Dynamic Models with Latent Variables

Jean-Michel Zakoian

CREST-ENSAE

The Kalman Filter

State-space models

General form:

$$\begin{cases} \boldsymbol{y}_t = \boldsymbol{M}_t\boldsymbol{\alpha}_t + \boldsymbol{d}_t + \boldsymbol{u}_t & \text{Measurement equation} \\ \boldsymbol{\alpha}_t = \boldsymbol{T}_t\boldsymbol{\alpha}_{t-1} + \boldsymbol{c}_t + \boldsymbol{R}_t\boldsymbol{v}_t, & \text{Transition equation} \end{cases}$$

where $\boldsymbol{y}_t \in \mathbb{R}^N$, $\boldsymbol{\alpha}_t \in \mathbb{R}^m$ (**state vector**), $(\boldsymbol{u}_t)$ and $(\boldsymbol{v}_t)$ are two sequences of independent variables, valued in $\mathbb{R}^N$ and $\mathbb{R}^K$ such that

$$E(\boldsymbol{u}_t) = \boldsymbol{0}_N, \quad E(\boldsymbol{v}_t) = \boldsymbol{0}_K, \quad \mathrm{Var}(\boldsymbol{u}_t) = \boldsymbol{H}_t, \quad \mathrm{Var}(\boldsymbol{v}_t) = \boldsymbol{Q}_t,$$

$\boldsymbol{M}_t, \boldsymbol{T}_t$ and $\boldsymbol{R}_t$ are non-random $N \times m$, $m \times m$ and $m \times K$ matrices, $\boldsymbol{d}_t \in \mathbb{R}^N$, $\boldsymbol{c}_t \in \mathbb{R}^m$ are non-random vectors.

Aims of the Kalman filter

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

(i) **predicting** the value of the state vector at time t, given observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}$;

(ii) **filtering**, that is, estimating $\boldsymbol{\alpha}_t$ given observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t$;

(iii) **smoothing**, that is, estimating $\boldsymbol{\alpha}_t$ given observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T$; with $T > t$.

Assumptions

To implement this algorithm, additional normality and independence assumptions will be made:

- $(\boldsymbol{u}_t, \boldsymbol{v}_t)$ is an **independent Gaussian** sequence such that

$$\begin{pmatrix} \boldsymbol{u}_t \\ \boldsymbol{v}_t \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{H}_t & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_t \end{pmatrix} \right),$$

- The initial distribution of the state vector is Gaussian and is independent from $(\boldsymbol{u}_t)$ and $(\boldsymbol{v}_t)$:

$$\boldsymbol{\alpha}_0 \rightsquigarrow \mathcal{N}(\boldsymbol{a}_0, \boldsymbol{P}_0), \quad \boldsymbol{\alpha}_0 \perp (\boldsymbol{u}_t), (\boldsymbol{v}_t).$$

- For all t, the matrix $\boldsymbol{H}_t$ is positive definite.

Notations: conditional moments with respect to observations

For $t \geq 1$,

$$\begin{aligned} \boldsymbol{\alpha}_{t|t} &= E(\boldsymbol{\alpha}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t), \\ \boldsymbol{P}_{t|t} &= \mathrm{Var}(\boldsymbol{\alpha}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t). \end{aligned}$$

For $t > 1$,

$$\begin{aligned} \boldsymbol{\alpha}_{t|t-1} &= E(\boldsymbol{\alpha}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}), \\ \boldsymbol{P}_{t|t-1} &= \mathrm{Var}(\boldsymbol{\alpha}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}). \end{aligned}$$

Let

$$\boldsymbol{\alpha}_{1|0} = E(\boldsymbol{\alpha}_1), \quad \boldsymbol{P}_{1|0} = \mathrm{Var}(\boldsymbol{\alpha}_1).$$

The aim is to compute recursively these sequences.

First step

Taking the conditional expectation with respect to $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}$ in the transition equation yields:

$$\boldsymbol{\alpha}_{t|t-1} = \boldsymbol{T}_t \boldsymbol{\alpha}_{t-1|t-1} + \boldsymbol{c}_t$$

and by taking the conditional variance:

$$\boldsymbol{P}_{t|t-1} = \boldsymbol{T}_t \boldsymbol{P}_{t-1|t-1} \boldsymbol{T}'_t + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}'_t.$$

These equations are called **prediction equations.**
Thus the conditional moments of $\boldsymbol{y}_t$:

$$\boldsymbol{y}_{t|t-1} := E(\boldsymbol{y}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}) = \boldsymbol{M}_t \boldsymbol{\alpha}_{t|t-1} + \boldsymbol{d}_t$$

and

$$\boldsymbol{F}_{t|t-1} := \mathrm{Var}(\boldsymbol{y}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}) = \boldsymbol{M}_t \boldsymbol{P}_{t|t-1} \boldsymbol{M}'_t + \boldsymbol{H}_t.$$

We also have

$$\mathrm{Cov}(\boldsymbol{\alpha}_t, \boldsymbol{y}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}) = \boldsymbol{P}_{t|t-1} \boldsymbol{M}'_t.$$

Conditional distributions of the components of a Gaussian vector

Let

$$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{XX}} & \boldsymbol{\Sigma}_{\boldsymbol{XY}} \\ \boldsymbol{\Sigma}_{\boldsymbol{YX}} & \boldsymbol{\Sigma}_{\boldsymbol{YY}} \end{pmatrix}\right).$$

Then the distribution of X conditional on $\boldsymbol{Y} = \boldsymbol{y}$ is

$$\mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{X}} + \boldsymbol{\Sigma}_{\boldsymbol{XY}} \boldsymbol{\Sigma}_{\boldsymbol{YY}}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{\boldsymbol{Y}}), \quad \boldsymbol{\Sigma}_{\boldsymbol{XX}} - \boldsymbol{\Sigma}_{\boldsymbol{XY}} \boldsymbol{\Sigma}_{\boldsymbol{YY}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{YX}}\right).$$

Conditional law of $(\boldsymbol{y}_t, \boldsymbol{\alpha}_t)$

We have

$$\left(\boldsymbol{y}_t, \boldsymbol{\alpha}_t, \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_1\right) = F\left(\boldsymbol{\alpha}_0, \boldsymbol{u}_t, \ldots, \boldsymbol{u}_1, \boldsymbol{v}_t, \ldots, \boldsymbol{v}_1\right)$$

where F is linear.

The vector $\left(\boldsymbol{y}_t, \boldsymbol{\alpha}_t, \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_1\right)$ is thus Gaussian.

The law of $(\boldsymbol{y}_t, \boldsymbol{\alpha}_t)$ conditional on $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}$ is thus also Gaussian.

Second step: updating the prediction formulas

New observation at time t: $\boldsymbol{y}_t$

$\boldsymbol{\alpha}_t$ is Gaussian cond. on $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}$ and $\boldsymbol{y}_t$

$$\boldsymbol{\alpha}_{t|t} = \boldsymbol{\alpha}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{M}_t'\boldsymbol{F}_{t|t-1}^{-1}(\boldsymbol{y}_t - \boldsymbol{M}_t\boldsymbol{\alpha}_{t|t-1} - \boldsymbol{d}_t)$$

and

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{M}_t'\boldsymbol{F}_{t|t-1}^{-1}\boldsymbol{M}_t\boldsymbol{P}_{t|t-1}.$$

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:

$$\boldsymbol{\alpha}_{1|0} = \boldsymbol{T}_1\boldsymbol{a}_0 + \boldsymbol{c}_1, \quad \boldsymbol{P}_{1|0} = \boldsymbol{T}_1\boldsymbol{P}_0\boldsymbol{T}_1' + \boldsymbol{R}_1\boldsymbol{Q}_1\boldsymbol{R}_1'.$$

The sequences $(\boldsymbol{\alpha}_{t|t-1})$, $(\boldsymbol{P}_{t|t-1})$, $(\boldsymbol{\alpha}_{t|t})$, and $(\boldsymbol{P}_{t|t})$ are computed recursively for $t = 1, \dots, n$

Initial values :

$$\boldsymbol{\alpha}_{1|0} = \boldsymbol{T}_1 \boldsymbol{a}_0 + \boldsymbol{c}_1, \quad \boldsymbol{P}_{1|0} = \boldsymbol{T}_1 \boldsymbol{P}_0 \boldsymbol{T}_1' + \boldsymbol{R}_1 \boldsymbol{Q}_1 \boldsymbol{R}_1'$$

Prediction equations: using $\boldsymbol{y}_1, \dots, \boldsymbol{y}_{t-1}$

$$\boldsymbol{\alpha}_{t|t-1} = \boldsymbol{T}_t \boldsymbol{\alpha}_{t-1|t-1} + \boldsymbol{c}_t$$

$$\boldsymbol{P}_{t|t-1} = \boldsymbol{T}_t \boldsymbol{P}_{t-1|t-1} \boldsymbol{T}_t' + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t'$$

$$\boldsymbol{F}_{t|t-1} = \boldsymbol{M}_t \boldsymbol{P}_{t|t-1} \boldsymbol{M}_t' + \boldsymbol{H}_t$$

Updating equations: using also $\boldsymbol{y}_t$,

$$\boldsymbol{\alpha}_{t|t} = \boldsymbol{\alpha}_{t|t-1} + \boldsymbol{P}_{t|t-1} \boldsymbol{M}_t' \boldsymbol{F}_{t|t-1}^{-1} (\boldsymbol{y}_t - \boldsymbol{M}_t \boldsymbol{\alpha}_{t|t-1} - \boldsymbol{d}_t)$$

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1} \boldsymbol{M}_t' \boldsymbol{F}_{t|t-1}^{-1} \boldsymbol{M}_t \boldsymbol{P}_{t|t-1}$$

Direct computation of the sequences $(\boldsymbol{\alpha}_{t|t-1})$ and $(\boldsymbol{P}_{t|t-1})$

$$\begin{cases} \boldsymbol{\alpha}_{t|t-1} = \boldsymbol{T}_t \boldsymbol{\alpha}_{t-1|t-2} + c_t + \boldsymbol{K}_t(\boldsymbol{y}_{t-1} - \boldsymbol{M}_{t-1}\boldsymbol{\alpha}_{t-1|t-2} - \boldsymbol{d}_{t-1}) \\ \boldsymbol{P}_{t|t-1} = \boldsymbol{T}_t \boldsymbol{P}_{t-1|t-2} \boldsymbol{T}_t' - \boldsymbol{K}_t \boldsymbol{F}_{t-1|t-2} \boldsymbol{K}_t' + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t' \end{cases}$$

where

$$\boldsymbol{F}_{t-1|t-2} = \boldsymbol{M}_{t-1}\boldsymbol{P}_{t-1|t-2}\boldsymbol{M}_{t-1}' + \boldsymbol{H}_{t-1}$$

$$\boldsymbol{K}_t = \boldsymbol{T}_t \boldsymbol{P}_{t-1|t-2}\boldsymbol{M}_{t-1}'\boldsymbol{F}_{t-1|t-2}^{-1}$$

$\boldsymbol{K}_t$ is called **gain matrix**.

Remark: if $\boldsymbol{H}_{t-1}$ is "large", $\boldsymbol{K}_t$ will be "small".

Correlation between the noise sequences

The assumption that the noises are mutually uncorrelated can be relaxed:

$$\begin{pmatrix} \boldsymbol{u}_t \\ \boldsymbol{v}_t \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{H}_t & \boldsymbol{G}_t' \\ \boldsymbol{G}_t & \boldsymbol{Q}_t \end{pmatrix}\right).$$

Prediction equations:

$$\boldsymbol{\alpha}_{t|t-1} = \boldsymbol{T}_t \boldsymbol{\alpha}_{t-1|t-1} + \boldsymbol{c}_t, \qquad \boldsymbol{P}_{t|t-1} = \boldsymbol{T}_t \boldsymbol{P}_{t-1|t-1} \boldsymbol{T}_t' + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t'$$

$$\boldsymbol{F}_{t|t-1} = \boldsymbol{M}_t P_{t|t-1} \boldsymbol{M}_t' + \boldsymbol{H}_t + \boldsymbol{M}_t \boldsymbol{R}_t \boldsymbol{G}_t + \boldsymbol{G}_t' \boldsymbol{R}_t' \boldsymbol{M}_t'$$

Updating equations:

$$\boldsymbol{\alpha}_{t|t} = \boldsymbol{\alpha}_{t|t-1} + (\boldsymbol{P}_{t|t-1} \boldsymbol{M}_t' + \boldsymbol{R}_t \boldsymbol{G}_t) \boldsymbol{F}_{t|t-1}^{-1} (\boldsymbol{y}_t - \boldsymbol{M}_t \boldsymbol{\alpha}_{t|t-1} - \boldsymbol{d}_t)$$

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - (\boldsymbol{P}_{t|t-1} \boldsymbol{M}_t' + \boldsymbol{R}_t \boldsymbol{G}_t) \boldsymbol{F}_{t|t-1}^{-1} (\boldsymbol{M}_t \boldsymbol{P}_{t|t-1} + \boldsymbol{G}_t' \boldsymbol{R}_t').$$

Can the normality assumption be relaxed?

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

$$\|\boldsymbol{X} - E(\boldsymbol{X} \mid \boldsymbol{Y})\|_2^2 = \min_{\phi \in \Phi} \|\boldsymbol{X} - \phi(\boldsymbol{Y})\|_2^2$$

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

The linear conditional expectation $EL(\boldsymbol{X} \mid \boldsymbol{Y})$ is characterized by the same program but with ϕ linear:

$$\|\boldsymbol{X} - EL(\boldsymbol{X} \mid \boldsymbol{Y})\|_2^2 = \min_{\boldsymbol{A},\boldsymbol{b}} \|\boldsymbol{X} - \boldsymbol{A}\boldsymbol{Y} - \boldsymbol{b}\|_2^2.$$

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 $(\boldsymbol{X}, \boldsymbol{Y})$.

It follows that

$$EL(\boldsymbol{X} \mid \boldsymbol{Y}) = \boldsymbol{\mu}_{\boldsymbol{X}} + \boldsymbol{\Sigma}_{\boldsymbol{X}\boldsymbol{X}} \boldsymbol{\Sigma}_{\boldsymbol{Y}\boldsymbol{Y}}^{-1} (\boldsymbol{Y} - \boldsymbol{\mu}_{\boldsymbol{Y}}).$$

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

$$EL(\boldsymbol{y}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}) = \boldsymbol{M}_t \boldsymbol{\alpha}_{t|t-1} + \boldsymbol{d}_t$$

and the variance of the prediction error:

$$\mathrm{var}(\boldsymbol{y}_t - EL(\boldsymbol{y}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1})) = \boldsymbol{F}_{t|t-1} = \boldsymbol{M}_t \boldsymbol{P}_{t|t-1} \boldsymbol{M}_t' + \boldsymbol{H}_t.$$

Prediction

The Kalman filter can be used to predict at any horizon.

To simplify, let $\boldsymbol{c}_t = \boldsymbol{d}_t = 0$, $\boldsymbol{T}_t = \boldsymbol{T}$ et $\boldsymbol{M}_t = \boldsymbol{M}$ for all t:

$$\left\{ \begin{array}{lcl} \boldsymbol{y}_t & = & \boldsymbol{M}\boldsymbol{\alpha}_t + \boldsymbol{u}_t, \\ \boldsymbol{\alpha}_t & = & \boldsymbol{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{R}_t \boldsymbol{v}_t. \end{array} \right.$$

For any $h \geq 0$,

$$\boldsymbol{\alpha}_{t+h} = \boldsymbol{T}^{h+1}\boldsymbol{\alpha}_{t-1} + \sum_{i=0}^{h} \boldsymbol{T}^{h-i}\boldsymbol{R}_{t+i}\boldsymbol{v}_{t+i},$$

hence

$$\boldsymbol{\alpha}_{t+h|t-1} = E(\boldsymbol{\alpha}_{t+h} \mid \boldsymbol{y}_1, \ldots \boldsymbol{y}_{t-1}) = \boldsymbol{T}^{h+1}\boldsymbol{\alpha}_{t-1|t-1}.$$

Prediction

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

$$\begin{aligned} \boldsymbol{P}_{t+h|t-1} &= \mathrm{Var}(\boldsymbol{\alpha}_{t+h} - \boldsymbol{\alpha}_{t+h|t-1}) \\ &= \boldsymbol{T}^{h+1}\boldsymbol{P}_{t-1|t-1}(\boldsymbol{T}^{h+1})' + \sum_{i=0}^{h} \boldsymbol{T}^{h-i}\boldsymbol{R}_{t+i}\boldsymbol{Q}_{t+i}(\boldsymbol{T}^{h-i}\boldsymbol{R}_{t+i})'. \end{aligned}$$

Moreover $\boldsymbol{y}_{t+h} = \boldsymbol{M}\boldsymbol{\alpha}_{t+h} + \boldsymbol{u}_{t+h}$, thus

$$\boldsymbol{y}_{t+h|t-1} = E(\boldsymbol{y}_{t+h} \mid \boldsymbol{y}_1, \ldots \boldsymbol{y}_{t-1}) = \boldsymbol{M}\boldsymbol{\alpha}_{t+h|t-1} = \boldsymbol{M}\boldsymbol{T}^{h+1}\boldsymbol{\alpha}_{t-1|t-1}.$$

The prediction error is $\boldsymbol{y}_{t+h} - \boldsymbol{y}_{t+h|t-1} = \boldsymbol{M}(\boldsymbol{\alpha}_{t+h} - \boldsymbol{\alpha}_{t+h|t-1}) + \boldsymbol{u}_{t+h}$ and its variance is

$$\mathrm{Var}(\boldsymbol{y}_{t+h} - \boldsymbol{y}_{t+h|t-1}) = \boldsymbol{M}\boldsymbol{P}_{t+h|t-1}\boldsymbol{M}' + \boldsymbol{H}_{t+h}.$$

Smoothing

The updating formula provides the filtered value $\boldsymbol{\alpha}_{t|t}$ of $\boldsymbol{\alpha}_t$.

For certain applications, it is important to "smooth" $\boldsymbol{\alpha}_t$ using the posterior observations.

Let

$$\boldsymbol{\alpha}_{t|n} \quad = \quad E(\boldsymbol{\alpha}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n), \quad \boldsymbol{P}_{t|n} \quad = \quad \mathrm{Var}(\boldsymbol{\alpha}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n).$$

Steps for computing $\boldsymbol{\alpha}_{t|n}$

- $E(\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t+1} | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t)$ [already known]

  $\Downarrow$ Using the normality

- $E(\boldsymbol{\alpha}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t, \boldsymbol{\alpha}_{t+1})$

  $\Downarrow$ Using a lemma

- $E(\boldsymbol{\alpha}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t, \boldsymbol{y}_{t+1}, \ldots, \boldsymbol{y}_n, \boldsymbol{\alpha}_{t+1})$

  $\Downarrow$ By deconditioning

- $E(\boldsymbol{\alpha}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$

Algorithm

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

$$\boldsymbol{\alpha}_{t|n} = \boldsymbol{\alpha}_{t|t} + \tilde{\boldsymbol{F}}_t(\boldsymbol{\alpha}_{t+1|n} - \boldsymbol{\alpha}_{t+1|t}), \qquad t < n$$

and

$$\boldsymbol{P}_{t|n} = \boldsymbol{P}_{t|t} + \tilde{\boldsymbol{F}}_t(\boldsymbol{P}_{t+1|n} - \boldsymbol{P}_{t+1|t})\tilde{\boldsymbol{F}}'_t, \qquad t < n$$

where

$$\tilde{\boldsymbol{F}}_t = \boldsymbol{P}_{t|t}\boldsymbol{T}'_{t+1}\boldsymbol{P}^{-1}_{t+1|t}, \qquad t < n.$$

Proof (1)

- Using the normality
  The law of $\boldsymbol{\alpha}_t$ conditional to $\boldsymbol{\alpha}_{t+1}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t$ is Gaussian with mean

$$E(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t+1}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t) = \boldsymbol{\alpha}_{t|t} + \underbrace{\boldsymbol{P}_{t|t} \boldsymbol{T}'_{t+1} P^{-1}_{t+1|t}}_{\tilde{F}_t} (\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_{t+1|t}),$$

  because

$$\mathrm{Cov}(\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t+1} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t) = \mathrm{Cov}(\boldsymbol{\alpha}_t, \boldsymbol{T}_{t+1} \boldsymbol{\alpha}_t \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t) = \boldsymbol{P}_{t|t} \boldsymbol{T}'_{t+1}.$$

Proof (2)

- To predict $\boldsymbol{\alpha}_t$, the knowledge of $\boldsymbol{y}_{t+1},\ldots,\boldsymbol{y}_n$ does not convey additional information with respect to $\boldsymbol{\alpha}_{t+1},\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t$:

$$E(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t+1},\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n) = E(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t+1},\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t).$$

$\boldsymbol{y}_{t+1} = f_1(\boldsymbol{\alpha}_{t+1},\boldsymbol{u}_{t+1})$ and

$$\boldsymbol{y}_{t+j} = f_j(\boldsymbol{\alpha}_{t+1},\boldsymbol{u}_{t+j},\boldsymbol{v}_{t+j},\ldots,\boldsymbol{v}_{t+2}), \quad j \geq 2,$$

where the f_j are linear.
We have

$$\boldsymbol{\alpha}_t = E(\boldsymbol{\alpha}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t,\boldsymbol{\alpha}_{t+1}) + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \perp (\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t,\boldsymbol{\alpha}_{t+1}).$$

$$\begin{aligned}
\boldsymbol{e}_t = g(\boldsymbol{\alpha}_t,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t,\boldsymbol{\alpha}_{t+1}) \quad &\Rightarrow \quad \boldsymbol{e}_t \perp \{(\boldsymbol{u}_{t+j})_{j\geq 1},(\boldsymbol{v}_{t+j})_{j\geq 2}\} \\
&\Rightarrow \quad \boldsymbol{e}_t \perp \boldsymbol{y}_{t+j} \quad \text{for} \quad j \geq 1 \\
&\Rightarrow \quad E(\boldsymbol{e}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n,\boldsymbol{\alpha}_{t+1}) = 0.
\end{aligned}$$

Proof (3)

- Thus

$$E(\boldsymbol{\alpha}_t \mid \boldsymbol{\alpha}_{t+1}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t, \boldsymbol{y}_{t+1} \ldots, \boldsymbol{y}_n) = \boldsymbol{\alpha}_{t|t} + \tilde{F}_t(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_{t+1|t}).$$

By deconditioning with respect to $\boldsymbol{\alpha}_{t+1}$ we get

$$\boldsymbol{\alpha}_{t|n} = \boldsymbol{\alpha}_{t|t} + \tilde{\boldsymbol{F}}_t(\boldsymbol{\alpha}_{t+1|n} - \boldsymbol{\alpha}_{t+1|t}), \qquad t < n.$$

Proof (4): variance of the smoothing error

$$\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|n} = \boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t} - \tilde{F}_t(\boldsymbol{\alpha}_{t+1|n} - \boldsymbol{\alpha}_{t+1|t})$$

$$\Rightarrow \boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|n} + \tilde{F}_t\boldsymbol{\alpha}_{t+1|n} = \boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t} + \tilde{F}_t\boldsymbol{\alpha}_{t+1|t}$$

$$\Rightarrow \mathsf{Var}(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|n}) + \tilde{F}_t\mathsf{Var}(\boldsymbol{\alpha}_{t+1|n})\tilde{F}_t' = \mathsf{Var}(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}) + \tilde{F}_t\mathsf{Var}(\boldsymbol{\alpha}_{t+1|t})\tilde{F}_t'.$$

We have $\mathsf{Cov}(\boldsymbol{\alpha}_{t+1}, \boldsymbol{\alpha}_{t+1|n}) = \mathsf{Var}(\boldsymbol{\alpha}_{t+1|n})$ hence

$$\begin{aligned}
\mathsf{Var}(\boldsymbol{\alpha}_{t+1|n} - \boldsymbol{\alpha}_{t+1}) &= \mathsf{Var}(\boldsymbol{\alpha}_{t+1|n}) + \mathsf{Var}(\boldsymbol{\alpha}_{t+1}) \\
&\quad - \mathsf{Cov}(\boldsymbol{\alpha}_{t+1|n}, \boldsymbol{\alpha}_{t+1}) - \mathsf{Cov}(\boldsymbol{\alpha}_{t+1}, \boldsymbol{\alpha}_{t+1|n}) \\
&= \mathsf{Var}(\boldsymbol{\alpha}_{t+1}) - \mathsf{Var}(\boldsymbol{\alpha}_{t+1|n}).
\end{aligned}$$

Similarly $\mathsf{Var}(\boldsymbol{\alpha}_{t+1|t} - \boldsymbol{\alpha}_{t+1}) = \mathsf{Var}(\boldsymbol{\alpha}_{t+1}) - \mathsf{Var}(\boldsymbol{\alpha}_{t+1|t})$. Then

$$\mathsf{Var}(\boldsymbol{\alpha}_{t+1|n} - \boldsymbol{\alpha}_{t+1}) - \mathsf{Var}(\boldsymbol{\alpha}_{t+1|t} - \boldsymbol{\alpha}_{t+1}) = \mathsf{Var}(\boldsymbol{\alpha}_{t+1|t}) - \mathsf{Var}(\boldsymbol{\alpha}_{t+1|n}).$$

Proof (5): variance of the smoothing error

$$\mathrm{Var}(\boldsymbol{\alpha}_{t+1|n}-\boldsymbol{\alpha}_{t+1})-\mathrm{Var}(\boldsymbol{\alpha}_{t+1|t}-\boldsymbol{\alpha}_{t+1})=\mathrm{Var}(\boldsymbol{\alpha}_{t+1|t})-\mathrm{Var}(\boldsymbol{\alpha}_{t+1|n}).$$

Now

$$\begin{aligned}
\boldsymbol{P}_{t+1|t} &= \mathrm{Var}(\boldsymbol{\alpha}_{t+1}-\boldsymbol{\alpha}_{t+1|t})=\mathrm{Var}(\boldsymbol{\alpha}_{t+1})-\mathrm{Var}(\boldsymbol{\alpha}_{t+1|t}),\\
\boldsymbol{P}_{t|n} &= \mathrm{Var}(\boldsymbol{\alpha}_{t}-\boldsymbol{\alpha}_{t|n})=\mathrm{Var}(\boldsymbol{\alpha}_{t})-\mathrm{Var}(\boldsymbol{\alpha}_{t|n}),\\
\boldsymbol{P}_{t|t} &= \mathrm{Var}(\boldsymbol{\alpha}_{t}-\boldsymbol{\alpha}_{t|t})=\mathrm{Var}(\boldsymbol{\alpha}_{t})-\mathrm{Var}(\boldsymbol{\alpha}_{t|t}).
\end{aligned}$$

Time-homogeneous model

Some simplifications appear in the stationary case. In particular, the updating of a variance matrix (and its inversion) can be avoided, at the price of a mild approximation.

Consider the model with fixed coefficients:

$$\left\{ \begin{array}{rcl} \boldsymbol{y}_t & = & \boldsymbol{M}\boldsymbol{\alpha}_t + \boldsymbol{d} + \boldsymbol{u}_t \\ \boldsymbol{\alpha}_t & = & \boldsymbol{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{c} + \boldsymbol{R}\boldsymbol{v}_t, \end{array} \right. \qquad \begin{pmatrix} \boldsymbol{u}_t \\ \boldsymbol{v}_t \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{H} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q} \end{pmatrix} \right).$$

Because the matrices $\boldsymbol{M}, \boldsymbol{T}, \boldsymbol{d}; \boldsymbol{c}, \boldsymbol{R}, \boldsymbol{H}$ do not change over time, the model is said to be time-homogeneous.

Stationarity of the time-homogeneous model

The state vector $\boldsymbol{\alpha}_t$ satisfies a VAR(1) model.
This model admits a second-order stationary solution if the eigenvalues of the matrix $\boldsymbol{T}$ have modulus strictly less than 1:

$$\rho(\boldsymbol{T}) < 1.$$

The first two moments of the stationary solution are given by

$$E\boldsymbol{\alpha}_t = (I - \boldsymbol{T})^{-1}\boldsymbol{c}, \qquad \mathrm{Var}(\boldsymbol{\alpha}_t) = \boldsymbol{T}\mathrm{Var}(\boldsymbol{\alpha}_t)\boldsymbol{T}' + \boldsymbol{RQR}',$$

Variance of $\boldsymbol{\alpha}_t$

Thus, using the vec operator and the Kronecker product*,

$$\mathrm{vec}\{\mathrm{Var}(\boldsymbol{\alpha}_t)\} = (\boldsymbol{I} - \boldsymbol{T} \otimes \boldsymbol{T})^{-1}\mathrm{vec}(\boldsymbol{RQR}').$$

If $\boldsymbol{\alpha}_0 \rightsquigarrow \mathcal{N}(\boldsymbol{a}_0, \boldsymbol{P}_0)$, with $\boldsymbol{a}_0 = (\boldsymbol{I} - \boldsymbol{T})^{-1}c$ and $\mathrm{vec}(\boldsymbol{P}_0) = (I - \boldsymbol{T} \otimes \boldsymbol{T})^{-1}\mathrm{vec}(\boldsymbol{RQR}')$, then

$$E\boldsymbol{\alpha}_t = (\boldsymbol{I} - \boldsymbol{T})^{-1}\boldsymbol{c}, \qquad \mathrm{Var}(\boldsymbol{\alpha}_t) = \boldsymbol{T}\mathrm{Var}(\boldsymbol{\alpha}_t)\boldsymbol{T}' + \boldsymbol{RQR}',$$

holds for any $t \geq 0$.

*For any matrices A and B, $A \otimes B = (a_{ij}B)$. We have $\mathrm{vec}(ABC) = (C' \otimes A)\mathrm{vec}(B)$ whenever ABC is well defined.

Convergence of the sequence $(\boldsymbol{P}_{t|t-1})$

The updating of the conditional variance of $\boldsymbol{\alpha}_t$ writes

$$\boldsymbol{P}_{t|t-1} = \boldsymbol{T}\{\boldsymbol{P}_{t-1|t-2} - \boldsymbol{P}_{t-1|t-2}\boldsymbol{M}'\boldsymbol{F}_{t-1|t-2}^{-1}\boldsymbol{M}\boldsymbol{P}_{t-1|t-2}\}\boldsymbol{T}' + \boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}'$$

where

$$\boldsymbol{F}_{t-1|t-2} = \boldsymbol{M}\boldsymbol{P}_{t-1|t-2}\boldsymbol{M}' + \boldsymbol{H}.$$

If this sequence converges, $\boldsymbol{P}^* = \lim \boldsymbol{P}_{t|t-1}$ must satisfy the algebraic Ricatti equation

$$\boldsymbol{P}^* = \boldsymbol{T}\{\boldsymbol{P}^* - \boldsymbol{P}^*\boldsymbol{M}'(\boldsymbol{M}\boldsymbol{P}^*\boldsymbol{M}' + \boldsymbol{H})^{-1}\boldsymbol{M}\boldsymbol{P}^*\}\boldsymbol{T}' + \boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}'. \quad (1)$$

Proposition

If $\rho(\boldsymbol{T}) < 1$ and if at least $\boldsymbol{H}$ or $\boldsymbol{Q}$ is positive definite, the sequence $(\boldsymbol{P}_{t|t-1})_{t\geq 1}$ initialized at any semi-positive definite matrix $\boldsymbol{P}_{1|0}$, converges to a unique matrix $\boldsymbol{P}^*$ unique (independent of $\boldsymbol{P}_{1|0}$), satisfying (1).

Consequences

- The rate of convergence of $(\boldsymbol{P}_{t|t-1})$ is shown to be exponential (see Harvey (1989) and references, Section 3.3.3).
- The sequences $(\boldsymbol{F}_{t|t-1})$ and (K_t) also converge, with limits

$$\boldsymbol{F}^* = \boldsymbol{M}\boldsymbol{P}^*\boldsymbol{M}' + \boldsymbol{H}, \quad \boldsymbol{K}^* = \boldsymbol{T}\boldsymbol{P}^*\boldsymbol{M}'\boldsymbol{F}^{*-1}.$$

- If $\boldsymbol{P}_{t|t-1}$ is sufficiently close to $\boldsymbol{P}^*$, the updating formula can be approximated by

$$\boldsymbol{\alpha}_{t|t-1} = \boldsymbol{T}\boldsymbol{\alpha}_{t-1|t-2} + \boldsymbol{K}^*(y_{t-1} - \boldsymbol{M}\boldsymbol{\alpha}_{t-1|t-2} - \boldsymbol{d}) + \boldsymbol{c}.$$

  This avoids the inversion of $\boldsymbol{F}_{t|t-1}$ at every step of the algorithm.
- Criteria for stopping the computations of $\boldsymbol{P}_{t|t-1}$: for instance

$$|\det \boldsymbol{P}_{t+1|t} - \det \boldsymbol{P}_{t|t-1}| < \tau$$

  where $\tau > 0$ is a small number.

Parametric model

The model is now parameterized by a vector $\boldsymbol{\theta}$ belonging to some parameter set $\Theta \in \mathbb{R}^d$. For $\boldsymbol{y}_t \in \mathbb{R}^N$,

$$\left\{ \begin{array}{lcl} \boldsymbol{y}_t & = & \boldsymbol{M}(\boldsymbol{\theta})\boldsymbol{\alpha}_t + \boldsymbol{d}(\boldsymbol{\theta}) + \boldsymbol{u}_t \\ & & \\ \boldsymbol{\alpha}_t & = & \boldsymbol{T}(\boldsymbol{\theta})\boldsymbol{\alpha}_{t-1} + \boldsymbol{c}(\boldsymbol{\theta}) + \boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{v}_t, \end{array} \right. ,$$

with

$$\begin{pmatrix} \boldsymbol{u}_t \\ \boldsymbol{v}_t \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{H}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}(\boldsymbol{\theta}) \end{pmatrix} \right).$$

Given observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$, and for some given functions $\boldsymbol{M}, \boldsymbol{d}, \boldsymbol{T}, \boldsymbol{c}, \boldsymbol{H}, \boldsymbol{Q}$, the problem is to estimate $\boldsymbol{\theta}$.

Likelihood

Conditionally to initial values $\boldsymbol{\epsilon}_1(\boldsymbol{\theta})$ and $\boldsymbol{F}_1(\boldsymbol{\theta})$, the Gaussian likelihood $L_n(\boldsymbol{\theta})$ writes

$$
\begin{aligned}
L_n(\boldsymbol{\theta}) &= L_n(\boldsymbol{\theta};\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n) \\
&= \prod_{t=1}^{n} \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{F}_t(\boldsymbol{\theta})|}} \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}'_t(\boldsymbol{\theta})\boldsymbol{F}_{t|t-1}^{-1}(\boldsymbol{\theta})\boldsymbol{\epsilon}_t(\boldsymbol{\theta})\right),
\end{aligned}
$$

where, for $t > 1$,

$$
\begin{aligned}
\boldsymbol{\epsilon}_t(\boldsymbol{\theta}) &= \boldsymbol{y}_t - E_{\boldsymbol{\theta}}(\boldsymbol{y}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1}) = \boldsymbol{y}_t - \boldsymbol{y}_{t|t-1}(\boldsymbol{\theta}), \\
\boldsymbol{F}_{t|t-1}(\boldsymbol{\theta}) &= \mathrm{Var}_{\boldsymbol{\theta}}(\boldsymbol{y}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1}).
\end{aligned}
$$

ML estimator

A maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ is defined as any measurable solution $\hat{\boldsymbol{\theta}}_n$ of

$$\hat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta} \in \Theta} L_n(\boldsymbol{\theta}).$$

Maximizing the likelihood is equivalent to minimizing

$$\mathbf{l}_n(\boldsymbol{\theta}) = n^{-1} \sum_{t=1}^{n} \ell_t,$$

where

$$\ell_t = \ell_t(\boldsymbol{\theta}) = \boldsymbol{\epsilon}'_t(\boldsymbol{\theta}) \boldsymbol{F}_{t|t-1}^{-1}(\boldsymbol{\theta}) \boldsymbol{\epsilon}_t(\boldsymbol{\theta}) + \log |\boldsymbol{F}_{t|t-1}(\boldsymbol{\theta})|.$$

Using the Kalman filter

The Kalman filter allows to compute $\boldsymbol{\epsilon}_t(\theta)$ and $\boldsymbol{F}_{t|t-1}(\boldsymbol{\theta})$, for any $\boldsymbol{\theta}$.

Numerical optimization procedures may be called for to solve the program.

The theoretical properties of the MLE (consistency and asymptotic normality) require additional assumptions on the model.

MA(1) model

Let

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

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

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

where

$$M = (1, b), \qquad \alpha_t = \begin{pmatrix} \epsilon_t \\ \epsilon_{t-1} \end{pmatrix}, \qquad T = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Thus

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

MA(1) model

- $E_{t-1}(y_t) = \mu + b\epsilon_{t-1|t-1}$.
- $\mathrm{Var}_{t-1}(y_t) = \sigma^2 + b^2 p_{t-1}$ where $p_{t-1} = \mathrm{Var}(\epsilon_{t-1|t-1})$.
- Law of $\begin{pmatrix} y_t \\ \epsilon_t \end{pmatrix}$ given $y_{t-1}, \dots, y_1$:

$$\mathcal{N}\left(\begin{pmatrix} \mu + b\epsilon_{t-1|t-1} \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 + b^2 p_{t-1} & \sigma^2 \\ \sigma^2 & \sigma^2 \end{pmatrix}\right).$$

Thus

$$\begin{cases} \epsilon_{t|t} & = & \frac{\sigma^2}{\sigma^2 + b^2 p_{t-1}}(y_t - b\epsilon_{t-1|t-1} - \mu), \qquad t \geq 1, \\ p_t & = & \sigma^2 - \frac{\sigma^4}{\sigma^2 + b^2 p_{t-1}} = \frac{\sigma^2 b^2 p_{t-1}}{\sigma^2 + b^2 p_{t-1}}, \qquad t \geq 1, \end{cases}$$

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

MA(1) model

Solving

$$p_t = \sigma^2 - \frac{\sigma^4}{\sigma^2 + b^2 p_{t-1}} = \frac{\sigma^2 b^2 p_{t-1}}{\sigma^2 + b^2 p_{t-1}}, \qquad t \geq 1,$$

yields

$$p_t = \frac{\sigma^2}{1 + \frac{1}{b^2} + \dots \frac{1}{b^{2t}}}.$$

Remark: Naive prediction of ϵ_t given $y_1, \dots, y_t$:

$$\hat{\epsilon}_t = y_t - \mu - b\hat{\epsilon}_{t-1}, \qquad t \geq 1,$$

with $\hat{\epsilon}_0 = 0$, instead of

$$\epsilon_{t|t} = \frac{\sigma^2}{\sigma^2 + b^2 p_{t-1}} (y_t - b\epsilon_{t-1|t-1} - \mu).$$

The naive prediction neglects the variability of the previous prediction.

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

When $|b| < 1$: we have

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

It follows that

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

in L^2 when $t \to \infty$.

The Kalman filter thus allows to **approximate** ϵ_t for t large enough.

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

When $|b| > 1$: we have $\lim_{t\to\infty} p_t = \sigma^2(1 - \frac{1}{b^2})$. Thus, in the L^2 sense

$$\epsilon_{t|t} - \frac{1}{b^2}(y_t - b\epsilon_{t-1|t-1} - \mu) \to 0.$$

Interpretation: If, for all t, $\epsilon_{t|t} - \frac{1}{b^2}(y_t - b\epsilon_{t-1|t-1} - \mu) = 0$ we would get,

$$y_{t+1|t} - \mu = \frac{1}{b}\{(y_t - \mu) - (y_{t|t-1} - \mu)\} = \sum_{i=0}^{t} \frac{(-1)^i}{b^i}(y_{t-i} - \mu).$$

The Kalman filter thus provides an approximation, for large t, of the prediction formula obtained with the canonical representation of the MA(1):

$$y_t = \mu + u_t + \frac{1}{b}u_{t-1},$$

where (u_t) is the linear innovation of (y_t) (with $\mathrm{Var}(u_t) = b^2\sigma^2$).

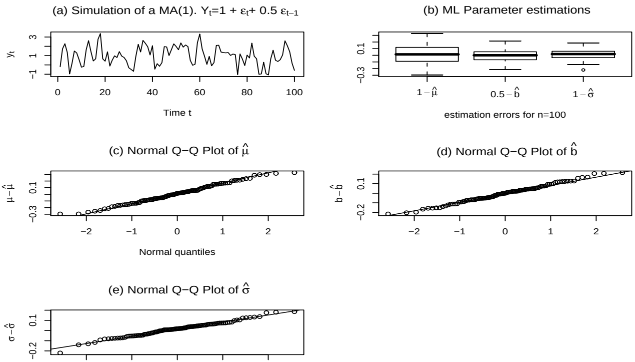
Estimation of the MA(1) model

Let $\theta = (\mu, b, \sigma^2)'$. The ML estimator is obtained by minimizing with respect to θ,

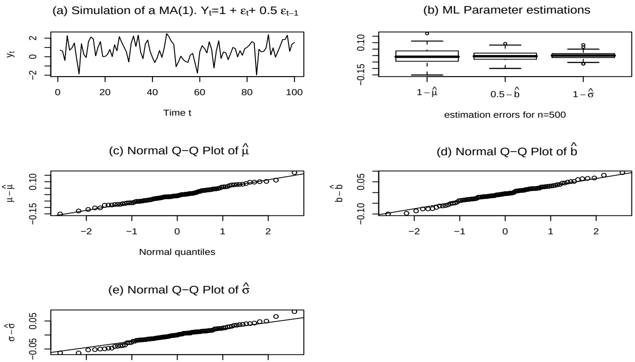
$$\mathbf{l}_n(\theta) = n^{-1} \sum_{t=1}^{n} \frac{(y_t - \mu - b\epsilon_{t-1|t-1})^2}{\sigma^2 + b^2 p_{t-1}} + \log|\sigma^2 + b^2 p_{t-1}|,$$

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

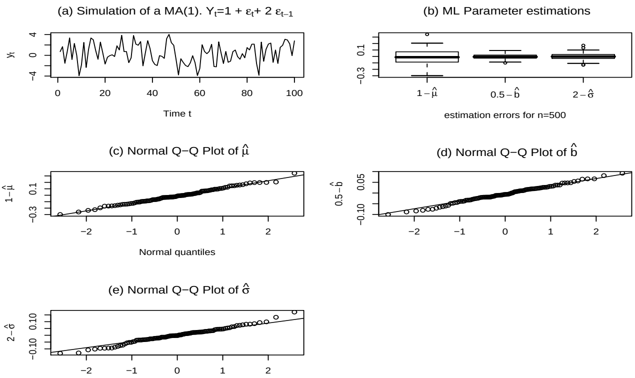
Simulations of length 100 of a MA(1) and ML estimation



Simulations of length 500 of a MA(1) and ML estimation



Simulations of length 500 of a non canonical MA(1) and ML estimation



References on the Kalman filter

- Harvey, A.C. (1990) *Forecasting, structural time series models and the Kalman filter.* Cambridge University Press.
- Harvey, A.C., E. Ruiz, and N. Shephard (1994) Multivariate stochastic variance models. *Review of Economic Studies* 61, 247–264.
- Harvey, A. C. et Pierse, R. G. (1984) Estimating Missing Observations in Economic Time Series. *J. of the American Statistical Association*, 79, 125–131.
- Kalman, R.E. (1960) A new approach of linear filtering and prediction problem. *J. of Basic Engineering* 82, 34-35.
- Kalman, R.E. et R.S. Bucy (1961) New results in linear filtering and prediction theory. *J. of Basic Engineering* 83, 95-108.
- Ruiz, E. (1994) Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models. *J. of Econometrics*, 63, 289–306.