Time Series Data Forecasting

Machine Learning Graz Meetup

Maximilian Toller

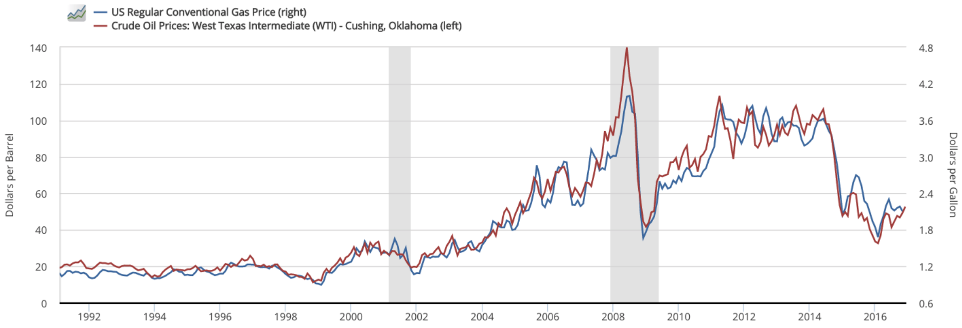
mtoller(at)know-center.at

2019-04-17

What are *time series*?

What are time series?

Data observed over time



What are time series?

Stochastic processes indexed by integers

- $\{X_t | t \in T\}$ $T = \mathbb{Z}$
- Confirmatory data analysis
- Goal: See if model is sound
- Mainly about: theorems, models, proofs
- Pros: Provably correct, theoretically sound
- Cons: "*All models are wrong*" - George Box

What are time series?

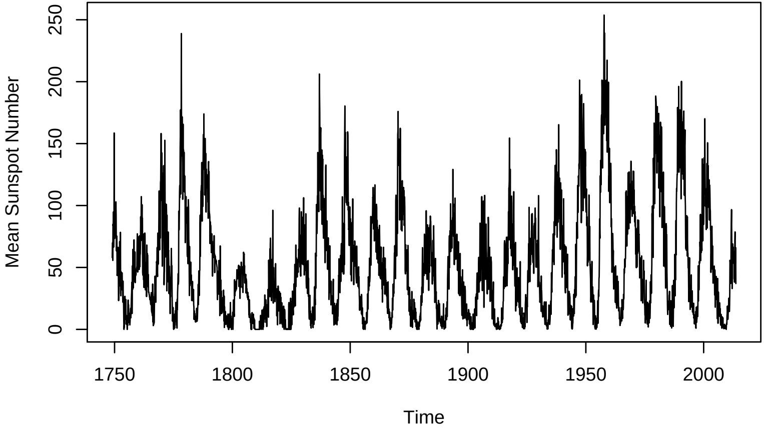
Data vs Process

- $x_t = \{114, 117, 104, \ldots\}$
- Exploratory data analysis
- Work with data
- Pros: fast, domain specific
- Cons: possibly unsound

- $\{X_t | t \in T\}$ $T = \mathbb{Z}$
- Confirmatory data analysis
- Work with models
- Pros: theoretically sound
- Cons: slow, simplification

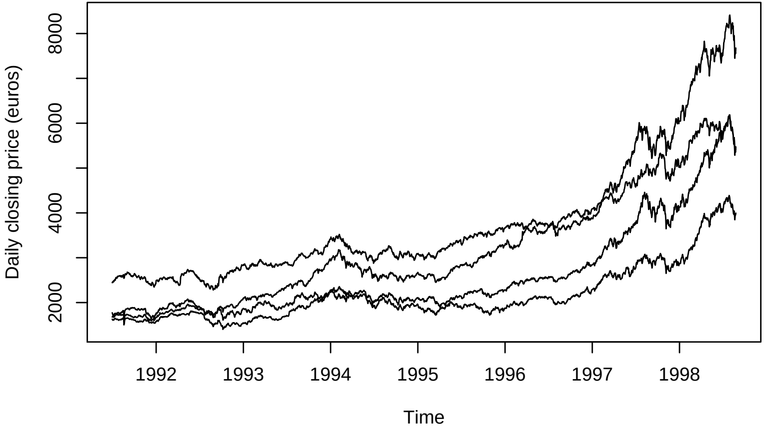
What are time series data?

Sunspot counts (monthly)



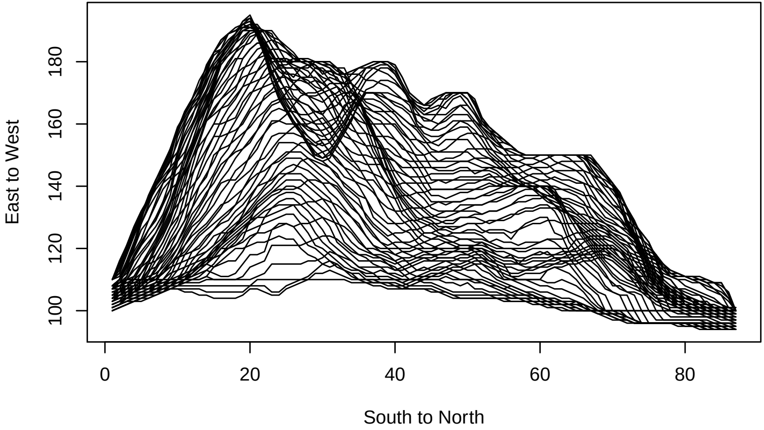
What are time series data?

EU stock market prices (daily)?



What are time series data?

Volcano topography?



Time series data vs other data

Some bad news, some good news

- $\downarrow$ High dimensionality
- $\downarrow$ Classical statistics don't work
- $\uparrow$ Temporal connection allows forecasting
- $\uparrow$ Solid theory

γ and ρ

Some math we'll need later

Some math we'll need later

Autocovariance

- $\mu_t = \mathbb{E}[X_t]$
- $\gamma(\tau, k) = \mathbb{E}[(X_\tau - \mu_\tau)(X_k - \mu_k)]$
- $\hat{\mu} = \text{undefined}, \quad \hat{\mu}_t = \frac{1}{m}\sum_{i=1}^{m} x_t^{(i)}$
- $\hat{\gamma}(\tau, k) = \frac{1}{n-1}\sum_{i=1}^{N}(x_{i\tau} - \mu_\tau)(x_{ik} - \mu_k)$

Some math we'll need later

Autocorrelation

- $\rho(\tau) = \frac{\gamma(\tau,k)}{\sqrt{\gamma(\tau,\tau)\gamma(k,k)}}$
- $\hat{\rho}(\tau, k) = \frac{\hat{\gamma}(\tau,k)}{\sqrt{\hat{\gamma}(\tau,\tau)\gamma(k,k)}}$
- *With only one realization x_t, we can't compute this*

Stationarity

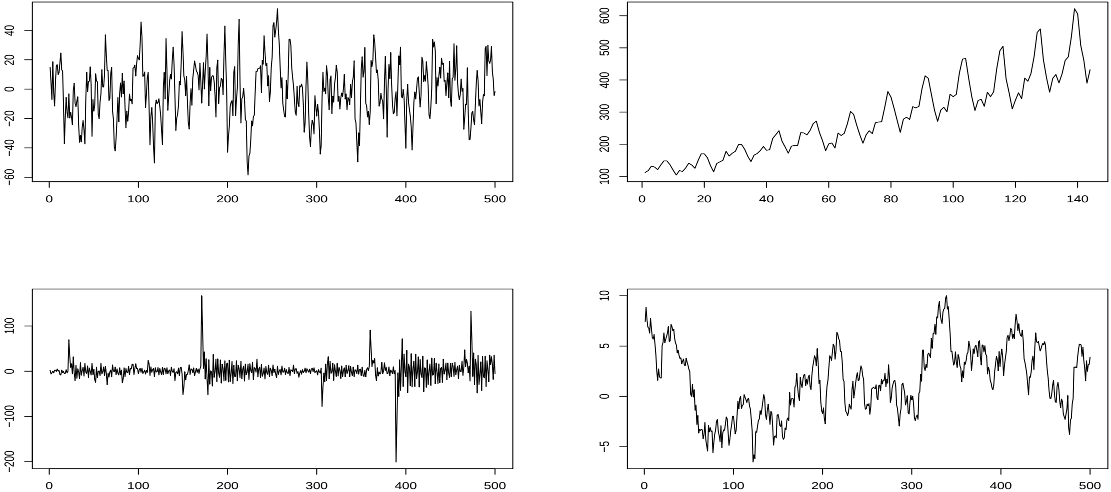
What, why and how?

- Strict stationarity
  - $F_X(X_t, \ldots, X_{t+k}) = F_X(X_{t+\tau}, \ldots, X_{t+\tau+k})$ for all $t, \tau, k \in \mathbb{Z}$
  - Time and order do not matter

- Weak stationarity
  - $E[X_t] = \mu$ for all t
  - $E[X^2] < \infty$
  - $E[(X_t - \mu)(X_{t+\tau} - \mu)] = \gamma(\tau)$ for all t and any τ

Stationarity

A short quiz



Stationarity

For ~~dummies~~ non-statisticians

- Data can't be stationary or non-stationary
- Stationarity is a property of processes
- Correct question: "*Was my data generated by a stationary process?*"
- Roughly: "no change over time"

Stationarity

Why?

- Classical statistics require strict stationarity
- Most models require at least weak stationarity
- Transformation to stationary form often possible
- Non-stationary theory is complex
- We can estimate autocorrelation

Stationarity

How?

- Augmented Dickey-Fuller test
- Priestley-Subba Rao test
- Hyndman's suggestion
- ~~Visual inspection~~

The ARMA model

Not the video game series. . .

- $\mathrm{AR}(1): \; X_t = c + \theta X_{t-1} + \epsilon_t$
- $\mathrm{AR}(p): \; X_t = c + \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t_p} + \epsilon_t$
- Simple linear model of past
- Stationary if $\sum \theta$ is small
- Least squares parameter fitting

Moving Average-Model (MA)

- $\mathrm{MA}(1): \; X_t = c + \epsilon_t + \phi\epsilon_{t-1}$
- $\mathrm{MA}(q): \; X_t = c + \epsilon_t + \phi_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + \ldots + \phi_q\epsilon_{t-q}$
- Don't confuse with rolling average
- Always weakly-stationary
- Assume distribution and maximize likelihood

ARMA model

Autoregressive Moving Average-Model

- $\mathrm{ARMA}(p, q): \; X_t = c + \sum_{i=1}^{p} \theta_i X_{t-i} + \sum_{j=1}^{q} \phi_j \epsilon_{t-j} + \epsilon_t$
- $\mathrm{ARMA}(p, q): \; x_t = \mathrm{AR}(p) + \mathrm{MA}(q) - c - \epsilon_t$
- Approximates large p or q
- Stationary if AR part stationary
- Parameter fitting as above

How can we choose p and q?

ARMA order estimation

ARMA order estimation

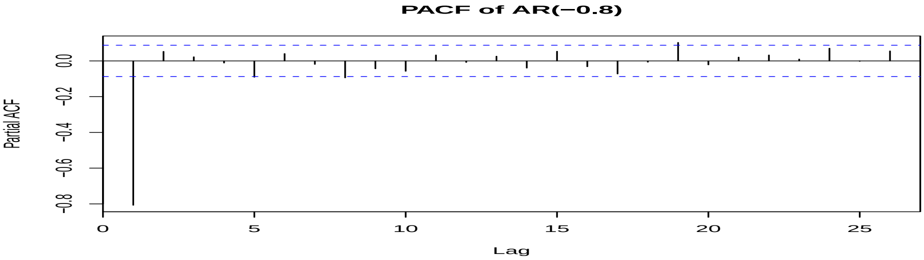
Partial autocorrelation

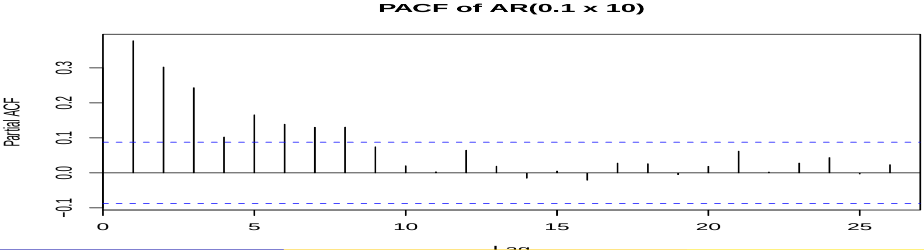
- $\alpha(1) = \rho(1)$
- $\alpha(\tau) = \frac{\mathbb{E}[(X_{\tau+1} - P_{\overline{\text{sp}}\{1, X_2, \ldots, X_\tau\}}(X_{\tau+1}) - \mu)(X_1 - P_{\overline{\text{sp}}\{1, X_2, \ldots, X_\tau\}}(X_1) - \mu)]}{\sqrt{\mathbb{E}[(X_{\tau+1} - P_{\overline{\text{sp}}\{1, X_2, \ldots, X_\tau\}}(X_{\tau+1}) - \mu)^2]\mathbb{E}[(X_1 - P_{\overline{\text{sp}}\{1, X_2, \ldots, X_\tau\}}(X_1) - \mu)^2]}}$
- ACF with lagged values estimated by linear model
- Usually Yule-Walker equations or OLS

- $\alpha(\tau \leq p)$ will be non-zero
- $\alpha(\tau > p)$ will be zero
- Compute $\hat{\alpha}$
- p is lag where $\hat{\alpha}$ enters confidence borders

ARMA order estimation

Estimating AR order p





ARMA order estimation

Estimating MA order q

- Plot ACF
- q is lag where ACF becomes zero
- Hyndman's method for stationary

ARMA order estimation

The Box-Jenkins Method

ACF Shape	**Indication**
Some spikes, almost zero	MA model, $q =$ time to first zero
Exponential decay to zero	AR model, plot PACF to find p
Alternating exp. decay to zero	AR model, plot PACF to find p
Delayed decay	ARMA model
Peaks at fixed intervals	Data are seasonal, use SARMA
Never reaches zero	Probably not stationary, detrend
Everything almost zero	Data are independent, noise

D_t and S_t

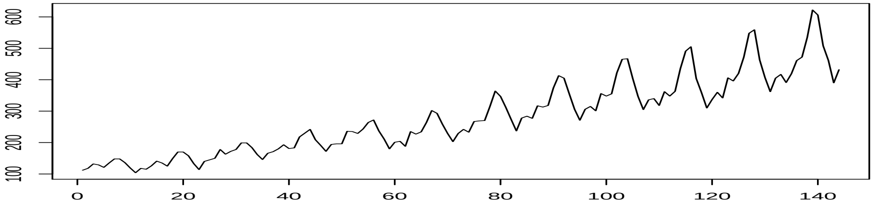
Trend and Seasonality

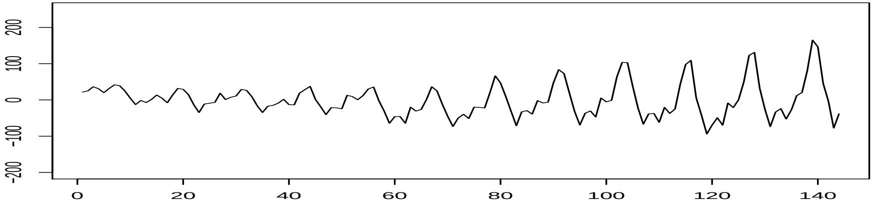
- $X_t = D_t + S_t + Y_t \quad D_t = f(t),\ S_t = g(t), S_t = S_{t+k}$
- $Y_t \ldots$stochastic residual
- Estimate $\hat{D}_t$ and $\hat{S}_t$
- Subtract and analyze residual

- Filters
  - Assume $S_t = 0 \ \forall t$
  - Remove arbitrary polynomial
- Regression
  - Linear
  - Non-isotonic
  - Isotonic
- Differencing
  - Stochastic trend
  - $\nabla(X_t) = X_t - X_{t-1}$
- $\mathrm{ARIMA}(p, d, q)$: Integrate $\mathrm{AR}(p) + \mathrm{MA}(q)$ d times

Trend and Seasonality

Detrending: Example





- Repeating events $\rightarrow$ Fourier Analysis
- Periodogram:
  - Fourier Sequence $\mathcal{F}_n(\omega)$
  - Fast Fourier Transform of ACF
- Peak Analysis: $s = \frac{1}{\arg\max_{\omega}(F_n)}$
- $\mathrm{SARIMA}(p, d, q)(P, \mathcal{D}, Q)_s$

Tools

Some software recommendations

- R
  - http://www.statmethods.net/advstats/timeseries.html
  - https://cran.r-project.org/web/views/TimeSeries.html
  - https://github.com/robjhyndman/
- Python
  - Prophet
  - TS-Fresh
  - Pandas, NumPy, scikit-learn, Statsmodels
- MatLab/Octave
  - TSA
  - Signal
  - ...
- Java
  - JMotif
  - Weka
  - ...

One last thing. . .

Remarks about artificial neural networks

- Feedforward ANN simulates nonlinear-$\mathrm{MA}(q)$
- Recurrent ANN simulates nonlinear -$\mathrm{ARMA}(p, q)$
- Autoregressive ANN $\neq \mathrm{AR}(p)$
- Long Short-Term Memory, Gated Recurrent Units

The End

(Now is the right time for questions)

```
library(forecast)

ts_data <- AirPassengers %>% c() %>% as.ts()
ts_data %>% plot.ts()


model1 <- Arima(ts_data,order = c(2,0,0))
model1 %>% forecast %>% plot(showgap=F)
model1$sigma2
model1$aic

ts_data %>% plot.ts()


model2 <- Arima(ts_data, order = c(0,0,2))
model2 %>% forecast %>% plot(showgap=F)
model2$sigma2
model2$aic

ts_data %>% plot.ts()

model3 <- Arima(ts_data, order = c(2,0,2))
model3 %>% forecast %>% plot(showgap=F)
model3$sigma2
model3$aic

ts_data %>% plot.ts()
ts_data %>% acf(lag.max = 50)
ts_data %>% pacf()

model4 <- Arima(ts_data, order = c(2,0,0))
model4 %>% forecast %>% plot(showgap=F)
model4$sigma2
model4$aic

detrended_data <- ts_data %>% diff()
detrended_data %>% plot()

model5 <- Arima(ts_data, order = c(2,1,0))
model5 %>% forecast %>% plot(showgap=F)
model5$sigma2
model5$aic

detrended_data %>% plot()
detrended_data %>% acf()
detrended_data %>% pacf()

model6 <- Arima(ts_data, order = c(2,1,1))
model6 %>% forecast() %>% plot(showgap=F)
model6$sigma2
model6$aic

detrended_data %>% acf(lag.max = 100)
pgram <- ts_data %>% spec.pgram()
{pgram$spec} %>% which.max() %>% {1/pgram$freq[.]}

model7 <- Arima(ts_data, order = c(2,1,1),seasonal = list(order=c(0,1,0),period=12))
model7 %>% forecast() %>% plot(showgap=F)
model7$sigma2
model7$aic

model8 <- auto.arima(ts_data)
model8 %>% forecast() %>% plot(showgap=F)
model8$sigma2
model8$aic
```