t = \sigma_\epsilon^2 / F_t$$

Smoothed state

Smoothing

Estimation of $\alpha_1, \ldots, \alpha_n$ given the entire sample $Y_n = (y_1, \ldots, y_n)$.

The conditional density

$$\alpha_t | Y_n \sim \mathcal{N}(\widehat{\alpha}_t, V_t)$$

where the *smoothed state*

$$\widehat{\alpha}_t = E(\alpha_t | Y_n)$$

and the *smoothed state variance*

$$V_t = \mathrm{Var}[\alpha_t | Y_n].$$

The operation of calculating $\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_n$ is called *state smoothing*.

Smoothed state

- The forecast errors $v_1, \ldots, v_n$ are mutually independent and are a linear transformation of $y_1, \ldots, y_n$
- the errors $(v_t, \ldots, v_n)$ are independent of $(y_1, \ldots, y_{t-1})$ with zero means.
- when $y_1, \ldots, y_n$ are fixed $\Rightarrow Y_{t-1}$, $v_t, \ldots, v_n$ are fixed and vice versa.

$$E[\mathbf{x}|\mathbf{y}, \mathbf{z}] = E[\mathbf{x}|\mathbf{y}] + \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{z}$$

$$\begin{aligned}
\widehat{\alpha}_t &= E[\alpha_t|Y_n] = E[\alpha_t|Y_{t-1}, v_t, \ldots, v_n] \\
&= E[\alpha_t|Y_{t-1}] + \mathrm{Cov}[\alpha_t, (v_t, \ldots, v_n)']\,\mathrm{Var}[(v_t, \ldots, v_n)']^{-1}(v_t, \ldots, v_n)' \\
&= a_t + \begin{bmatrix} \mathrm{Cov}(\alpha_t, v_t) \\ \vdots \\ \mathrm{Cov}(\alpha_t, v_n) \end{bmatrix}' \begin{bmatrix} F_t & & 0 \\ & \ddots & \\ 0 & & F_n \end{bmatrix}^{-1} \begin{bmatrix} v_t \\ \vdots \\ v_n \end{bmatrix} \\
&= a_t + \sum_{j=t}^{n} \mathrm{Cov}(\alpha_t, v_j) F_j^{-1} v_j
\end{aligned}$$

Smoothed state

$\text{Cov}(\alpha_t, v_j) = \text{Cov}(x_t, v_j)$ for $j = t, \ldots, n$ and

$$\begin{aligned} \text{Cov}(x_t, v_t) &= E[x_t(x_t + \epsilon_t)] = \text{Var}(x_t) = P_t \\ \text{Cov}(x_t, v_{t+1}) &= E[x_t(x_{t+1} + \epsilon_{t+1})] = E[x_t(L_t x_t + \eta_t - K_t \epsilon_t)] = P_t L_t \end{aligned}$$

Similarly,

$$\begin{aligned} \text{Cov}(x_t, v_{t+2}) &= P_t L_t L_{t+1} \\ &\vdots \\ \text{Cov}(x_t, v_n) &= P_t L_t L_{t+1} \ldots L_{n-1} \end{aligned}$$

Smoothed state

Substituting in $\widehat{\alpha}_t$:

$$\begin{aligned}\widehat{\alpha}_t &= a_t + P_t\frac{v_t}{F_t} + P_tL_t\frac{v_{t+1}}{F_{t+1}} + P_tL_tL_{t+1}\frac{v_{t+2}}{F_{t+2}} + \ldots \\ &= a_t + P_t r_{t-1}\end{aligned}$$

where

$$r_{t-1} = \frac{v_t}{F_t} + L_t\frac{v_{t+1}}{F_{t+1}} + L_tL_{t+1}\frac{v_{t+2}}{F_{t+2}} + \ldots + L_tL_{t+1}\ldots L_{n-1}\frac{v_n}{F_n}$$

is a weighted sum of innovations after $t-1$.
$r_n = 0$ since no observations are available after time n.

Smoothed state

The value of r_{t-1} can be evaluated using the backward recursion

$$r_{t-1} = \frac{v_t}{F_t} + L_t r_t \quad t = n, n-1, \ldots, 1$$

The smoothed stare can be calculated by the backwards recursion (*smoothing state recursion*)

$$\begin{aligned} r_{t-1} &= \frac{v_t}{F_t} + L_t r_t \\ \widehat{\alpha}_t &= a_t + P_t r_{t-1} \quad t = n, n-1, \ldots, 1 \end{aligned}$$

with $r_n = 0$.

Smoothed state variance

The error variance of the smoothed state

$$
\begin{aligned}
V_t &= \mathrm{Var}[\alpha_t|Y_n] = \mathrm{Var}[\alpha_t|Y_{t-1}, v_t, \ldots, v_n] \\
&= \mathrm{Var}[\alpha_t|Y_{t-1}] \\
&\quad - \mathrm{Cov}[\alpha_t, (v_t, \ldots, v_n)'] \,\mathrm{Var}[(v_t, \ldots, v_n)']^{-1} \,\mathrm{Cov}[\alpha_t, (v_t, \ldots, v_n)']' \\
&= P_t - \begin{bmatrix} \mathrm{Cov}(\alpha_t, v_t) \\ \vdots \\ \mathrm{Cov}(\alpha_t, v_n) \end{bmatrix}' \begin{bmatrix} F_t & & 0 \\ & \ddots & \\ 0 & & F_n \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{Cov}(\alpha_t, v_t) \\ \vdots \\ \mathrm{Cov}(\alpha_t, v_n) \end{bmatrix} \\
&= P_t - \sum_{j=1}^{n} \mathrm{Cov}[(\alpha_t, v_j)]^2 F_j^{-1}
\end{aligned}
$$

Given that $\mathrm{Cov}(\alpha_t, v_j) = \mathrm{Cov}(x_t, v_j)$ for $j = t, \ldots, n$ and

$$
\begin{aligned}
\mathrm{Cov}(x_t, v_t) &= P_t \\
\mathrm{Cov}(x_t, v_{t+1}) &= P_t L_t
\end{aligned}
$$

Smoothed state variance

We can obtain

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

where

$$N_{t-1} = \frac{1}{F_t} + L_t^2 \frac{1}{F_{t+1}} + L_t^2 L_{t+1}^2 \frac{1}{F_{t+2}} + \ldots + L_t^2 L_{t+1}^2 \ldots L_{n-1}^2 \frac{1}{F_n}$$

$$N_t = \frac{1}{F_{t+1}} + L_{t+1}^2 \frac{1}{F_{t+2}} + L_{t+1}^2 L_{t+2}^2 \frac{1}{F_{t+3}} + \ldots + L_{t+1}^2 L_{t+2}^2 \ldots L_{n-1}^2 \frac{1}{F_n}$$

$N_n = 0$. N_{t-1} can be calculated using the backward recursion

$$N_{t-1} = \frac{1}{F_t} + L_t^2 N_t$$

$$N_t = \mathrm{Var}(r_t)$$

since the forecast errors v_t are independent.

Smoothed state variance

The error variance of the smoothed state can be calculated by the backwards recursion

$$N_{t-1} = \frac{1}{F_t} + L_t^2 N_t \quad V_t = P_t - P_t^2 N_{t-1} \quad t = n, \ldots, 1$$

From the standard error $\sqrt{V_t}$ of $\widehat{\alpha}_t$ we can construct confidence intervals for α_t for $t = 1, \ldots, n$. It is also possible to derive the smoothed covariances between the states $\text{Cov}(\alpha_s, \alpha_t | Y_n), t \neq s$.

Disturbance smoothing

Smoothed observation disturbance

$$\widehat{\epsilon}_t = E[\epsilon_t | Y_n] = y_t - \widehat{\alpha}_t$$

smoothed state disturbance

$$\widehat{\eta}_t = E[\eta_t | Y_n] = \widehat{\alpha}_{t+1} - \widehat{\alpha}_t$$

the estimates of $\widehat{\epsilon}_t$ and $\widehat{\eta}_t$ are useful for detecting outliers and structural breaks.

Missing observations

Missing observations are very easy to handle in Kalman filtering.

- Suppose the observations $y_j, j = \tau, \ldots, \tau^* - 1$ are missing for $1 < \tau < \tau^* \leq n$.
- The most ideal way to deal with it is to define a new series

$$\begin{aligned} y_t^* &= y_t & t &= 1, \ldots, \tau - 1 & & \\ y_t^* &= y_{t+\tau^*-\tau} & t &= \tau, \ldots, n^* & n^* &= n - (\tau^* - \tau) \end{aligned}$$

The model is the same local level model with $y_t = y_t^*$ except that

$$\alpha_\tau = \alpha_{\tau-1} + \eta_{\tau-1}$$

where

$$\eta_{\tau-1} \sim \mathcal{N}(0, (\tau^* - \tau)\sigma_\eta^2)$$

Missing observations

Filtering at time $t = \tau, \ldots, \tau^* - 1$

$$E[\alpha_t|Y_{t-1}] = E[\alpha_t|Y_{\tau-1}] = E\left[\alpha_\tau + \sum_{j=\tau}^{t-1} \eta_j|Y_{\tau-1}\right] = a_\tau$$

and

$$\mathrm{Var}[\alpha_t|Y_{t-1}] = \mathrm{Var}[\alpha_t|Y_{\tau-1}] = \mathrm{Var}\left[\alpha_\tau + \sum_{j=\tau}^{t-1} \eta_j|Y_{\tau-1}\right] = P_\tau + (t-\tau)\sigma_\eta^2$$

giving

$$a_{t+1} = a_t \quad P_{t+1} = P_t + \sigma_\eta^2 \quad t = \tau, \ldots, \tau^* - 1$$

We can use the original KF for all t, by taking $v_t = 0$ and $K_t = 0$ at the missing time points.

Forecasting

Let $\overline{y}_{n+j}$ be the minimum MSE forecast given the time series $\{y_1, \ldots, y_n\}, j = 1, 2, \ldots, J$, with $J > 0$. Then

$$\overline{y}_{n+j} = E[y_{n+j}|Y_n]$$

The theory of forecasting for the local level model: we regard forecasting as filtering the observations $(y_1, \ldots, y_n, y_{n+1}, \ldots, y_{n+J})$ using KF and treating the last J observations $y_{n+1}, \ldots, y_{n+J}$ as missing.

Letting

$$\begin{aligned} \overline{a}_{n+j} &= E[\alpha_{n+j}|Y_n] \\ \overline{P}_{n+j+1} &= \overline{P}_{n+j} + \sigma_\eta^2 \quad j = 1, \ldots, J-1 \end{aligned}$$

with $\overline{a}_{n+1} = a_{n+1}$ and $\overline{P}_{n+1} = P_{n+1}$ obtained from the KF.

Forecasting

The forecast of y:

$$\begin{aligned} \overline{y}_{n+j} &= E[y_{n+j}|Y_n] = E[\alpha_{n+j}|Y_n] + E[\epsilon_{n+j}|Y_n] = \overline{a}_{n+j} \\ \overline{F}_{n+j} &= \mathrm{Var}[y_{n+j}|Y_n] + \mathrm{Var}[\epsilon_{n+j}|Y_n] = \overline{P}_{n+j} + \sigma_\epsilon^2 \end{aligned}$$

for $j = 1, \ldots, J$.

The KF can be applied for $t = 1, \ldots, n+J$ where the observations at times $n+1, \ldots, n+J$ are treated as missing.

Initialization

How to start up the filter when nothing is known about the distribution of α_1.

Diffuse prior:

$$\alpha_1 \sim N(a_1, P_1)$$

where a_1 is set at an arbitrary value and $P_1 \to \infty$.

$$v_1 = y_1 - a_1 \quad F_1 = P_1 + \sigma_\epsilon^2$$

substituting into the equations for a_2 and P_2:

$$a_2 = a_1 + \frac{P_1}{P_1 + \sigma_\epsilon^2}(y_1 - a_1)$$

$$\begin{aligned} P_2 &= P_1 + \left(1 - \frac{P_1}{P_1 + \sigma_\epsilon^2}\right) + \sigma_\eta^2 \\ &= \frac{P_1}{P_1 + \sigma_\epsilon^2}\sigma_\epsilon^2 + \sigma_\eta^2 \end{aligned}$$

letting $P_1 \to \infty$, we obtain $a_2 = y_1, P_2 = \sigma_\epsilon^2 + \sigma_\eta^2$, then we proceed with the KF (diffuse KF).

Parameter estimation

Since

$$p(y_1, \ldots, y_n) = p(Y_{t-1}) p(y_t | Y_{t-1}) \quad t = 2, \ldots, n$$

the joint density of $y_1, \ldots, y_n$ can be expressed as

$$p(y) = \prod_{t=1}^{n} p(y_t | Y_{t-1})$$

where $p(y_1 | Y_0) = p(y_1)$.

$$p(y_t | Y_{t-1}) = \mathcal{N}(a_t, F_t) \quad v_t = y_t - a_t$$

the log-likelihood is given by

$$\log L = \log p(y) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{n} \left(\log F_t + \frac{v_t^2}{F_t} \right)$$

with v_t and F_t from the KF. This the *prediction error decomposition* of the likelihood. Estimation proceeds by numerically maximizing $\log L$.

Parameter estimation

The log-likelihood in the *diffuse case*.

- All terms in the log-lik expression remain finite as $P_1 \to \infty$ with y fixed except the term for $t = 1$.
- To remove the influence of P_1 define the diffuse log-likelihood as:

$$\begin{aligned}\log L_d &= \lim_{P_1 \to \infty} \left(\log L + \frac{1}{2} \log P_1 \right) \\ &= -\frac{1}{2} \lim_{P_1 \to \infty} \left(\log \frac{F_1}{P_1} + \frac{v_1^2}{F_1} \right) \\ &\quad - \frac{n}{2} \log (2\pi) - \frac{1}{2} \sum_{t=2}^{n} \left(\log F_t + \frac{v_t^2}{F_t} \right)\end{aligned}$$

since $F_1/P_1 \to 1$ and $v_1^2/F_1 \to 0$ as $P_1 \to \infty$.
P_1 does not depend on σ_ϵ^2 and σ_η^2, the values of σ_ϵ^2 and σ_η^2 that maximize $\log L$ are identical to the values that maximize $\log L + \frac{1}{2} \log P_1$.

Linear Gaussian state space model

Linear Gaussian state space model is defined in *three parts*:

- *State equation*:

$$\alpha_{t+1} = T_t \alpha_t + R_t \zeta_t \quad \eta_t \sim i.i.d.\mathcal{N}(0, Q_t)$$

- *Observation equation*:

$$y_t = Z_t \alpha_t + \epsilon_t \quad \epsilon_t \sim i.i.d.\mathcal{N}(0, H_t)$$

- *Initial state distribution*:

$$\alpha_1 \sim \mathcal{N}(a_1, P_1)$$

The matrices Z_t, T_t, R_t, H_t, Q_t are independent of $\{\epsilon_1, \ldots, \epsilon_n\}$ and $\{\eta_1, \ldots, \eta_n\}$. In many applications $R_t = I_m$, the theory remains valid if R_t $(m \times r)$.

Linear Gaussian state space model

- The idea underlying the model is that the development of the system over time is determined by α_t according to the state equation.
- Because α_t cannot be observed directly we must base the analysis on observations y_t.
- The matrices Z_t, T_t, R_t, H_t and Q_t are initially assumed to be known and the error terms ϵ_t and η_t are assumed to be serially independent and independent of each other at all time points.
- Matrices Z_t, and T_{t-1}, are permitted to depend on $y_1, \ldots, y_{t-1}$
- The initial state vector is assumed to be $\mathcal{N}(a_1, P_1)$ independently of $\epsilon_1, \ldots, \epsilon_n$ and $\eta_1, \ldots, \eta_n$.
- a_1 and P_1 are assumed to be known.

Linear Gaussian state space model

- State space model is linear and Gaussian: therefore properties and results of multivariate normal distribution apply;
- State vector α_t evolves as a VAR(1) process;
- System matrices usually contain unknown parameters;
- Estimation has therefore two aspects:
  1. measuring the unobservable state (prediction, filtering and smoothing);
  2. estimation of unknown parameters (maximum likelihood estimation);
- State space methods offer a unified approach to a wide range of models and techniques: dynamic regression, ARIMA, UC models, latent variable models, spline-fitting and many ad-hoc filters;

Regression with Time Varying Coefficients

Regressors in $Z_t = X_t$

$$T_t = I$$
$$R_t = I$$

regression model with coefficient α_t following a random walk.

Unobserved Component models

Local level model:

$$\begin{aligned} y_t &= \mu_t + \epsilon_t & \epsilon_t &\sim N(0, \sigma_\epsilon^2) \\ \mu_{t+1} &= \mu_t + \eta_t & \eta_t &\sim N(0, \sigma_\eta^2) \end{aligned}$$

State equation:

$$\alpha_{t+1} = T_t \alpha_t + R_t \zeta_t \quad \zeta_t \sim i.i.d.\mathcal{N}(0, Q_t)$$

$$\alpha_t = \mu_t \quad T_t = 1 \quad R_t = 1 \quad Q_t = \sigma_\eta^2$$

Observation equation:

$$y_t = Z_t \alpha_t + \epsilon_t \quad \epsilon_t \sim i.i.d.\mathcal{N}(0, H_t)$$

$$Z_t = 1 \quad H_t = \sigma_\epsilon^2$$

Local linear trend model

$$
\begin{aligned}
y_t &= \mu_t + \epsilon_t & \epsilon_t &\sim N(0, \sigma_\epsilon^2) \\
\mu_{t+1} &= \mu_t + \nu_t + \eta_t & \eta_t &\sim N(0, \sigma_\eta^2) \\
\nu_{t+1} &= \nu_t + \xi_t & \xi_t &\sim N(0, \sigma_\xi^2)
\end{aligned}
$$

If $\xi_t = \eta_t = 0$

$$\nu_{t+1} = \nu_t = \nu$$

$$\mu_{t+1} = \mu_t + \nu$$

this entails

$$
\begin{aligned}
y_t &= \mu_t + \epsilon_t \\
\mu_{t+1} &= \mu_t + \nu
\end{aligned}
$$

deterministic trend plus noise.

Local linear trend model

State equation:

$$\alpha_{t+1} = T_t \alpha_t + R_t \zeta_t \quad \zeta_t \sim i.i.d.\mathcal{N}(0, Q_t)$$

$$\alpha_t = \begin{bmatrix} \mu_t \\ \nu_t \end{bmatrix} \quad T_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad R_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad Q_t = \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\xi^2 \end{bmatrix}$$

Observation equation:

$$y_t = Z_t \alpha_t + \epsilon_t \quad \epsilon_t \sim i.i.d.\mathcal{N}(0, H_t)$$

$$Z_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad H_t = \sigma_\epsilon^2$$

ARIMA

ARIMA(p,d,q): All ARIMA(p,d,q) models have a (non-unique) state space representation.

$$y_t = \Delta^d x_t$$

$$\phi(L) y_t = \theta(L) \eta_t$$

$$y_t = \sum_{j=1}^{r} \phi_j y_{t-j} + \eta_t + \sum_{j=1}^{r-1} \theta_j \eta_{t-j}$$

$$Z_t = [1, 0, 0, \ldots, 0]$$

$$\alpha_t = \begin{bmatrix} y_t \\ \phi_2 y_{t-1} + \ldots + \phi_r y_{t-r+1} + \theta_1 \eta_t + \ldots + \theta_{r-1} \eta_{t-r+2} \\ \phi_3 y_{t-1} + \ldots + \phi_r y_{t-r+2} + \theta_1 \eta_t + \ldots + \theta_{r-1} \eta_{t-r+3} \\ \vdots \\ \phi_r y_{t-1} + \theta_{r-1} \eta_t \end{bmatrix}$$

$r = \max(p, q+1)$

ARIMA

$$T_t = \begin{bmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ \vdots & & & & \\ \phi_{r-1} & 0 & 0 & \dots & 1 \\ \phi_r & 0 & 0 & \dots & 0 \end{bmatrix}$$

$$R_t = R = \begin{bmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{r-1} \end{bmatrix}$$

$$\zeta_t = \eta_{t+1}$$

Observation equation:

$$y_t = Z_t \alpha_t$$

$H_t = 0 \Rightarrow \epsilon_t = 0.$

ARIMA

MA(1) model:

$$y_t = \eta_t + \theta_1 \eta_{t-1}$$

Observation equation:

$$y_t = Z_t \alpha_t \quad Z_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad H_t = 0$$

State equation:

$$\left[\begin{array}{c} y_{t+1} \\ \theta_1 \eta_{t+1} \end{array} \right] = \left[\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array} \right] \left[\begin{array}{c} y_t \\ \theta_1 \eta_t \end{array} \right] + \left[\begin{array}{c} 1 \\ \theta_1 \end{array} \right] \eta_{t+1}$$

$$Q_t = \sigma^2_\eta$$

ARIMA

ARMA(2,1), $r = 2$,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \eta_t + \theta_1 \eta_{t-1}$$

Observation equation:

$$y_t = Z_t \alpha_t \quad Z_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad H_t = 0$$

State equation:

$$\begin{bmatrix} y_{t+1} \\ \phi_2 y_t + \theta_1 \eta_{t+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ \phi_2 y_{t-1} + \theta_1 \eta_t \end{bmatrix} + \begin{bmatrix} 1 \\ \theta_1 \end{bmatrix} \eta_{t+1}$$

$$\begin{aligned} y_{t+1} &= \phi_1 y_t + \phi_2 y_{t-1} + \theta_1 \eta_t + \eta_{t+1} \\ \phi_2 y_t + \theta_1 \eta_{t+1} &= \phi_2 y_t + \theta_1 \eta_{t+1} \end{aligned}$$

ARIMA

ARIMA(2,1,1):

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \eta_t + \theta_1 \eta_{t-1}$$

Observation equation:

$$y_t = [1, 1, 0]\, \alpha_t$$

State equation:

$$\alpha_{t+1} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & \phi_1 & 0 \\ 0 & \phi_2 & 0 \end{bmatrix} \alpha_t + \begin{bmatrix} 0 \\ 1 \\ \theta_1 \end{bmatrix} \eta_{t+1}$$

State vector:

$$\alpha_t = \begin{bmatrix} y_{t-1} \\ \Delta y_t \\ \phi_2 \Delta y_{t-1} + \theta_1 \eta_t \end{bmatrix}$$

ARIMA

1st equation

$$y_t = y_{t-1} + \Delta y_t$$

2nd equation

$$\Delta y_{t+1} = \phi_1 \Delta y_t + \phi_2 \Delta y_{t-1} + \eta_{t+1} + \theta_1 \eta_t$$

3rd equation

$$\phi_2 \Delta y_t + \theta_1 \eta_{t+1} = \phi_2 \Delta y_t + \theta_1 \eta_{t+1}$$

ARIMA

ARIMA(2,2,1):

$$\Delta^2 y_t = \phi_1 \Delta^2 y_{t-1} + \phi_2 \Delta^2 y_{t-2} + \eta_t + \theta_1 \eta_{t-1}$$

$$\alpha_{t+1} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & \phi_1 & 1 \\ 0 & 0 & \phi_2 & 0 \end{bmatrix} \alpha_t + \begin{bmatrix} 0 \\ 0 \\ 1 \\ \theta_1 \end{bmatrix} \eta_{t+1}$$

$$\alpha_t = \begin{bmatrix} y_{t-1} \\ \Delta y_{t-1} \\ \Delta^2 y_t \\ \phi_2 \Delta^2 y_{t-1} + \theta_1 \eta_t \end{bmatrix}$$

3rd equation

$$\Delta^2 y_{t+1} = \phi_1 \Delta^2 y_t + \phi_2 \Delta^2 y_{t-1} + \eta_{t+1} + \theta_1 \eta_t$$

The need to facilitate the initialization explains why the S-S model is set in this form. The elements y_0 and Δy_0 are treated as diffuse random elements while the other elements, including $\Delta^2 y_t$ are stationary which have proper unconditional means and variances.

Kalman filter recursions

The unobserved state α_t can be estimated from the observations with the Kalman filter:

$$\begin{aligned} v_t &= y_t - Z_t E[y_t|Y_{t-1}] = y_t - Z_t a_t \\ F_t &= Z_t P_t Z_t' + H_t \\ K_t &= T_t P_t Z_t' F_t^{-1} \\ a_{t+1} &= T_t a_t + K_t v_t \\ L_t &= T_t - K_t Z_t \\ P_{t+1} &= T_t P_t L_t' + R_t Q_t R_t' \end{aligned}$$

for $t = 1, \ldots, n$ and starting with given values for a_1 and P_1.

Kalman filter recursions

The contemporaneous filtering equations incorporate the computation of the state vector estimator $a_{t|t} \equiv E[\alpha_t|Y_t]$ and its variance $P_{t|t}$ These equations are just a re-formulation of the Kalman filter and me given by

$$\begin{aligned} M_t &= P_t Z_t' \\ a_{t|t} &= a_t + M_t F_t^{-1} v_t \\ P_{t|t} &= P_t - M_t F_t^{-1} M_t' \\ a_{t+1} &= T_t a_{t|t} \\ P_{t+1} &= T_t P_{t|t} T_t' + R_t Q_t R_t' \end{aligned}$$