

Macroeconomic Forecasting: Introduction
Belgische Francqui-leerstoel 2019-2020
Universiteit Antwerpen

Siem Jan Koopman

Vrije Universiteit Amsterdam, Tinbergen Institute

Tuesday 1 December 2020

<http://sjkoopman.net>

Introduction – Local Level Model

Program :

- Introduction
- Local level model
- Statistical dynamic properties
- Signal extraction, filtering and prediction
- Likelihood function and parameter estimation.
- Literature : J. Durbin and S.J. Koopman (2012), "Time Series Analysis by State Space Methods", Second Edition, Oxford: Oxford University Press. Chapter 2.

Time Series

A time series is a set of observations y_t , each one recorded at a specific time t .

The observations are ordered over time.

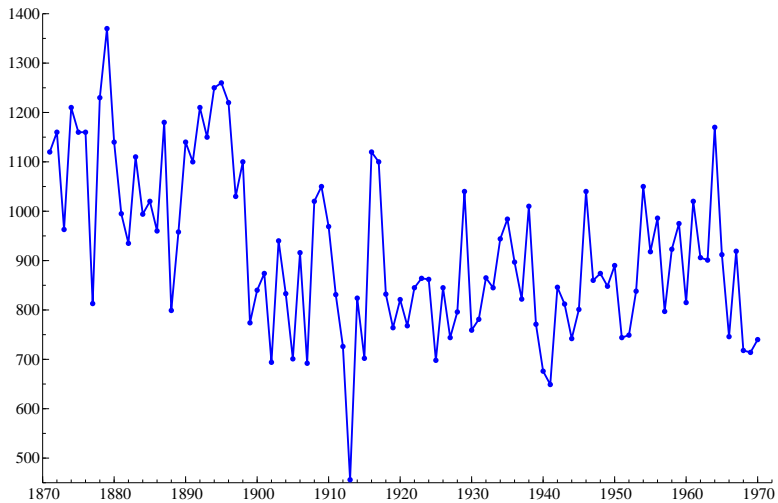
We assume to have n observations, $t = 1, \dots, n$.

Examples of time series are:

- Number of cars sold, every year
- Gross Domestic Product, of a country, every quarter
- Stock price changes, tick-by-tick, within one trading day
- CO₂ emissions, of a country, every month

Time series modeling is relevant for a wide variety of tasks and fields, including economic policy, financial decision making, climate change monitoring, and forecasting

Nile Data



Time Series

A time series for a single entity is typically denoted by

$$y_1, \dots, y_n \Leftrightarrow y_t, \quad t = 1, \dots, n,$$

where t is the time index and n is time series length.

The current value is y_t .

The first lagged value, or **first lag**, is y_{t-1} .

The τ th lagged value, or τ -th lag, is $y_{t-\tau}$ for $\tau = 1, 2, 3, \dots$

The change between period $t - 1$ and period t is $y_t - y_{t-1}$.

This is called the **first difference** denoted by $\Delta y_t = y_t - y_{t-1}$.

In economic time series, we often take the first difference of the logarithm, or the **log-difference**, that is

$$\Delta \log y_t = \log y_t - \log y_{t-1} = \log(y_t/y_{t-1}),$$

is a proxy of **proportional change**, see Appendix.

Percentage change is then $100\Delta \log y_t$.

Autoregressive model: AR(1)

The AR(1) model is given by

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2),$$

with three parameter coefficients μ , ϕ and σ_ε^2 with $0 < \sigma_\varepsilon < \infty$.

Stationarity condition: $|\phi| < 1$.

Statistical dynamic properties:

- Mean $\mathbb{E}(y_t) = \mu / (1 - \phi)$; in case $\mu = 0$, $\mathbb{E}(y_t) = 0$;
- Variance $\mathbb{V}\text{ar}(y_t) = \sigma_\varepsilon^2 / (1 - \phi^2)$;
- Autocovariance lag 1 is $\mathbb{C}\text{ov}(y_t, y_{t-1}) = \phi \sigma_\varepsilon^2 / (1 - \phi^2)$;
- and for lag $\tau = 2, 3, 4, \dots$ is $\mathbb{C}\text{ov}(y_t, y_{t-\tau}) = \phi^\tau \sigma_\varepsilon^2 / (1 - \phi^2)$;
- Autocorrelation lag $\tau = 1, 2, 3, \dots$ is $\mathbb{C}\text{orr}(y_t, y_{t-\tau}) = \phi^\tau$.

Autoregressive model: AR(1)

The AR(1) model is given by

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2),$$

with three parameter coefficients μ , ϕ and σ_ε^2 with $0 < \sigma_\varepsilon < \infty$.

- Unconditional distribution:

$$\mathbb{E}(y_t) = \mu / (1 - \phi), \quad \mathbb{V}\text{ar}(y_t) = \sigma_\varepsilon^2 / (1 - \phi^2)$$

- Conditional distribution:

$$\mathbb{E}(y_t | Y_{t-1}) = \mu + \phi y_{t-1}, \quad \mathbb{V}\text{ar}(y_t | Y_{t-1}) = \sigma_\varepsilon^2$$

where $Y_t = \{y_1, \dots, y_t\}$.

Moving Average model: MA(1)

The MA(1) model is given by

$$y_t = \mu + \theta \varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2),$$

with three parameter coefficients μ , θ and σ_ε^2 with $0 < \sigma_\varepsilon < \infty$.

Invertibility condition: $|\theta| < 1$.

Statistical dynamic properties:

- Mean $\mathbb{E}(y_t) = \mu$; in case $\mu = 0$, $\mathbb{E}(y_t) = 0$;
- Variance $\mathbb{V}\text{ar}(y_t) = \sigma_\varepsilon^2 (1 + \theta^2)$;
- Autocovariance lag 1 is $\mathbb{C}\text{ov}(y_t, y_{t-1}) = \theta \sigma_\varepsilon^2$;
- ... for lag $\tau = 2, 3, 4, \dots$ is $\mathbb{C}\text{ov}(y_t, y_{t-\tau}) = 0$;
- Autocorrelation lag 1 is $\mathbb{C}\text{orr}(y_t, y_{t-1}) = \theta / (1 + \theta^2)$.

Moving Average model: MA(1)

The MA(1) model is given by

$$y_t = \mu + \theta\varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2),$$

with three parameter coefficients μ , θ and σ_ε^2 with $0 < \sigma_\varepsilon < \infty$.

- Unconditional distribution:

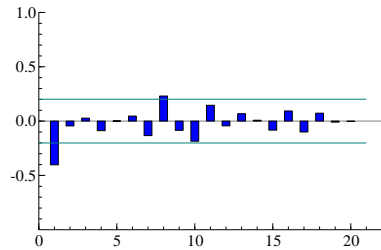
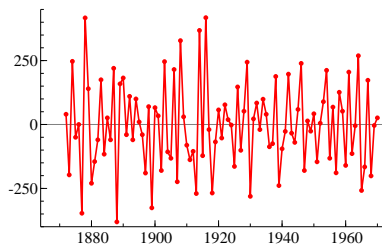
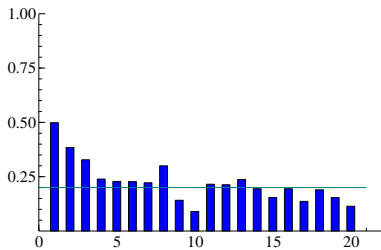
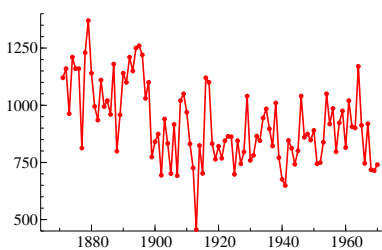
$$\mathbb{E}(y_t) = \mu, \quad \mathbb{V}\text{ar}(y_t) = \sigma_\varepsilon^2 (1 + \theta^2)$$

- Conditional distribution:

$$\mathbb{E}(y_t | Y_{t-1}) = \mu + \theta\varepsilon_{t-1}, \quad \mathbb{V}\text{ar}(y_t | Y_{t-1}) = \sigma_\varepsilon^2$$

where ε_{t-1} can be reconstructed from Y_{t-1} and with $\varepsilon_0 = 0$.

Example: Nile in levels and Nile in differences



Local Level:

model formulation and statistical properties

Constant vs Time Varying Mean Model

- Constant mean:
 - fixed level μ :

$$y_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2)$$

- Time Varying mean: replace μ by μ_t with
 - deterministic function of time:

$$\mu_t = a + b t + c t^2 + \dots$$

- stochastic function of time, for example:

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{NID}(0, \sigma_\eta^2)$$

Local Level Model

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2)$$
$$\mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim \mathcal{NID}(0, \sigma_\eta^2)$$

- Time-varying level is modelled as a random walk process;
- Notice the updating in terms of t or $t + 1$;
- The disturbances ε_t, η_s are independent for all s, t ;
- The model is incomplete without initial specification for μ_1 ;
- The process μ_t is nonstationary and y_t is nonstationary.

Local Level Model

The local level model or random walk plus noise model :

$$\begin{aligned}y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{NID}(0, \sigma_\varepsilon^2) \\ \mu_{t+1} &= \mu_t + \eta_t, & \eta_t &\sim \mathcal{NID}(0, \sigma_\eta^2)\end{aligned}$$

- The level μ_t and irregular ε_t are both unobserved;
- We still need to define μ_1 ;
- Parameters σ_ε^2 and σ_η^2 are unknown;
- Define q as the *signal-to-noise ratio* : $q = \sigma_\eta^2 / \sigma_\varepsilon^2$.

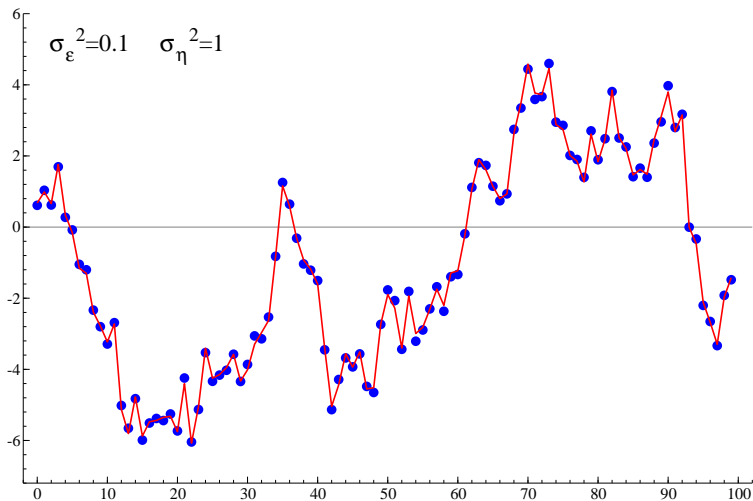
Local Level Model

The local level model or random walk plus noise model :

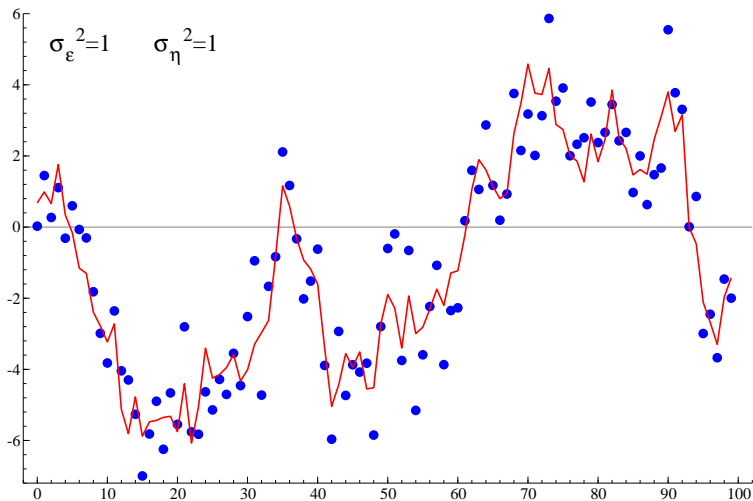
$$\begin{aligned}y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{NID}(0, \sigma_\varepsilon^2) \\ \mu_{t+1} &= \mu_t + \eta_t, & \eta_t &\sim \mathcal{NID}(0, \sigma_\eta^2)\end{aligned}$$

- Trivial special cases:
 - $\sigma_\eta^2 = 0 \implies y_t \sim \mathcal{NID}(\mu_1, \sigma_\varepsilon^2)$ (IID, global level);
 - $\sigma_\varepsilon^2 = 0 \implies y_{t+1} = y_t + \eta_t$ (random walk);
- Local Level model is basic illustration of **state space model**.
- It is very easy to simulate data from the local level model.

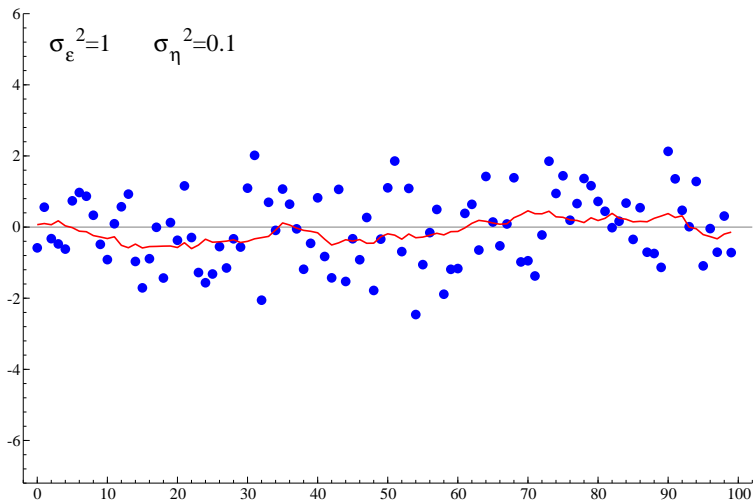
Simulated data, LL model, $\sigma_{\varepsilon}^2 = 0.1$, $\sigma_{\eta}^2 = 1$, $q = 10$



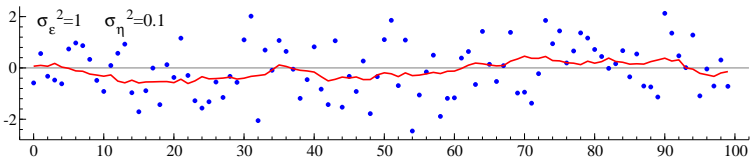
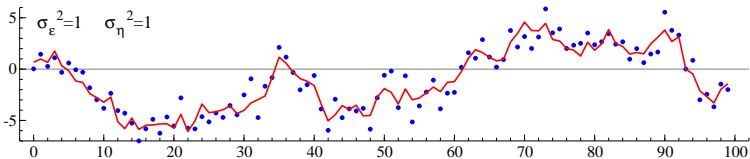
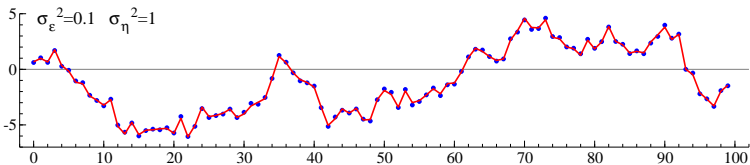
Simulated data, LL model, $\sigma_{\varepsilon}^2 = 1$, $\sigma_{\eta}^2 = 1$, $q = 1$



Simulated data, LL model, $\sigma_{\varepsilon}^2 = 1$, $\sigma_{\eta}^2 = 0.1$, $q = 0.1$



Simulated data, LL model



Statistical Properties

- Local Level Model, non-stationary y_t :

$$\begin{aligned}y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{NID}(0, \sigma_\varepsilon^2), \\ \mu_{t+1} &= \mu_t + \eta_t, & \eta_t &\sim \mathcal{NID}(0, \sigma_\eta^2).\end{aligned}$$

- First difference of y_t , stationary Δy_t :

$$\Delta y_t = \Delta \mu_t + \Delta \varepsilon_t = \eta_{t-1} + \varepsilon_t - \varepsilon_{t-1}.$$

- Dynamic properties of Δy_t : $\mathbb{E}(\Delta y_t) = 0$ and

$$\begin{aligned}\gamma_0 &= \mathbb{E}(\Delta y_t \Delta y_t) &= \sigma_\eta^2 + 2\sigma_\varepsilon^2, \\ \gamma_1 &= \mathbb{E}(\Delta y_t \Delta y_{t-1}) &= -\sigma_\varepsilon^2, \\ \gamma_\tau &= \mathbb{E}(\Delta y_t \Delta y_{t-\tau}) &= 0, & \text{for } \tau \geq 2.\end{aligned}$$

Properties of Local Level model

- Define theoretical autocorrelation function (acf)

$$\rho_\tau = \gamma_\tau / \gamma_0, \quad \tau = 1, 2, \dots$$

- We have q as the *signal-to-noise ratio* : $q = \sigma_\eta^2 / \sigma_\varepsilon^2$.
- We have $\gamma_0 = \sigma_\eta^2 + 2\sigma_\varepsilon^2$, $\gamma_1 = -\sigma_\varepsilon^2$, $\gamma_\tau = 0$ for $\tau \geq 2$.
- The theoretical acf of Δy_t in local level model is

$$\rho_1 = \frac{-\sigma_\varepsilon^2}{\sigma_\eta^2 + 2\sigma_\varepsilon^2} = -\frac{1}{q + 2},$$
$$\rho_\tau = 0, \quad \tau \geq 2.$$

- This acf is the same as acf of AR(1) or MA(1) ?
- We notice that

$$-1/2 \leq \rho_1 \leq 0$$

Properties of Local Level model (ctd)

- The local level model implies that $\Delta y_t \sim \text{MA}(1)$ but with the acf function $\rho_1 = -1/(q + 2)$, it is restricted;
- Hence $y_t \sim \text{ARIMA}(0, 1, 1)$.
- An alternative representation for Δy_t is the MA(1) model

$$\Delta y_t = \xi_t + \theta \xi_{t-1}, \quad \xi_t \sim \mathcal{NID}(0, \sigma^2).$$

- The acf of an MA(1) process is $\rho_1 = \theta / (1 + \theta^2)$.
- When y_t comes from a local level model, we have a restricted parameter space for θ : $-1 < \theta < 0$.
- To express θ as function of q , solve equality for the two ρ_1 's:

$$\theta = \frac{1}{2} \left(\sqrt{q^2 + 4q} - 2 - q \right).$$

Local Level Model

The Local Level model is given by

$$y_t = \mu_t + \varepsilon_t, \quad \mu_{t+1} = \mu_t + \eta_t, \quad t = 1, \dots, n.$$

- The parameters σ_ε^2 and σ_η^2 are unknown and need to be estimated, typically via maximum likelihood estimation;
- Parameter estimation via maximum likelihood will be discussed soon.
- When we treat parameters σ_ε^2 and σ_η^2 as known, how to "estimate" the unobserved series μ_1, \dots, μ_n ?
- This "estimation" is referred to as **signal extraction**.
- We base this "estimation" on **conditional expectations**.
- Signal extraction is the recursive evaluation of conditional means and variances of the unobserved μ_t for $t = 1, \dots, n$.

What are the take-aways so far ?

- Statistical formulation of local level model
- Role of signal-to-noise ratio
- Statistical (dynamic) properties and relation to ARIMA model

Signal extraction and prediction

some basics on the bivariate normal distribution

Normal density

Consider a random variable x that is normally distributed

$$x \sim \mathcal{N}(\mu_x, \sigma_x^2).$$

The density function for x is given by

$$f(x) = c \exp\left(-\frac{1}{2} Q_x\right),$$

where

$$c^{-1} = \sigma_x \sqrt{2\pi}, \quad Q_x = (x - \mu_x)^2 / \sigma_x^2.$$

The logdensity function for x is given by

$$\log f(x) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma_x^2 - \frac{1}{2} Q_x.$$

Bivariate normal distribution

Consider two random variable x and y that are normally distributed

$$x \sim \mathcal{N}(\mu_x, \sigma_x^2), \quad y \sim \mathcal{N}(\mu_y, \sigma_y^2), \quad \text{Cov}(x, y) = \sigma_{xy},$$

where we let $\sigma_{xy} = \sigma_x \sigma_y \rho$ with $-1 \leq \rho \leq 1$.

In case $\rho = 0$ and hence $\sigma_{xy} = 0$, the variables are *independent* or *uncorrelated*. The joint normal density function is simply

$$f(x, y) = f(x) f(y).$$

Also,

$$\mathbb{E}(x|y) = \mathbb{E}(x) = \mu_x, \quad \text{Var}(x|y) = \text{Var}(x) = \sigma_x^2.$$

Bivariate normal distribution, ctd.

Consider two random variable x and y that are normally distributed

$$x \sim \mathcal{N}(\mu_x, \sigma_x^2), \quad y \sim \mathcal{N}(\mu_y, \sigma_y^2), \quad \text{Cov}(x, y) = \sigma_{xy},$$

where $\sigma_{xy} = \sigma_x \sigma_y \rho$ with $-1 \leq \rho \leq 1$.

In case $-1 \leq \rho \leq 1$ and hence $\sigma_{xy} \neq 0$, the variables are *dependent* and hence *correlated*. The joint density function is normal and (to avoid matrix algebra for now) it can be expressed by

$$f(x, y) = f(x|y) f(y).$$

What are the expressions for $\mathbb{E}(x|y)$ and $\mathbb{V}\text{ar}(x|y)$?

Conditional mean and variance

Consider two random variable x and y that are normally distributed

$$x \sim \mathcal{N}(\mu_x, \sigma_x^2), \quad y \sim \mathcal{N}(\mu_y, \sigma_y^2), \quad \text{Cov}(x, y) = \sigma_{xy} \neq 0,$$

where $\sigma_{xy} = \sigma_x \sigma_y \rho$. To obtain expressions for the conditional mean $\mathbb{E}(x|y)$ and variance $\mathbb{V}\text{ar}(x|y)$, we define

$$y = \mu_y + \sigma_y z_y, \quad x = \mu_x + \sigma_x [\rho z_y + \sqrt{1 - \rho^2} z_x],$$

where $z_x, z_y \sim \mathcal{N}(0, 1)$ are independently distributed such that

$$f(z_x, z_y) = \frac{1}{2\pi} \exp \left[-\frac{1}{2} (z_x^2 + z_y^2) \right].$$

Please verify that statistical properties of x and y (mean, var, cov) are the same when based on the two expressions above.

Conditional mean, derivation for bivariate \mathcal{N}

Given that

$$y = \mu_y + \sigma_y z_y, \quad x = \mu_x + \sigma_x [\rho z_y + \sqrt{1 - \rho^2} z_x],$$

we have

$$\begin{aligned} \mathbb{E}(x|y) &= \mathbb{E} \left[\mu_x + \sigma_x \left(\rho z_y + \sqrt{1 - \rho^2} z_x \right) \mid y \right] \\ &= \mu_x + \sigma_x \mathbb{E} \left[\left(\rho \frac{y - \mu_y}{\sigma_y} + \sqrt{1 - \rho^2} z_x \right) \mid y \right] \\ &= \mu_x + \sigma_x \sigma_y \rho \frac{y - \mu_y}{\sigma_y^2} + \sqrt{1 - \rho^2} \mathbb{E}(z_x | z_y) \\ &= \mu_x + \sigma_{xy} (y - \mu_y) / \sigma_y^2. \end{aligned}$$

Conditional variance, derivation for bivariate \mathcal{N}

Given that

$$y = \mu_y + \sigma_y z_y, \quad x = \mu_x + \sigma_x [\rho z_y + \sqrt{1 - \rho^2} z_x],$$

we have

$$\begin{aligned} \text{Var}(x|y) &= \text{Var} \left[\mu_x + \sigma_x \left(\rho z_y + \sqrt{1 - \rho^2} z_x \right) | y \right] \\ &= \text{Var} \left[\mu_x + \sigma_x \left(\rho \frac{y - \mu_y}{\sigma_y} + \sqrt{1 - \rho^2} z_x \right) | y \right] \\ &= \text{Var} \left[\sigma_x \sqrt{1 - \rho^2} z_x | z_y \right] \\ &= \sigma_x^2 (1 - \rho^2) \\ &= \sigma_x^2 - \sigma_{xy}^2 / \sigma_y^2. \end{aligned}$$

Conditional mean and variance of bivariate \mathcal{N}

Consider two random variable x and y that are normally distributed

$$x \sim \mathcal{N}(\mu_x, \sigma_x^2), \quad y \sim \mathcal{N}(\mu_y, \sigma_y^2), \quad \text{Cov}(x, y) = \sigma_{xy} \neq 0.$$

The conditional mean and variance are given by

$$\mathbb{E}(x|y) = \mu_x + \sigma_{xy}(y - \mu_y) / \sigma_y^2, \quad \mathbb{V}\text{ar}(x|y) = \sigma_x^2 - \sigma_{xy}^2 / \sigma_y^2.$$

Verify these results and get familiar with these basic principles.

Notice that (1) $\mathbb{E}(x|y)$ is a function of y but $\mathbb{V}\text{ar}(x|y)$ is not;
(2) when $\sigma_{xy} = 0$, $\mathbb{E}(x|y) = \mu_x$ and $\mathbb{V}\text{ar}(x|y) = \sigma_x^2$.

Estimation Error / Prediction Error

Consider two random variable x and y that are normally distributed

$$x \sim \mathcal{N}(\mu_x, \sigma_x^2), \quad y \sim \mathcal{N}(\mu_y, \sigma_y^2), \quad \text{Cov}(x, y) = \sigma_{xy} \neq 0.$$

When considering the conditional mean $\mathbb{E}(x|y)$ as an **estimate** of x , the estimation error is

$$e = x - \mathbb{E}(x|y) = (x - \mu_x) - \sigma_{xy}(y - \mu_y) / \sigma_y^2,$$

and its properties are

$$\mathbb{E}(e) = 0, \quad \text{Var}(e) = \sigma_x^2 - \sigma_{xy}^2 / \sigma_y^2 = \text{Var}(x|y),$$

and

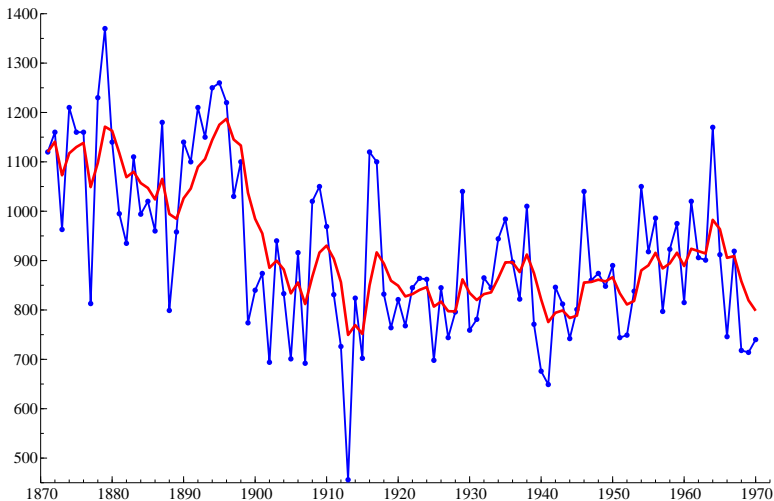
$$\text{Cov}(e, y) = \mathbb{E}[e(y - \mu_y)] = \sigma_{xy} - \sigma_{xy} = 0,$$

implying $\mathbb{E}(e|y) = \mathbb{E}(e) = 0$ and $\text{Var}(e|y) = \text{Var}(e) = \text{Var}(x|y)$.

Signal extraction and prediction

back to local level model

Signal Extraction for Nile Data: filtered estimate of level



Local Level Model: signal extraction

Local Level model :

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad \mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2).$$

We are at time point t .

Assume we have collected observations for y_1, \dots, y_{t-1} and that the conditional density $f(\mu_t | y_1, \dots, y_{t-1})$ is normal with known mean a_t and known variance p_t , we have

$$\mu_t | y_1, \dots, y_{t-1} \sim f(\mu_t | y_1, \dots, y_{t-1}) \equiv \mathcal{N}(a_t, p_t).$$

We collect observation for y_t , the conditional density of interest is

$$f(\mu_t | y_1, \dots, y_t),$$

This conditional density turns out to be normal as well

$$f(\mu_t | y_1, \dots, y_t) \equiv \mathcal{N}(a_{t|t}, p_{t|t}),$$

What are the expressions for $a_{t|t}$ and $p_{t|t}$? This is next !

Prediction error

Local Level model :

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad \mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2).$$

Given $f(\mu_t | y_1, \dots, y_{t-1}) \equiv \mathcal{N}(a_t, p_t)$, we construct a forecast for y_t , that is

$$\hat{y}_t = \mathbb{E}(y_t | Y_{t-1}) = \mathbb{E}(\mu_t + \varepsilon_t | Y_{t-1}) = a_t.$$

The corresponding **prediction error** is

$$v_t = y_t - \hat{y}_t = y_t - a_t,$$

with $\mathbb{E}(v_t) = \mathbb{E}(y_t - a_t) = \mathbb{E}[(\mu_t - a_t) + \varepsilon_t] = 0$ and

$$\text{Var}(v_t) = \text{Var}(y_t - a_t) = \text{Var}[(\mu_t - a_t) + \varepsilon_t] = p_t + \sigma_\varepsilon^2,$$

see results in Slide “Estimation Error / Prediction Error”.

Prediction error

We have prediction error

$$v_t = y_t - a_t, \quad \text{with} \quad \mathbb{E}(v_t) = 0, \quad \mathbb{V}\text{ar}(v_t) = p_t + \sigma_\varepsilon^2.$$

It also follows from Slide “Estimation Error / Prediction Error” that

$$\mathbb{C}\text{ov}(v_t, y_{t-j}) = 0, \quad \text{for } j = 1, \dots, t-1.$$

Hence, we also have

$$\begin{aligned} \mathbb{E}(v_t | Y_{t-1}) &= \mathbb{E}(v_t) = 0, \\ \mathbb{V}\text{ar}(v_t | Y_{t-1}) &= \mathbb{V}\text{ar}(v_t) = p_t + \sigma_\varepsilon^2, \end{aligned}$$

and

$$f(y_t | Y_{t-1}) = \mathcal{N}(a_t, p_t + \sigma_\varepsilon^2), \quad f(v_t) = f(v_t | Y_{t-1}) = \mathcal{N}(0, p_t + \sigma_\varepsilon^2).$$

Local Level Model: signal extraction

To obtain an expression for $a_{t|t}$ and $p_{t|t}$ in $f(\mu_t|Y_t) = \mathcal{N}(a_{t|t}, p_{t|t})$, we next make the point that we can “*re-shuffle*” linearly the conditional information set: we have $f(a|b, c) \equiv f(a|b, d)$, when, for example, $d = c - b$. Hence we can also have

$$f(\mu_t|Y_t) \equiv f(\mu_t|v_t, Y_{t-1})$$

since $v_t = y_t - \mathbb{E}(\mu_t|Y_{t-1})$ is a fixed linear function of Y_t , more on this below. Then we have

$$\begin{aligned} f(\mu_t|v_t, Y_{t-1}) &= f(\mu_t, v_t|Y_{t-1})/f(v_t|Y_{t-1}) \\ &= f(\mu_t|Y_{t-1})f(v_t|\mu_t, Y_{t-1})/f(v_t|Y_{t-1}), \end{aligned}$$

where all $f()$'s are normal densities.

Local Level Model: signal extraction

We have

$$f(\mu_t | v_t, Y_{t-1}) = f(\mu_t | Y_{t-1})f(v_t | \mu_t, Y_{t-1})/f(v_t | Y_{t-1}),$$

with

$$f(\mu_t | Y_{t-1}) = \mathcal{N}(a_t, p_t), \quad f(v_t | Y_{t-1}) = \mathcal{N}(0, p_t + \sigma_\varepsilon^2),$$

but what about $f(v_t | \mu_t, Y_{t-1})$?

Given that $v_t = y_t - a_t = \mu_t + \varepsilon_t - a_t$, we have

$$\mathbb{E}(v_t | \mu_t, Y_{t-1}) = \mu_t - a_t, \quad \mathbb{V}\text{ar}(v_t | \mu_t, Y_{t-1}) = \sigma_\varepsilon^2.$$

and hence

$$f(v_t | \mu_t, Y_{t-1}) = \mathcal{N}(\mu_t - a_t, \sigma_\varepsilon^2).$$

Local Level Model: signal extraction

It follows that

$$\begin{aligned} f(\mu_t | v_t, Y_{t-1}) &= f(\mu_t | Y_{t-1}) \times f(v_t | \mu_t, Y_{t-1}) / f(v_t | Y_{t-1}) \\ &= \mathcal{N}(a_t, p_t) \times \mathcal{N}(\mu_t - a_t, \sigma_\varepsilon^2) / \mathcal{N}(0, p_t + \sigma_\varepsilon^2). \end{aligned}$$

Given the functional form of the normal density, we have

$f(\mu_t | Y_t) = \text{const.} \times \exp\left(-\frac{1}{2} Q_t\right)$ with

$$Q_t = (\mu_t - a_t)^2 / p_t + (v_t - \mu_t + a_t)^2 / \sigma_\varepsilon^2 - v_t^2 / (p_t + \sigma_\varepsilon^2).$$

After some algebra ("completing the square"), we have

$$Q_t = \frac{p_t + \sigma_\varepsilon^2}{p_t \sigma_\varepsilon^2} \left(\mu_t - a_t - \frac{p_t v_t}{p_t + \sigma_\varepsilon^2} \right)^2.$$

Local Level Model: filter density

The **filter** density function $f(\mu_t|Y_t)$ is normal and has functional form,

$$f(\mu_t|Y_t) = \text{const.} \times \exp\left(-\frac{1}{2}Q_t\right),$$

with

$$Q_t = \frac{p_t + \sigma_\varepsilon^2}{p_t \sigma_\varepsilon^2} \left(\mu_t - a_t - \frac{p_t v_t}{p_t + \sigma_\varepsilon^2}\right)^2.$$

It implies that

$$f(\mu_t|Y_t) \equiv \mathcal{N}(a_{t|t}, p_{t|t}),$$

with

$$a_{t|t} = a_t + k_t v_t, \quad p_{t|t} = k_t \sigma_\varepsilon^2, \quad k_t = \frac{p_t}{p_t + \sigma_\varepsilon^2}.$$

Presto ! We have expressions for $a_{t|t}$ and $p_{t|t}$, congratulations !!

Local Level Model: prediction

Local Level model :

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad \mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2).$$

In addition, we are typically interested in the *predicted* signal density

$$f(\mu_{t+1} | Y_t) \equiv \mathcal{N}(a_{t+1}, p_{t+1}),$$

where

$$\begin{aligned} a_{t+1} &= \mathbb{E}(\mu_{t+1} | Y_t) = \mathbb{E}(\mu_t + \eta_t | Y_t) = a_{t|t}, \\ p_{t+1} &= \text{Var}(\mu_t + \eta_t | Y_t) = p_{t|t} + \sigma_\eta^2. \end{aligned}$$

We have obtained the updating equations

$$a_{t+1} = a_t + k_t v_t, \quad p_{t+1} = k_t \sigma_\varepsilon^2 + \sigma_\eta^2, \quad k_t = \frac{p_t}{p_t + \sigma_\varepsilon^2}.$$

These are **recursive equations**.

Kalman filter for the Local Level Model

Local Level model :

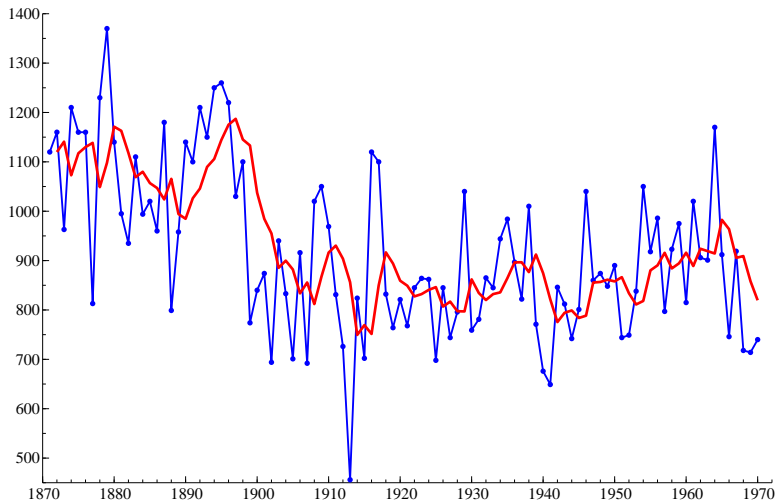
$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad \mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2).$$

For $f(\mu_t | Y_{t-1}) = \mathcal{N}(a_t, p_t)$, with given values of a_t and p_t , the Kalman filter update equation are given by

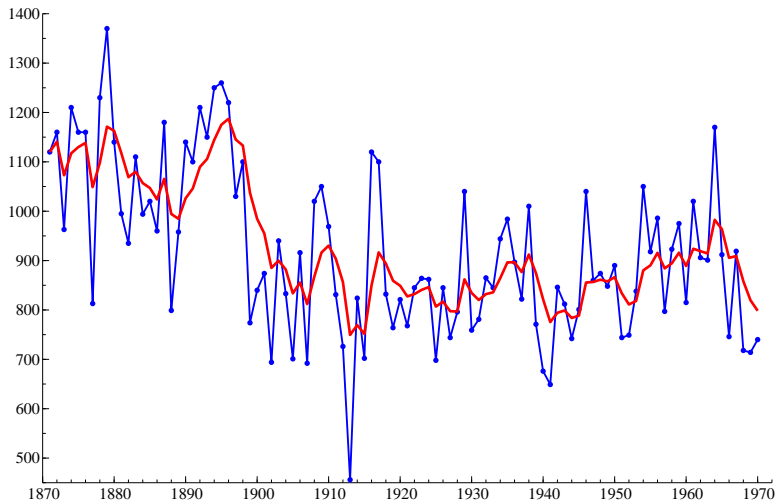
$$\begin{aligned} v_t &= y_t - a_t, & k_t &= p_t / (p_t + \sigma_\varepsilon^2), \\ a_{t|t} &= a_t + k_t v_t, & p_{t|t} &= k_t \sigma_\varepsilon^2, \\ a_{t+1} &= a_{t|t}, & p_{t+1} &= p_{t|t} + \sigma_\eta^2. \end{aligned}$$

We repeat this for each t , starting with $t = 1$: we let $t = 1, \dots, n$.
What initial values for a_1 and p_1 should we consider ?
Later !!

Signal Extraction: predicted estimates of local level



Signal Extraction: filtered estimates of local level



What are the take-aways here ?

- We derived a way to estimate μ_t from the Local Level model.
- The estimate a_t is the mean of the predicted conditional density $f(\mu_t | Y_{t-1}) = \mathcal{N}(a_t, p_t)$.
- The estimate $a_{t|t}$ is the mean of the filtered conditional density $f(\mu_t | Y_t) = \mathcal{N}(a_{t|t}, p_{t|t})$.
- The estimates are computed using recursive equations.
- The derivations rely on basic principles from the bivariate normal distribution.

Textbooks

- A.C.Harvey (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press
- G.Kitagawa & W.Gersch (1996). *Smoothness Priors Analysis of Time Series*. Springer-Verlag
- J.Harrison & M.West (1997). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag
- J.Durbin & S.J.Koopman (2012). *Time Series Analysis by State Space Methods, Second Edition*. Oxford University Press
- J.J.F.Commandeur & S.J.Koopman (2007). *An Introduction to State Space Time Series Analysis*. Oxford University Press

APPENDICES

Appendix – Taylor series

The Taylor expansion for function $f(x)$ around some value x^* is

$$f(x) = f(x = x^*) + f'(x = x^*)[x - x^*] + \frac{1}{2}f''(x = x^*)[x - x^*]^2 + \dots,$$

where

$$f'(x) = \frac{\partial f(x)}{\partial x}, \quad f''(x) = \frac{\partial^2 f(x)}{\partial x \partial x},$$

and $g(x = x^*)$ means that we evaluate function $g(x)$ at $x = x^*$.

Example: consider $f(x) = \log(1 + x)$ with $f'(x) = (1 + x)^{-1}$ and $f''(x) = -(1 + x)^{-2}$; the expansion of $f(x)$ around $x^* = 0$ is

$$\log(1 + x) = 0 + 1 \cdot (x - 0) + \frac{1}{2}(-1) \cdot (x - 0)^2 + \dots = x - \frac{1}{2}x^2 + \dots$$

Notice that $f(x = 0) = 0$, $f'(x = 0) = 1$ and $f''(x = 0) = -1$. For small enough x (when x is close to $x^* = 0$), we have

$$\log(1 + x) \approx x.$$

Check: $\log(1.01) = .00995 \approx 0.01$ and $\log(1.1) = 0.0953 \approx 0.1$.

Appendix – Percentage growth

Observation at time t is y_t and observation at time $t - 1$ is y_{t-1} .

We define rate r_t as the **proportional change** of y_t wrt y_{t-1} , that is

$$r_t = \frac{y_t - y_{t-1}}{y_{t-1}} \Rightarrow y_t - y_{t-1} = y_{t-1} \cdot r_t \Rightarrow y_t = y_{t-1} \cdot (1 + r_t).$$

We notice that r_t can be positive and negative !

When we take logs of $y_t = y_{t-1} \cdot (1 + r_t)$, we obtain

$$\log y_t = \log y_{t-1} + \log(1 + r_t) \Rightarrow \log y_t - \log y_{t-1} = \log(1 + r_t) \Rightarrow$$

$$\Delta \log y_t = \log(1 + r_t).$$

Since $\log(1 + r_t) \approx r_t$, see previous slide, when r_t is small, we have

$$r_t \approx \Delta \log y_t.$$

The **percentage growth** is defined as $100 \times r_t \approx 100 \cdot \Delta \log y_t$.

Appendix – Lag operators and polynomials

- Lag operator $Ly_t = y_{t-1}$, $L^\tau y_t = y_{t-\tau}$, for $\tau = 1, 2, 3, \dots$
- Difference operator $\Delta y_t = (1 - L)y_t = y_t - y_{t-1}$
- Autoregressive polynomial $\phi(L)y_t = (1 - \phi L)y_t = y_t - \phi y_{t-1}$
- Other polynomial $\theta(L)\varepsilon_t = (1 + \theta L)\varepsilon_t = \varepsilon_t + \theta\varepsilon_{t-1}$
- Second difference
$$\Delta^2 y_t = \Delta(\Delta y_t) = \Delta(y_t - y_{t-1}) = y_t - 2y_{t-1} + y_{t-2}$$
- Seasonal difference $\Delta_s y_t = y_t - y_{t-s}$ for typical
 $s = 2, 4, 7, 12, 52$
- Seasonal sum operator
$$S(L)y_t = (1 + L + L^2 + \dots + L^{s-1})y_t = y_t + y_{t-1} + \dots + y_{t-s+1}$$
- Show that $\Delta S(L) = \Delta_s$.