Introduction to Time Series Analysis

M2 Data Science & Artificial Intelligence

Juliette Chevallier

January 14, 2021

Introduction

Statistical Forecasting

Time series: Sequence of observations of a phenomenon over time

Continuous or discrete regular time

Example: Cryptocurrencies, electricity consumption, oil prices, French population, heart rate, seismograph readings, Internet traffic, cell phone sales, flood heights of the Nile, ocean temperature, carbon dioxide concentration in the atmosphere, blood glucose levels, the President's popularity rating, *etc*.

Idea: Signal *vs.* noise

Example: Cryptocurrencies, electricity consumption, oil prices, French population, heart rate, seismograph readings, Internet traffic, cell phone sales, flood heights of the Nile, ocean temperature, carbon dioxide concentration in the atmosphere, blood glucose levels, the President's popularity rating, *etc*.

Idea: Signal *vs.* noise

Example: Cryptocurrencies, electricity consumption, oil prices, French population, heart rate, seismograph readings, Internet traffic, cell phone sales, flood heights of the Nile, ocean temperature, carbon dioxide concentration in the atmosphere, blood glucose levels, the President's popularity rating, *etc*.

Idea: Signal *vs.* noise

"Prediction is very difficult, especially if it's about the future."
Nils Bohr, Nobel laureate in Physics

Risks of forecasting:

- Intrinsic risk: random variation, beyond explanation;
- Parameter risk: errors in estimating the parameters;
- Model risk: choosing the wrong model.

Example: Cryptocurrencies, electricity consumption, oil prices, French population, heart rate, seismograph readings, Internet traffic, cell phone sales, flood heights of the Nile, ocean temperature, carbon dioxide concentration in the atmosphere, blood glucose levels, the President's popularity rating, *etc*.

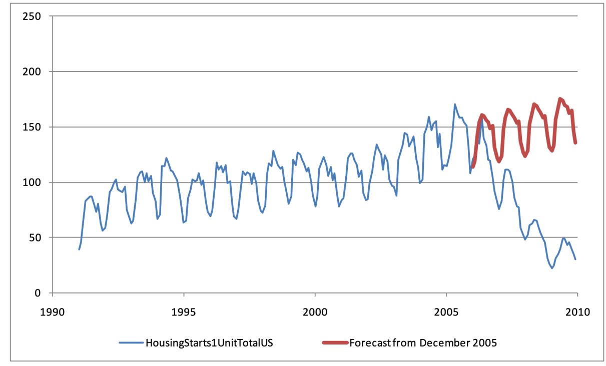
Idea: Signal *vs.* noise

"Prediction is very difficult, especially if it's about the future."
Nils Bohr, Nobel laureate in Physics

Risks of forecasting:

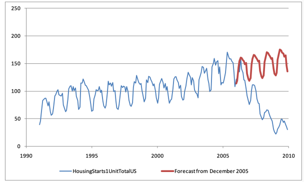
- Intrinsic risk: random variation, beyond explanation;
- Parameter risk: errors in estimating the parameters;
- Model risk: choosing the wrong model.

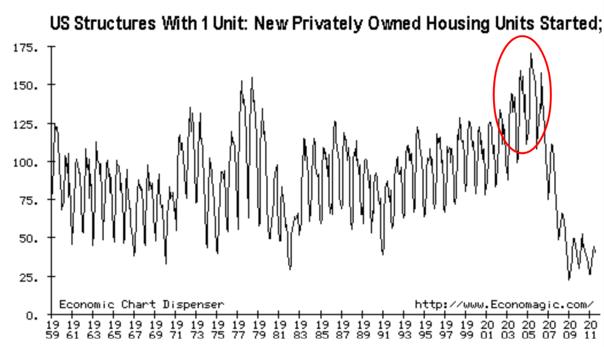
Risks of Forecasting



Example: US housing after 2005

Risks of Forecasting

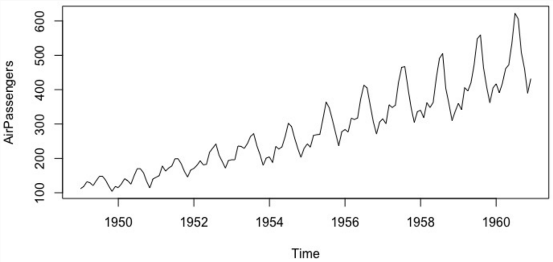




Example: US housing after 2005

Introduction

Descriptive Analysis

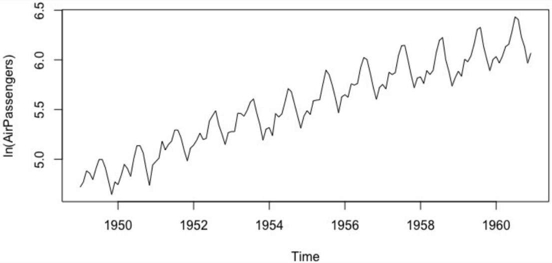


Monthly number of airline passengers (in thousands)

A **non-stationary** series:

- Trend,
- Seasonality,
- Variance

Descriptive Analysis

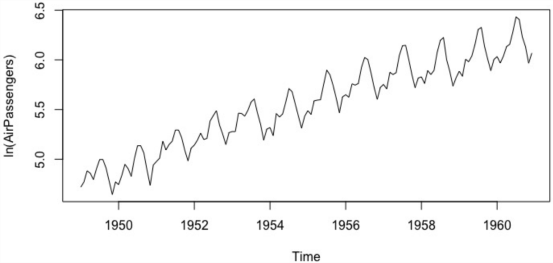


Monthly number of airline passengers (in thousands)

A **non-stationary** series:

- Trend m_t,
- Seasonality s_t,
- Variance

Descriptive Analysis



Monthly number of airline passengers (in thousands)

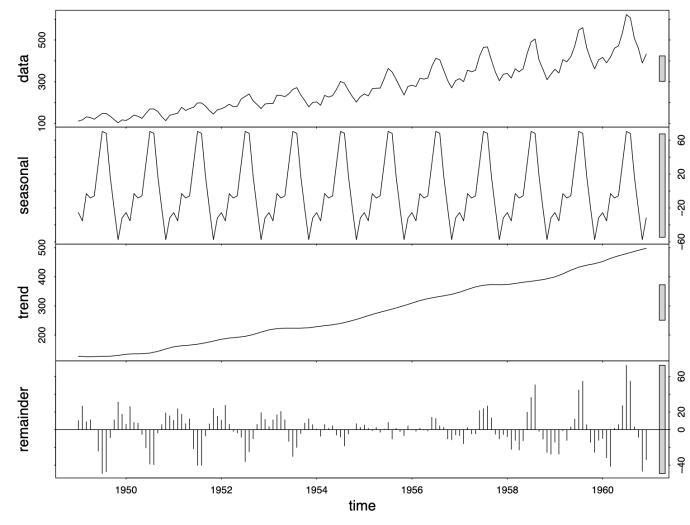
A **non-stationary** series:

- Trend m_t,
- Seasonality s_t,
- Variance

We observe $(y_t)_{t \in T}$ a trajectory of a stochastic process

$$Y_t = m_t + s_t + Z_t \quad , \quad \text{where } t \in T \subset \mathbb{Z} \text{ or } \mathbb{N}$$

and $(Z_t)_{t \in T}$ is a random noise that one hopes is stationary



Monthly number of airline passengers (in thousands)

Trend and Seasonality

Trend m_t

$$Y_t = m_t + s_t + Z_t \quad , \quad \text{where } t \in T \subset \mathbb{Z} \text{ or } \mathbb{N}$$

- **Expectation**: Slow changes that capture long-term variations;
- **Some examples**: Polynomial trend: $m_t = a_0 + a_1 t + \ldots + a_d t^d$,
  Exponential trend: $m_t = a_0 + a_1 \alpha^t$,
  Logistic trend: $m_t = \frac{1}{a_0 + a_1 t}$
- **Detrending**: Remove the trend component from the time series
  → Trend estimation, average, moving average, exponential smoothing

$$\widehat{y_t} = \frac{1}{2\ell + 1} \sum_{i=t-\ell}^{t+\ell} y_i$$

In case of parametric representation of the trend → Regression

Polynomial trend → Linear regression, *i.e.* least square estimation

$$(\widehat{a_0}, \ldots, \widehat{a_d}) = \underset{(a_0,\ldots,a_d)\in\mathbb{R}^d}{\operatorname{argmin}} \sum_{t=1}^{n} (y_t - m_t)^2 \quad , \quad \text{where } m_t = a_0 + a_1 t + \ldots + a_d t^d \, .$$

Exercise: $\boxed{\widehat{a} = ({}^t A A)^{-1} ({}^t A Y)}$

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad a = \begin{pmatrix} a_0 \\ \vdots \\ a_d \end{pmatrix} \qquad A = \begin{pmatrix} 1 & t_1 & t_1^2 & \ldots & t_1^d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & \ldots & t_n^d \end{pmatrix}$$

Trend Estimation: Non-Parametric Estimation

$$m_t = f(t)\,, \text{ where } f \text{ regular}$$

Several approaches are possible $\rightsquigarrow$ kernel and local polynomials estimators.

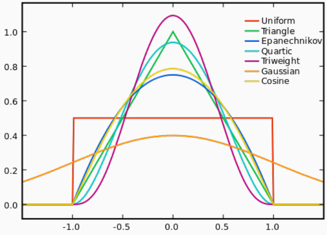
$$m_t = f(t)\,, \text{ where } f \text{ regular}$$

Several approaches are possible $\rightsquigarrow$ kernel and local polynomials estimators.

Kernel: Function $K\colon \mathbb{R}^d \to \mathbb{R}$ such that $\int K^2 < \infty$ and $\int K = 1$.

Kernel estimator associated with window $h \in \mathbb{R}^+$ and kernel K

$$\widehat{f}_h(x) = \frac{\sum_{t=1}^{n} y_t K(\frac{x-t}{h})}{\sum_{t=1}^{n} K(\frac{x-t}{h})}$$



Examples: Gaussian, uniform, triangle, logistic, Epanechnikov, *etc.*

Trend Estimation: Non-Parametric Estimation

$$m_t = f(t)\,,\ \text{where } f \text{ regular}$$

Several approaches are possible $\rightsquigarrow$ kernel and local polynomials estimators.

local polynomial estimator of degree q associated with window h and kernel K

$$\widehat{f}_h(x) = \underset{P}{\operatorname{argmin}} \sum_{t=1}^{n} W_t(x) \|y_t - P(x_t - x)\|^2$$

where $W_t(x) = \dfrac{K(\frac{x-t}{h})}{\sum_{t=1}^n K(\frac{x-t}{h})}$ and $P(x) = \displaystyle\sum_{j=0}^{q} a_j x^j$.

Another techniques: projection on adapted function bases, *etc.*

Trend Estimation by Exponential Smoothing

Exponential smoothing of parameter $\alpha \in [0, 1]$

$$\hat{m}_t = \alpha y_t + (1 - \alpha)\hat{m}_{t-1} \quad \text{and} \quad \hat{m}_1 = y_1$$

Exercise:

1. Show that exponential smoothing is a moving average, specify its nature and its coefficients.
2. What can be said about the evolution of weights according to the past considered? What happens when α is close to 1, close to 0?

Trend Estimation by Exponential Smoothing

Exponential smoothing of parameter $\alpha \in [0, 1]$

$$\hat{m}_t = \alpha y_t + (1 - \alpha)\hat{m}_{t-1} \quad \text{and} \quad \hat{m}_1 = y_1$$

Exercise:

1. Show that exponential smoothing is a moving average, specify its nature and its coefficients.
2. What can be said about the evolution of weights according to the past considered? What happens when α is close to 1, close to 0?

Exponential smoothing of Holt-Winters of parameter $\alpha \in [0, 1]$ and $\beta \in [0, 1]$

$$\hat{y}_{t+h} = \ell_t + h b_t$$

Level: $\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1} + b_{t-1}$
Trend: $b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$

Trend and Seasonality

Seasonality s_t

$$Y_t = m_t + s_t + Z_t \quad , \quad \text{where } t \in T \subset \mathbb{Z} \text{ or } \mathbb{N}$$

- **Expectation**: Periodic deterministic function of period r such that

$$\forall t \in T, \quad \sum_{i=1}^{r} s_{t+i} = 0\,;$$

- **Some examples**: combination of sinusoidal functions, Indicator functions;
- **Least square estimation**: $s_t = a_0 + \sum_{j=1}^{k} a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)$, where the a_j and b_j are unknown and λ_i and λ_j are known integer multiples of $\frac{2\pi}{d}$.

Trend and Seasonality

Delay and Difference Operators

Delay operator $(BY)_t = Y_{t-1}$;

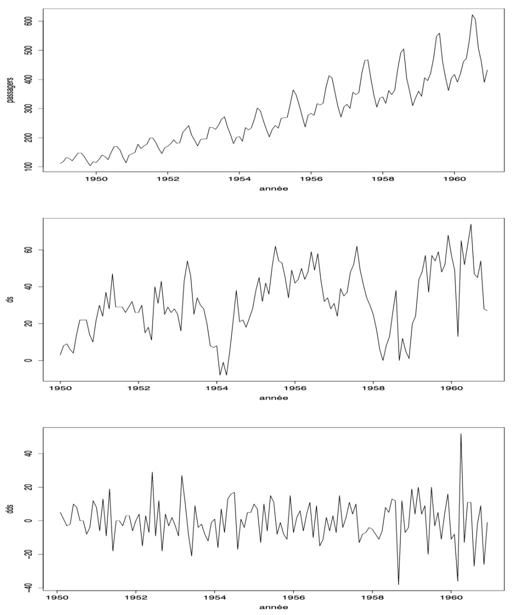
Difference operator: $(\Delta Y)_t = Y_t - Y_{t-1} = (1 - B)Y_t$

Seasonal difference operator: $(\Delta_d Y)_t = Y_t - Y_{t-d} = ((1 - B^d)Y)_t$

Difference operator of order n: $\Delta^n = (1 - B)^n$

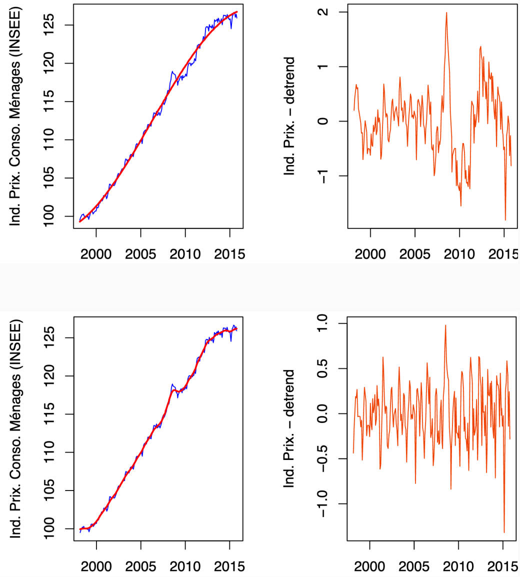
Proposition:

- n-order-difference operator eliminates polynomial trend of degree $< n$;
- Seasonal difference operator eliminates a seasonal component of period d.


passagers
année
ds
année
dds
année

Trend and Seasonality

Stationarize the Series



Test of Stationarity: Dickey Fuller, Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

Random time series modeling

Stochastic Process: Family $(X_t)_{t\in\mathbb{Z}}$ of random variables with values in $\mathbb{R}$

$$\Omega \times \mathbb{Z} \to \mathbb{R}$$
$$(\omega, t) \mapsto X_t(\omega)$$

- $\forall t \in \mathbb{Z}$, $X_t(\omega)$ is a random variable;
- $\forall \omega \in \Omega$, $t \mapsto X_t(\omega)$ is a trajectory of the process.

Example: Gaussian white noise is a sequence of independent and identically distributed variables (i.i.d.) according to a Gaullian law $\mathcal{N}(0, \sigma^2)$.

Second Order process: $(X_t)_{t\in\mathbb{Z}}$ is said of second order if $\forall t \; X_t \in \mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$.

For second order process:

- Mean $\mu_X : \mathbb{Z} \to \mathbb{R}$, $\mu_X(t) = \mathbb{E}[X_t]$
- Autocovariance $\gamma_X : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$, $\gamma_X(s,t) = Cov(X_s, X_t)$

Random time series modeling

Strongly stationary process: For all $h \in \mathbb{Z}$ and all sequence $(t_1, \ldots, t_n) \in \mathbb{Z}^n$, $(X_{t_1}, \ldots, X_{t_n})$ and $(X_{t_1+h}, \ldots, X_{t_n+h})$ have the same law.

Stationary process: A second order process is stationary if μ_X is constant and γ_X is invariant by translation.

$$\forall s, t, h \in \mathbb{Z}, \quad \mu_X(t+h) = \mu_X(t) \qquad \text{and} \qquad \gamma_X(s,t) = \gamma_X(s+h, t+h)$$

Exercise: What implication(s) exist between strong stationarity and stationarity?

Let $(X_t)_{t\in\mathbb{Z}}$ a stationary process.

Autocovariance function:

$$\gamma_X : \mathcal{Z} \to \mathbb{R}$$
$$h \mapsto \gamma_X(h) = \gamma_X(0,h) = Cov(X_t, X_t + h) \qquad (\forall t \in \mathbb{Z})$$

Autocorrelation function:

$$\rho_X : \mathcal{Z} \to [-1,1]$$
$$h \mapsto \rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \frac{Cov(X_t, X_t + h)}{\sqrt{Var(X_t)}\sqrt{Var(X_{t+h})}} \qquad (\forall t \in \mathbb{Z})$$

Random time series modeling

Auto Regressive Moving Average (ARMA) Models

ARMA: $(X_t)_{t\in\mathbb{Z}}$ admits an ARMA(p,q) representation if

$$\forall t \in \mathbb{Z}, \quad \Phi(B)X_t = \Theta(B)Z_t$$

where $(Z_t)_{t\in\mathbb{Z}}$ is a (Gaussian) white noise and

$$\Phi(B) = I - \phi_1 B - \phi_2 B^2 \ldots - \phi_p B^p \qquad \text{and} \qquad \Theta(B) = I + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$$

Theorem: If Φ has **no** module 1 root then ARMA(p,q) has a single stationary solution

⇝ Rational fraction of an ARMA, depending on whether it is *causal* and *reversible*

Moving Average (MA) Models

MA: $(Z_t)_{t\in\mathbb{Z}}$ admits an MA(q) representation if it is of the second order, stationary, and solution of the recurrence equation

$$\forall t \in \mathbb{Z}, \quad Z_t = \varepsilon_t + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} = \Theta(B)\varepsilon_t$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a (Gaussian) white noise and

$$\Theta(B) = I + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q \,.$$

- q is the order of the process and (θ, σ^2) its parameters
- Fully specified,
- *Several representations* but only one canonical representation.

Moving Average (MA) Models

MA: $(Z_t)_{t\in\mathbb{Z}}$ admits an MA(q) representation if it is of the second order, stationary, and solution of the recurrence equation

$$\forall t \in \mathbb{Z}, \quad Z_t = \varepsilon_t + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} = \Theta(B)\varepsilon_t$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a (Gaussian) white noise and

$$\Theta(B) = I + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q \,.$$

- q is the order of the process and (θ, σ^2) its parameters
- Fully specified,
- *Several representations* but only one canonical representation.

Exercise:

1. Show that $Var(Z_t) = \sigma^2(1 + \sum_{j=1}^{q} \theta_j^2)$
2. What about $\gamma_Z(h)$ for $h > q$?
3. Compute $\gamma_Z(1)$, $\gamma_Z(2)$ and derive a general expression from $\gamma_Z(h)$ for $h \leqslant q$

Auto Regressive (AR) Models

AR: $(Z_t)_{t\in\mathbb{Z}}$ admits an AR(p) representation if it is of the second order, stationary, and solution of the recurrence equation

$$\boxed{\forall t \in \mathbb{Z}, \quad Z_t = \varepsilon_t + \sum_{k=1}^{p} \phi_k Z_{t-k}} \qquad \textit{i.e.} \qquad \boxed{\forall t \in \mathbb{Z}, \quad \varepsilon_t = \Phi(B) Z_t}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a (Gaussian) white noise and

$$\Phi(B) = I - \phi_1 B - \phi_2 B^2 \ldots - \phi_p B^p \,.$$

Theorem:

- An *infinite number* of second order processes verifying the equation;
- If Φ has **no** module 1 root then AR(p) has a single stationary solution

Auto Regressive (AR) Models

AR: $(Z_t)_{t\in\mathbb{Z}}$ admits an AR(p) representation if it is of the second order, stationary, and solution of the recurrence equation

$$\forall t \in \mathbb{Z}, \quad Z_t = \varepsilon_t + \sum_{k=1}^{p} \phi_k Z_{t-k} \qquad \textit{i.e.} \qquad \forall t \in \mathbb{Z}, \quad \varepsilon_t = \Phi(B) Z_t$$

Exercise: Consider $(Z_t)_{t\in\mathbb{Z}}$ of canonical representation $\Phi(B)Z_t = \varepsilon_t$

1. Show that $Var(Z_t) = \sum_{i=1}^{p} \phi_i \gamma_Z(i) + Var(\varepsilon_t)$ and $\gamma_Z(0) = \dfrac{\sigma^2}{1 - \sum_{i=1}^{p} \phi_i \rho_Z(i)}$;
2. Show that, for all $h \in \mathbb{N}^*$, $Cov(Z_t, Z_{t+h}) = \sum_{i=1}^{p} \phi_i \gamma_Z(h-i) + Var(\varepsilon_t)$ and $\rho_Z(h) = \dfrac{\sigma^2}{1 - \sum_{i=1}^{p} \phi_i \gamma_Z(h-i)}$;
3. Check the exponential decay of autocorrelations on an AR(1) of canonical representation $Z_t = \phi Z_{t-1} + \varepsilon_t$

Take home message

In practice, here are the steps we can try to follow:

- Plot the time series and graphically look for a trend or a seasonal component;
- Model the trend and seasonal component. Differentiation can be used;
- Model the remainders.