Dynamic Statistical Models with Hidden Variables

Benjamin Poignard

CREST & PSL (Paris Dauphine University, CEREMADE)

Chapter 2: The Kalman Filter

State-space models

General form:

$$\begin{cases} y_t &= M_t\alpha_t + d_t + u_t, \text{ Measurement equation} \\ \alpha_t &= T_t\alpha_{t-1} + c_t + R_t v_t, \text{ Transition equation} \end{cases}$$

with $y_t \in \mathbb{R}^N$, $\alpha_t \in \mathbb{R}^m$ (the state-vector), (u_t) and (v_t) are two sequences of independent variables, respectively valued in $\mathbb{R}^N$ and $\mathbb{R}^K$ such that

$$\mathbb{E}[u_t] = 0_N,\ \mathbb{E}[v_t] = 0_K,\ \text{Var}(u_t) = H_t,\ \text{Var}(v_t) = Q_t.$$

M_t, T_t and R_t are non-random $N \times n$, $m \times m$ and $m \times K$ matrices, $d_t \in \mathbb{R}^N$, $c_t \in \mathbb{R}^m$ are non-random vectors.

Objectives of the Kalmnan filter

The Kalman filter (Kalman, 1960) is an algorithm used for

(i) **predicting** the value of the state vector at time t, given observations $y_1, \cdots, y_{t-1}$.

(ii) **filtering**, that is, estimating α_t given observations $y_1, \cdots, y_t$.

(iii) **smoothing**, that is, estimating α_t given observations $y_1, \cdots, y_T$, with $T > t$.

Assumptions

To implement this algorithm, we need further assumptions: normality and independence:

- (u_t, v_t) is an **independent Gaussian** sequence such that

$$\mathbb{P}_{(u_t,v_t)'} = \mathcal{N}_{\mathbb{R}^{N+K}}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}H_t & 0\\0 & Q_t\end{pmatrix}\right)$$

- The initial distribution of the state vector is Gaussian and is independent from (u_t) and (v_t):

$$\mathbb{P}_{\alpha_0} = \mathcal{N}_{\mathbb{R}^K}(a_0, P_0),\ \alpha_0 \perp (u_t), (v_t).$$

- The matrix H_t is positive definite (for any t).

1. General form of the filter
   - Prediction and updating formulas
   - Prediction at any horizon and smoothing

2. Statistical inference

Notations: conditional moments w.r.t. bservations

For $t \geq 1$, then

$$\begin{aligned} \alpha_{t|t} &= \mathbb{E}[\alpha_t | y_1, \cdots, y_t], \\ P_{t|t} &= \mathrm{Var}(\alpha_t | y_1, \cdots, y_t). \end{aligned}$$

For $t > 1$, then

$$\begin{aligned} \alpha_{t|t-1} &= \mathbb{E}[\alpha_t | y_1, \cdots, y_{t-1}], \\ P_{t|t-1} &= \mathrm{Var}(\alpha_t | y_1, \cdots, y_{t-1}). \end{aligned}$$

Let

$$\alpha_{1|0} = \mathbb{E}[\alpha_1], \; P_{1|0} = \mathrm{Var}(\alpha_1).$$

The objective consists of computing these sequences recursively.

First step

Taking the conditional expectation w.r.t. $y_1, \cdots, y_{t-1}$ in the transition equation gives

$$\alpha_{t|t-1} = T_t \alpha_{t-1|t-1} + c_t,$$

and by taking the conditional variance

$$P_{t|t-1} = T_t P_{t-1|t-1} T_t' + R_t Q_t R_t'.$$

Both equations are called **prediction equations**. Hence the conditional moments of y_t are

$$\begin{aligned} y_{t|t-1} &:= \mathbb{E}[y_t | y_1, \cdots, y_{t-1}] = M_t \alpha_{t|t-1} + d_t, \\ F_{t|t-1} &:= \mathrm{Var}(y_t | y_1, \cdots, y_{t-1}) = M_t P_{t|t-1} M_t' + H_t. \end{aligned}$$

We also have

$$\mathrm{Cov}(\alpha_t, y_t | y_1, \cdots, y_{t-1}) = P_{t|t-1} M_t'.$$

Conditional distributions of the components of a Gaussian vector

Let

$$\mathbb{P}_{(X,Y)'} = \mathcal{N}_{\mathbb{R}^{d_X \times d_Y}} \left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} \right).$$

Then the distribution of X conditional on $Y = y$ is

$$\mathbb{P}_X^{Y=y} = \mathcal{N}_{\mathbb{R}^{d_X}}(\mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y - \mu_Y), \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}).$$

Conditional law of (y_t, α_t)

We have

$$(y_t, \alpha_t, y_{t-1}, \cdots, y_1) = F(\alpha_0, u_t, \cdots, u_1, v_t, \cdots, v_1),$$

where $F(.)$ is a linear mapping.

Consequently, the vector $(y_t, \alpha_t, y_{t-1}, \cdots, y_1)$ is gaussian.

The law of (y_t, α_t) conditional on $y_1, \cdots, y_{t-1}$ is also gaussian.

Second step: updating the prediction formulas

New observation at time t: y_t.
α_t is gaussian conditionally on $y_1, \cdots, y_{t-1}$ and y_t:

$$\begin{aligned} \alpha_{t|t} &= \alpha_{t|t-1} + P_{t|t-1} M_t' F_{t|t-1}^{-1} (y_t - M_t \alpha_{t|t-1} - d_t), \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1} M_t' F_{t|t-1}^{-1} M_t P_{t|t-1}. \end{aligned}$$

These equations are called **updating equations**.
Remark: The normality assumption is only used in the second step.
Initialization: At time 1, the conditional moments coincide with the unconditional ones, ie

$$\alpha_{1|0} = T_1 a_0 + c_1, \; P_{1|0} = T_1 P_0 T_1' + R_1 Q_1 R_1'.$$

Recursive computations

The sequences $(\alpha_{t|t_1}), (P_{t|t-1}), (\alpha_{t|t})$ and $(P_{t|t})$ are computed recursively for $t = 1, \cdots, T$.

Initial values:

$$\alpha_{1|0} = T_1 a_0 + c_1, \; P_{1|0} = T_1 P_0 T_1' + R_1 Q_1 R_1'.$$

Prediction equations:

$$\begin{aligned} \alpha_{t|t-1} &= T_t \alpha_{t-1|t-1} + c_t, \\ P_{t|t-1} &= T_t P_{t-1|t-1} T_t' + R_t Q_t R_t', \\ F_{t|t-1} &= M_t P_{t|t-1} M_t' + H_t. \end{aligned}$$

Updating equations: using also y_t, ie

$$\begin{aligned} \alpha_{t|t} &= \alpha_{t|t-1} + P_{t|t-1} M_t' F_{t|t-1}^{-1} (y_t - M_t \alpha_{t|t-1} - d_t), \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1} M_t' F_{t|t-1}^{-1} M_t P_{t|t-1}. \end{aligned}$$

Direct computation of $(\alpha_{t|t-1})$ and $(P_{t|t-1})$

$$\begin{cases} \alpha_{t|t-1} & = & T_t \alpha_{t-1|t-2} + c_t + K_t (y_t - M_{t-1} - \alpha_{t-1|t-2} - d_{t-1}), \\ P_{t|t-1} & = & T_t P_{t-1|t-2} T_t' - K_t F_{t-1|t-2} K_t' + R_t Q_t R_t', \end{cases}$$

where

$$F_{t-1|t-2} = M_{t-1} P_{t-1|t-2} M_{t-1}' + H_{t-1},$$
$$K_t = T_t P_{t-1|t-2} M_{t-1}' F_{t-1|t-2}^{-1}.$$

K_t is the gain matrix.

Correlated noise sequences

We can relax the assumption regarding the noncorrelation between the noises:

$$\mathbb{P}_{(u_t,v_t)'} = \mathcal{N}_{R^{N+K}}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}H_t & G_t'\\ G_t & Q_t\end{pmatrix}\right).$$

Prediction equations:

$$\alpha_{t|t-1} = T_t\alpha_{t-1|t-1} + c_t,\ P_{t|t-1} = T_tP_{t-1|t-1}T_t' + R_tQ_tR_t',$$
$$F_{t|t-1} = M_tP_{t|t-1}M_t' + H_t + M_tR_tG_t + G_t'R_t'M_t'.$$

Updating equations:

$$\begin{aligned}\alpha_{t|t} &= \alpha_{t|t-1} + (P_{t|t-1}M_t' + R_tG_t)F_{t|t-1}^{-1}(y_t - M_t\alpha_{t|t-1} - d_t),\\ P_{t|t} &= P_{t|t-1} - (P_{t|t-1}M_t' + R_tG_t)F_{t|t-1}^{-1}(M_tP_{t|t-1} + G_t'R_t').\end{aligned}$$

Can the normality assumption be relaxed?

For random vectors $X \in L^2(\mathbb{R}^m)$ and $Y \in \mathbb{R}^n$, the conditional expectation $\mathbb{E}[X|Y]$ is characterized by

$$\|X - \mathbb{E}[X|Y]\|_2^2 = \min_{\varphi \in \Phi} \|X - \varphi(Y)\|_2^2,$$

where Φ is the set of measurable functions $\varphi : \mathbb{R}^n \mapsto \mathbb{R}^m$ such that $\varphi(Y) \in L^2(\mathbb{R}^n)$.

The linear conditional expectation $\mathbb{E}L[X|Y]$ is characterized by the same program but with φ linear, ie

$$\|X - \mathbb{E}L[X|Y]\|_2^2 = \min_{A,b} \|X - AY - b\|_2^2.$$

For gaussian vectors, the two conditional expectations coincide.

Can the normality assumption be relaxed?

The linear conditional expectation only depends on the L^2 structure of (X, Y).
It follows that

$$\mathbb{E}L[X|Y] = \mu_X + \Sigma_{XX}\Sigma_{YY}^{-1}(Y - \mu_Y).$$

Without the gaussian assumption, the Kalman filter provides the linear prediction

$$\mathbb{E}L[y_t|y_1, \cdots, y_{t-1}] = M_t\alpha_{t|t-1} + d_t,$$

and the variance of the prediction error

$$\mathrm{Var}(y_t - \mathbb{E}L[y_t|y_1, \cdots, y_{t-1}]) = F_{t|t-1} = M_t P_{t|t-1} M_t' + H_t.$$

Prediction

The Kalman filter can be used to predict at any horizon.
To simplify, let $\forall t, c_t = d_t = 0$, $T_t = T$ and $M_t = M$:

$$\begin{cases} y_t &= M\alpha_t + u_t, \\ \alpha_t &= T\alpha_{t-1} + R_t v_t. \end{cases}$$

For any $h \geq 0$, then

$$\alpha_{t+h} = T^{h+1}\alpha_{t-1} + \sum_{i=0}^{h} T^{h-i} R_{t+i} v_{t+i},$$

then

$$\alpha_{t+h|t-1} = \mathbb{E}[\alpha_{t+h}|y_1, \cdots, y_{t-1}] = T^{h+1}\alpha_{t-1|t-1}.$$

Prediction

The variance of the prediction error at horizon $h+1$ is

$$\begin{aligned} P_{t+h|t-1} &= \text{Var}(\alpha_{t+h} - \alpha_{t+h|t-1}) \\ &= T^{h+1} P_{t-1|t-1} (T^{h+1})' + \sum_{i=0}^{h} T^{h-i} R_{t+i} Q_{t+i} (T^{h-i} R_{t+i})'. \end{aligned}$$

Moreover, $y_{t+h} = M\alpha_{t+h} + u_{t+h}$, and consequently

$$y_{t+h|t-1} = \mathbb{E}[y_{t+h}|y_1, \cdots, y_{t-1}] = M\alpha_{t+h|t-1} = MT^{h+1}\alpha_{t-1|t-1}.$$

The prediction error is
$y_{t+h} - y_{t+h|t-1} = M(\alpha_{t+h} - \alpha_{t+h|t-1}) + u_{t+h}$, and its variance is

$$\text{Var}(y_{t+h} - y_{t+h|t-1}) = MP_{t+h|t-1}M' + H_{t+h}.$$

Smoothing

The updating formula provides the filtered value $\alpha_{t|t}$ of α_t.

For certain applications, it is important to smooth α_t using the posterior observations.

Let

$$\alpha_{t|T} = \mathbb{E}[\alpha_t | y_1, \cdots, y_T], \; P_{t|T} = \mathrm{Var}(\alpha_t | y_1, \cdots, y_T).$$

Steps for computing $\alpha_{t|T}$

(i) $\mathbb{E}[\alpha_t, \alpha_{t+1} | y_1, \cdots, y_t]$ (Already known).
$\Downarrow$ Using the normality.

(ii) $\mathbb{E}[\alpha_t | y_1, \cdots, y_t, \alpha_{t+1}]$
$\Downarrow$ Using a Lemma.

(iii) $\mathbb{E}[\alpha_t | y_1, \cdots, y_t, y_{t+1}, \cdots, y_T, \alpha_{t+1}]$
$\Downarrow$ By deconditioning.

(iv) $\mathbb{E}[y_t | y_1, \cdots, y_T]$.

Algorithm

The algorithm is initialized at $\alpha_{T|T}$ and is used in a descending recurrence:

$$\alpha_{t|T} = \alpha_{t|t} + \bar{F}_t(\alpha_{t+1|T} - \alpha_{t+1|t}),\ t < T,$$

and

$$P_{t|T} = P_{t|t} + \bar{F}_t(P_{t+1|T} - P_{t+1|t})\bar{F}_t',\ t < T,$$

with

$$\bar{F}_t = P_{t|t} T_{t+1}' P_{t+1|t}^{-1}, t < T.$$

Sketch of the proof (1)

Normality assumption

$\mathbb{P}_{\alpha_t}^{\alpha_{t+1}, y_1, \cdots, y_t}$ is Gaussian with

$$\mathbb{E}[\alpha_t | \alpha_{t+1}, y_1, \cdots, y_t] = \alpha_{t|t} + P_{t|t} T'_{t+1} P_{t+1|t}^{-1} (\alpha_{t+1} - \alpha_{t+1|t}),$$

since

$$\text{Cov}(\alpha_t, \alpha_{t+1} | y_1, \cdots, y_t) = \text{Cov}(\alpha_t, T_{t+1}\alpha_t | y_1, \cdots, y_t) = P_{t|t} T'_{t+1}.$$

Sketch of the proof (2)

$$\mathbb{E}[\alpha_t|\alpha_{t+1}, y_1, \cdots, y_T] = \mathbb{E}[\alpha_t|\alpha_{t+1}, y_1, \cdots, y_t].$$

$$\begin{array}{rcl} y_{t+1} & = & f_1(\alpha_{t+1}, u_{t+1}), \\ y_{t+j} & = & f_j(\alpha_{t+1}, u_{t+j}, v_{t+j}, \cdots, v_{t+2}), j \geq 2, \end{array}$$

with each $f_j(.)$ linear functions. We have

$$\alpha_t = \mathbb{E}[\alpha_t|y_1, \cdots, y_t, \alpha_{t+1}] + e_t,\ e_t \perp (y_1, \cdots, y_t, \alpha_{t+1}).$$

Besides

$$\begin{array}{rcl} e_t = g(\alpha_t, y_1, \cdots, y_t, \alpha_{t+1}) & \Rightarrow & e_t \perp \{(u_{t+j})_{j\geq 1}, (v_{t+j})_{j\geq 2}\} \\ & \Rightarrow & e_t \perp y_{t+j} \text{ for } j \geq 1 \\ & \Rightarrow & \mathbb{E}[e_t|y_1, \cdots, y_T, \alpha_{t+1}] = 0. \end{array}$$

Sketch of the proof (3)

Consequently, we obtain

$$\mathbb{E}[\alpha_t|\alpha_{t+1}, y_1, \cdots, y_t, y_{t+1}, \cdots, y_T] = \alpha_{t|t} + \bar{F}_t(\alpha_{t+1} - \alpha_{t+1|t}).$$

By deconditioning with respect to α_{t+1}, we get

$$\alpha_{t|T} = \alpha_{t|t} + \bar{F}_t(\alpha_{t+1|T} - \alpha_{t+1|t}),\ t < T.$$

Sketch of the proof (4): variance of the smoothing error

$$\alpha_t - \alpha_{t|T} = \alpha_t - \alpha_{t|t} - \bar{F}_t(\alpha_{t+1|T} - \alpha_{t+1|t})$$
$$\Rightarrow \alpha_t - \alpha_{t|T} + \bar{F}_t\alpha_{t+1|T} = \alpha_t - \alpha_{t|t} + \bar{F}_t\alpha_{t+1|t}$$
$$\Rightarrow \mathrm{Var}(\alpha_t - \alpha_{t|n}) + \bar{F}_t \mathrm{Var}(\alpha_{t+1|n})\bar{F}_t' = \mathrm{Var}(\alpha_t - \alpha_{t|t}) + \bar{F}_t \mathrm{Var}(\alpha_{t+1|t})\bar{F}_t'.$$

We have $\mathrm{Cov}(\alpha_{t+1}, \alpha_{t+1|T}) = \mathrm{Var}(\alpha_{t+1|T})$. Consequently

$$\begin{aligned}\mathrm{Var}(\alpha_{t+1|T} - \alpha_{t+1}) &= \mathrm{Var}(\alpha_{t+1|T}) + \mathrm{Var}(\alpha_{t+1}) \\ &- \mathrm{Cov}(\alpha_{t+1|T}, \alpha_{t+1}) - \mathrm{Cov}(\alpha_{t+1}, \alpha_{t+1|T}) \\ &= \mathrm{Var}(\alpha_{t+1}) - \mathrm{Var}(\alpha_{t+1|T}).\end{aligned}$$

And $\mathrm{Var}(\alpha_{t+1|t} - \alpha_{t+1}) = \mathrm{Var}(\alpha_{t+1}) - \mathrm{Var}(\alpha_{t+1|t})$. Hence

$$\mathrm{Var}(\alpha_{t+1|T} - \alpha_{t+1}) - \mathrm{Var}(\alpha_{t+1|t} - \alpha_{t+1}) = \mathrm{Var}(\alpha_{t+1|t}) - \mathrm{Var}(\alpha_{t+1|T}).$$

Sketch of the proof (5): variance of the smoothing error

$$\mathrm{Var}(\alpha_{t+1|T}-\alpha_{t+1})-\mathrm{Var}(\alpha_{t+1|t}-\alpha_{t+1}) = \mathrm{Var}(\alpha_{t+1|t})-\mathrm{Var}(\alpha_{t+1|T}).$$

Now

$$\begin{aligned}
P_{t+1|t} &= \mathrm{Var}(\alpha_{t+1}-\alpha_{t+1|t}) = \mathrm{Var}(\alpha_{t+1}) - \mathrm{Var}(\alpha_{t+1|t}),\\
P_{t|T} &= \mathrm{Var}(\alpha_t-\alpha_{t|T}) = \mathrm{Var}(\alpha_t) - \mathrm{Var}(\alpha_{t|T}),\\
P_{t|t} &= \mathrm{Var}(\alpha_t-\alpha_{t|t}) = \mathrm{Var}(\alpha_t) - \mathrm{Var}(\alpha_{t|t}).
\end{aligned}$$

1. General form of the filter

2. Statistical inference
   - ML estimation

Parametric model

Suppose the model is parameterized by $\theta \in \Theta \subset \mathbb{R}^d$. Then

$$\begin{cases} y_t &= M(\theta)\alpha_t + d(\theta) + u_t, \\ \alpha_t &= T(\theta)\alpha_{t-1} + c(\theta) + R(\theta)v_t, \end{cases}$$

where

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \mathcal{N}_{\mathbb{R}^{N+K}}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H(\theta) & 0 \\ 0 & Q(\theta) \end{pmatrix}\right).$$

We observe $y_1, \cdots, y_T$ and for some given functions M, d, T, c, H, Q, the problem consists in estimating θ.

Likelihood function

Inital values: $\epsilon_1(\theta)$ and $F_1(\theta)$, the Gaussian likelihood corresponds to

$$\begin{aligned}\mathcal{L}_T(\theta) &= \mathcal{L}_T(y_1, \cdots, y_T; \theta) \\ &= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}|F_t(\theta)|^{\frac{1}{2}}} \exp\{-\frac{1}{2}\epsilon'_t(\theta) F'_{t|t-1} \epsilon_t(\theta)\},\end{aligned}$$

with

$$\begin{aligned}\epsilon_t(\theta) &= y_t - \mathbb{E}_\theta[y_t|y_1, \cdots, y_{t-1}] = y_t - y_{t|t-1}(\theta), \\ F_{t|t-1}(\theta) &= \mathrm{Var}_\theta(y_t|y_1, \cdots, y_{t-1}).\end{aligned}$$

M-estimator

A M-estimator (MLE) of θ satisfies the optimization problem

$$\begin{aligned} \hat{\theta}_T &= \arg\max_{\theta\in\Theta} \mathcal{L}_T(\theta) \\ \Leftrightarrow \hat{\theta}_T &= \arg\min_{\theta\in\Theta} \{-\log(\mathcal{L}_T(\theta))\} \\ &= \arg\min_{\theta\in\Theta} \frac{1}{T}\sum_{t=1}^{T} l(y_t, \cdots, y_1; \theta), \end{aligned}$$

such that

$$l(y_t, \cdots, y_1; \theta) = \epsilon_t'(\theta) F_{t|t-1}^{-1}(\theta)\epsilon_t(\theta) + \log(|F_{t|t-1}(\theta)|).$$

Kalman filtering

The Kalman filter enables to compute $\epsilon_t(\theta)$ and $F_{t|t-1}(\theta)$ for any θ.

Numerical procedure to solve the problem: Newton-Raphson (several approximations of the Hessian), stochastic algorithm.

The theoretical properties of the estimator require additional assumptions regarding the model.

Application: MA(1)

Let

$$y_t = \mu + \epsilon_t + b\epsilon_{t-1},$$

with (ϵ_t) a white noise with variance σ^2. The state-space representation is

$$\begin{cases} y_t &= \mu + M\alpha_t, \\ \alpha_t &= T\alpha_{t-1} + (\epsilon_t, 0)', \end{cases}$$

with $M = (1, b)$, $\alpha_t = (\epsilon_t, \epsilon_{t-1})'$ and $T = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. Regarding the general state space model:

$$d_t = \mu, u_t = 0, c_t = (0,0)', v_t = (\epsilon_t, 0)', H_t = 0, Q_t = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Application: MA(1)

$$
\begin{array}{rcl}
\mathbb{E}_{t-1}[y_t] & = & \mu + b\epsilon_{t-1|t-1}. \\
\mathrm{Var}_{t-1}(y_t) & = & \sigma^2 + b^2 p_{t-1},\ p_{t-1} = \mathrm{Var}(\epsilon_{t-1|t-1}). \\
\mathbb{P}_{(y_t,\epsilon_t)'}^{|y_1,\cdots,y_{t-1}} & = & \mathcal{N}_{\mathbb{R}^2}\left(\begin{pmatrix} \mu + b\epsilon_{t-1|t-1} \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 + b^2 p_{t-1} & \sigma^2 \\ \sigma^2 & \sigma^2 \end{pmatrix}\right).
\end{array}
$$

We then obtain

$$
\left\{
\begin{array}{rcl}
\epsilon_{t|t} & = & \dfrac{\sigma^2}{\sigma^2 + b^2 p_{t-1}}(y_t - b\epsilon_{t-1|t-1} - \mu), t \geq 1, \\
p_t & = & \sigma^2 - \dfrac{\sigma^4}{\sigma^2 + b^2 p_{t-1}}, t \geq 1,
\end{array}
\right.
$$

with initial values $\epsilon_{0|0} = 0$ and $p_0 = \sigma^2$.

Asymptotic behavior of (p_t) and $\epsilon_{t|t}$

When $|b| < 1$, then

$$\lim_{t\to\infty} p_t = \lim_{t\to\infty} \mathbb{E}[(\epsilon_t - \epsilon_{t|t})^2] = 0.$$

This implies

$$\epsilon_t - \epsilon_{t|t} \to 0.$$

This implies that we can approximate ϵ_t for t large enough.

Estimation of the MA(1)

Let $\theta = (\mu, b, \sigma^2)'$. The M-estimator minimizes

$$-\log(\mathcal{L}_T(\theta)) = \frac{1}{T}\sum_{t=1}^{T}\frac{y_t - \mu - b\epsilon_{t-1|t-1}}{\sigma^2 + b^2 p_{t-1}} + \log|\sigma^2 + b^2 p_{t-1}|,$$

where p_{t-1} and $\epsilon_{t-1|t-1}$ are computed using the Kalman filter.