Machine Learning: Advanced

Introduction to Time-series modeling, forecasting

SUVRIT SRA

Massachusetts Institute of Technology

[ml.mit.edu](ml.mit.edu)

What is a time series?

- Collection of observations $x_1, \ldots, x_n$ indexed by time

What is a time series?

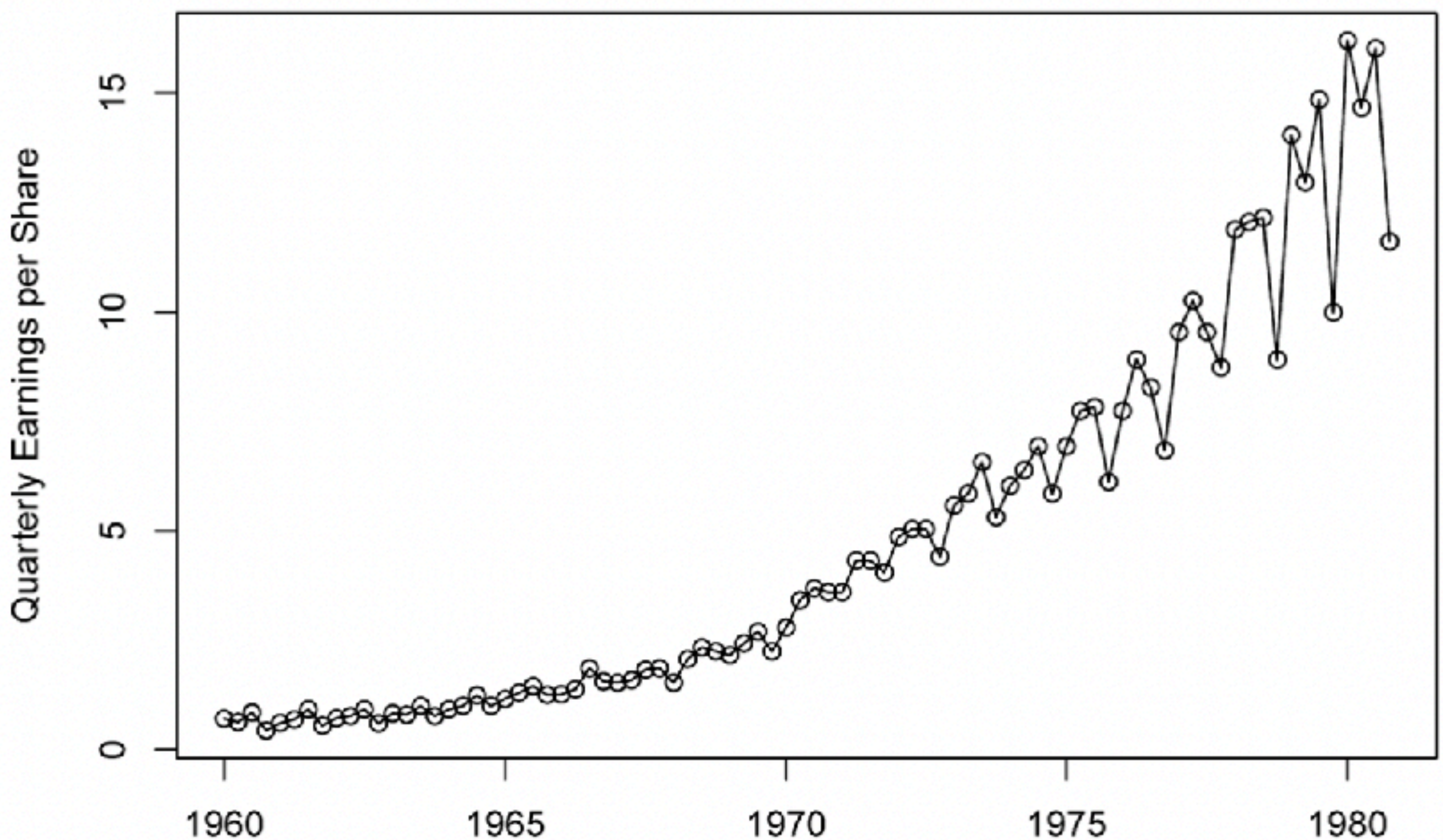
- Collection of observations $x_1, \ldots, x_n$ indexed by time
- Come up in
  - economics, finance
  - social sciences
  - epidemiology
  - fMRI/neuroscience
  - environmental measurements
  - speech analysis …

What is a time series?

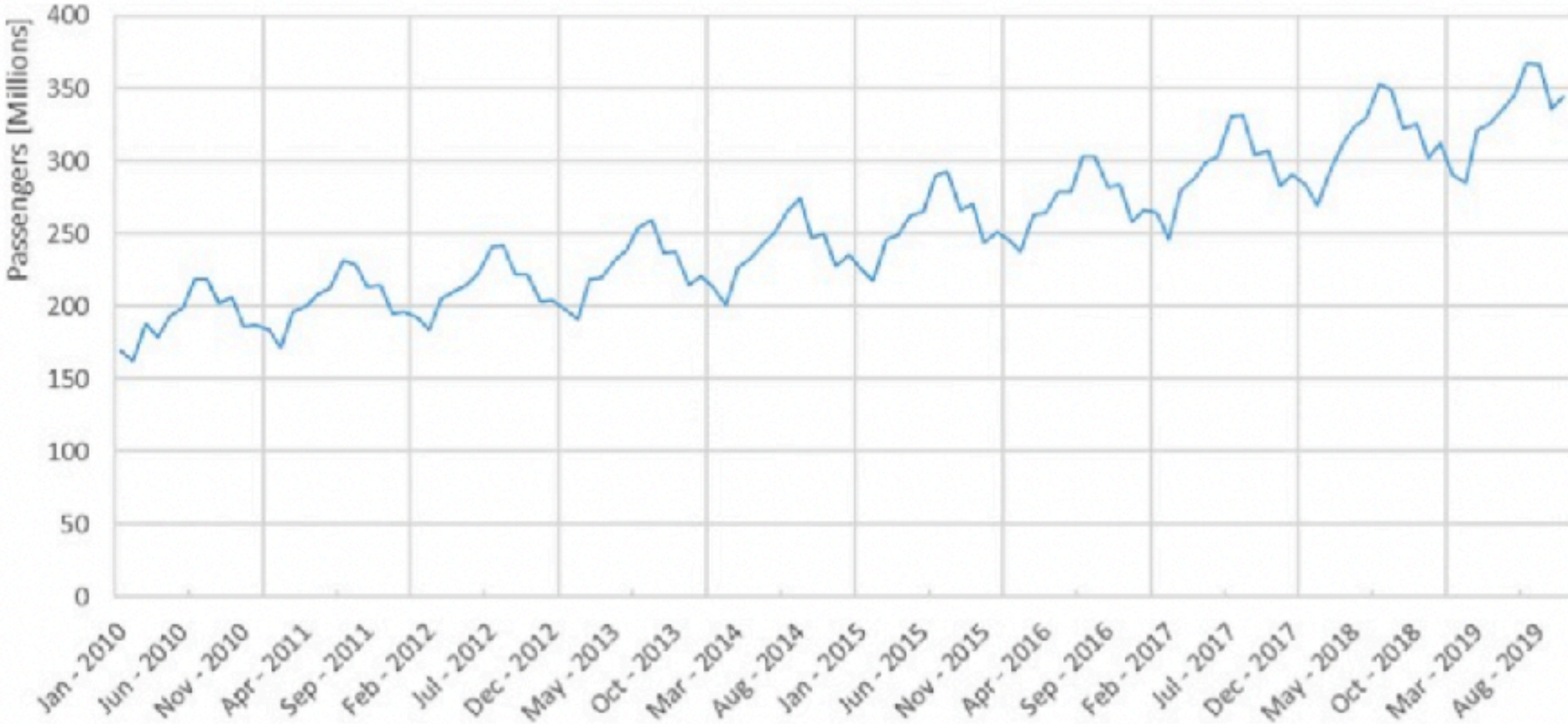
- Collection of observations $x_1, \ldots, x_n$ indexed by time
- Come up in
  - economics, finance
  - social sciences
  - epidemiology
  - fMRI/neuroscience
  - environmental measurements
  - speech analysis …

- What do time series look like?

Johnson & Johnson quarterly earnings per share

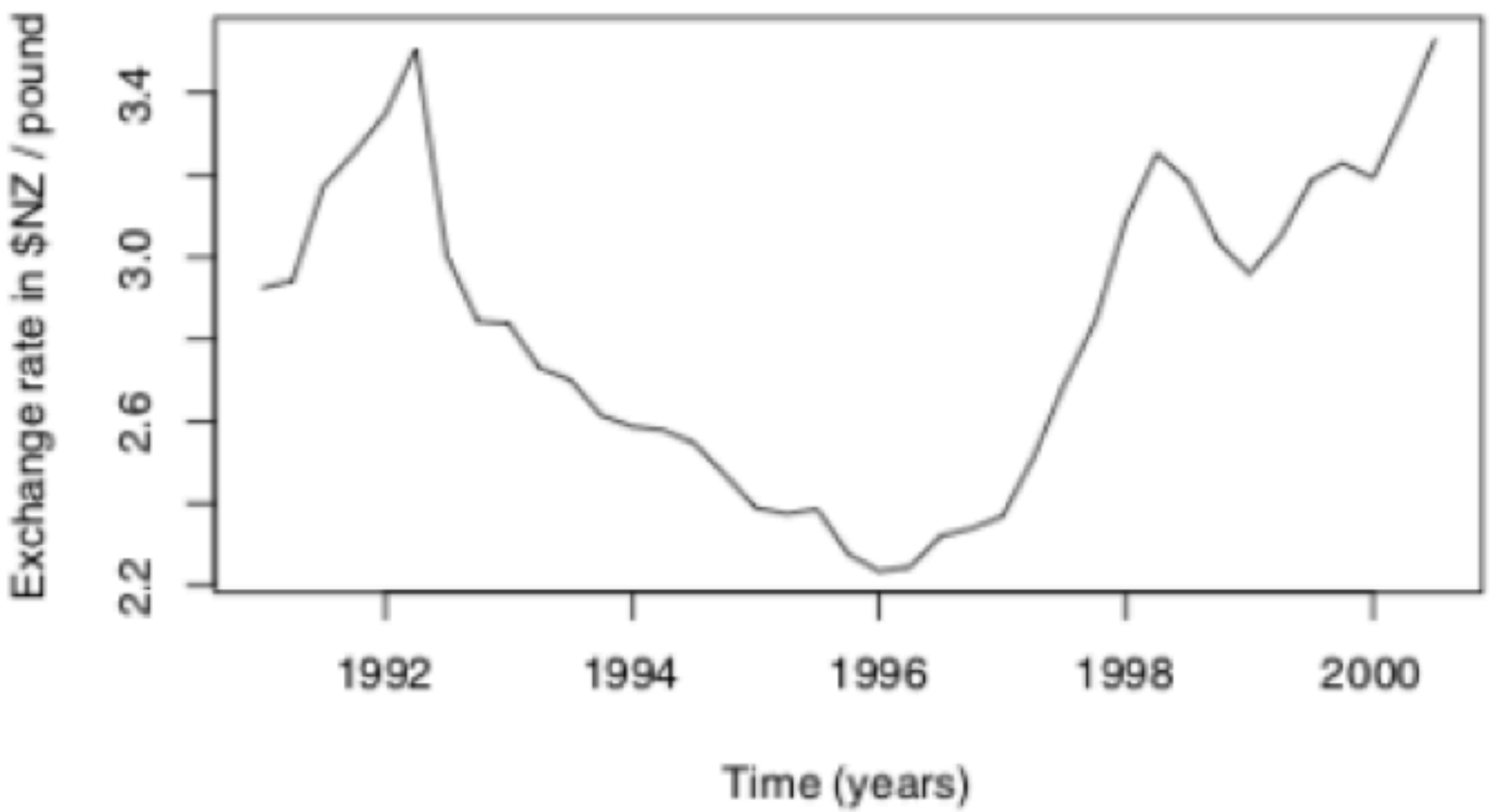


Global Airline Passengers

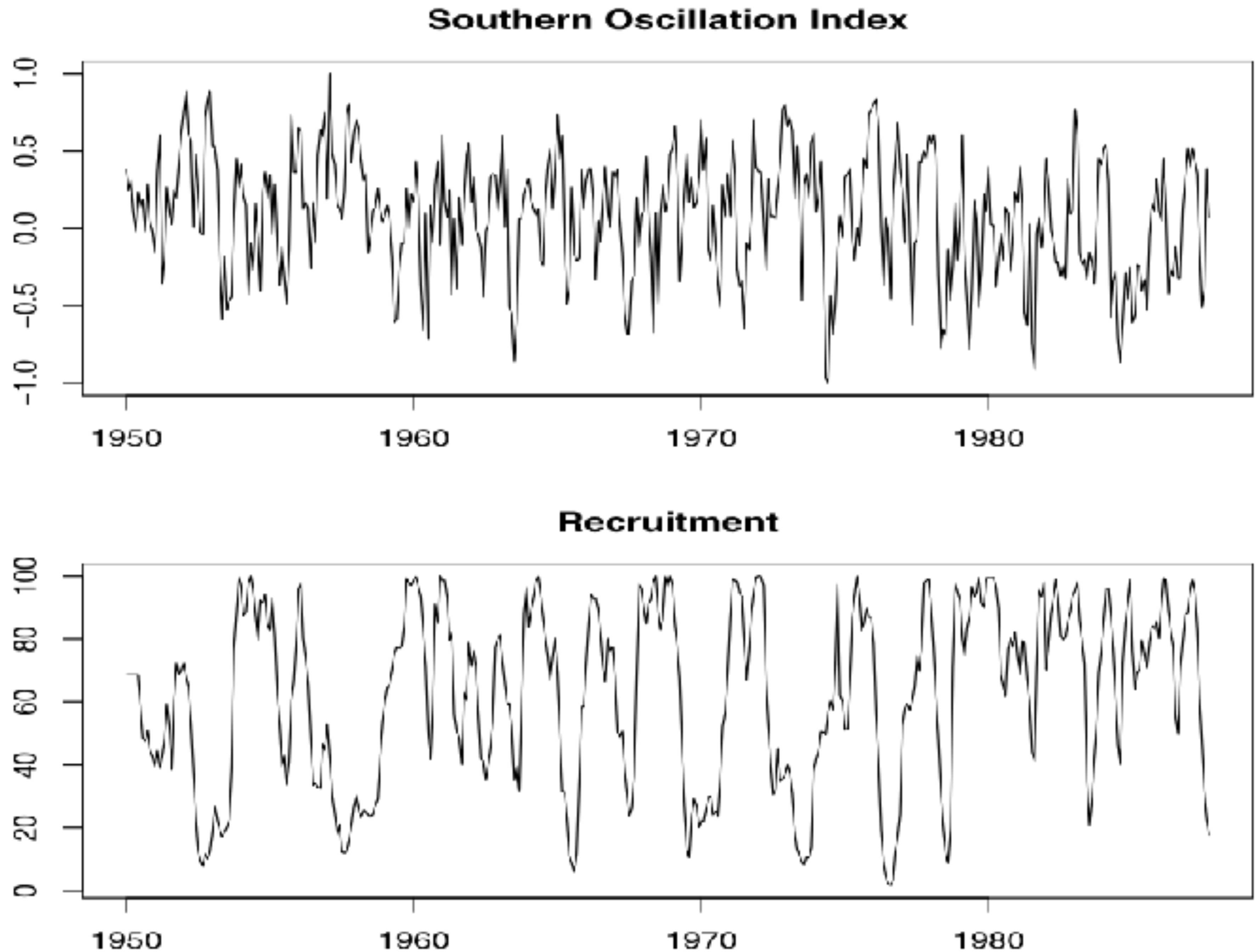


source: S. M. Iacus, F. Natale, C. Santamaria, S. Spyratos, M. Vespe. Estimating and projecting air passenger traffic during the COVID-19 coronavirus outbreak and its socio-economic impact. *Safety Science* vol 129, 2020.

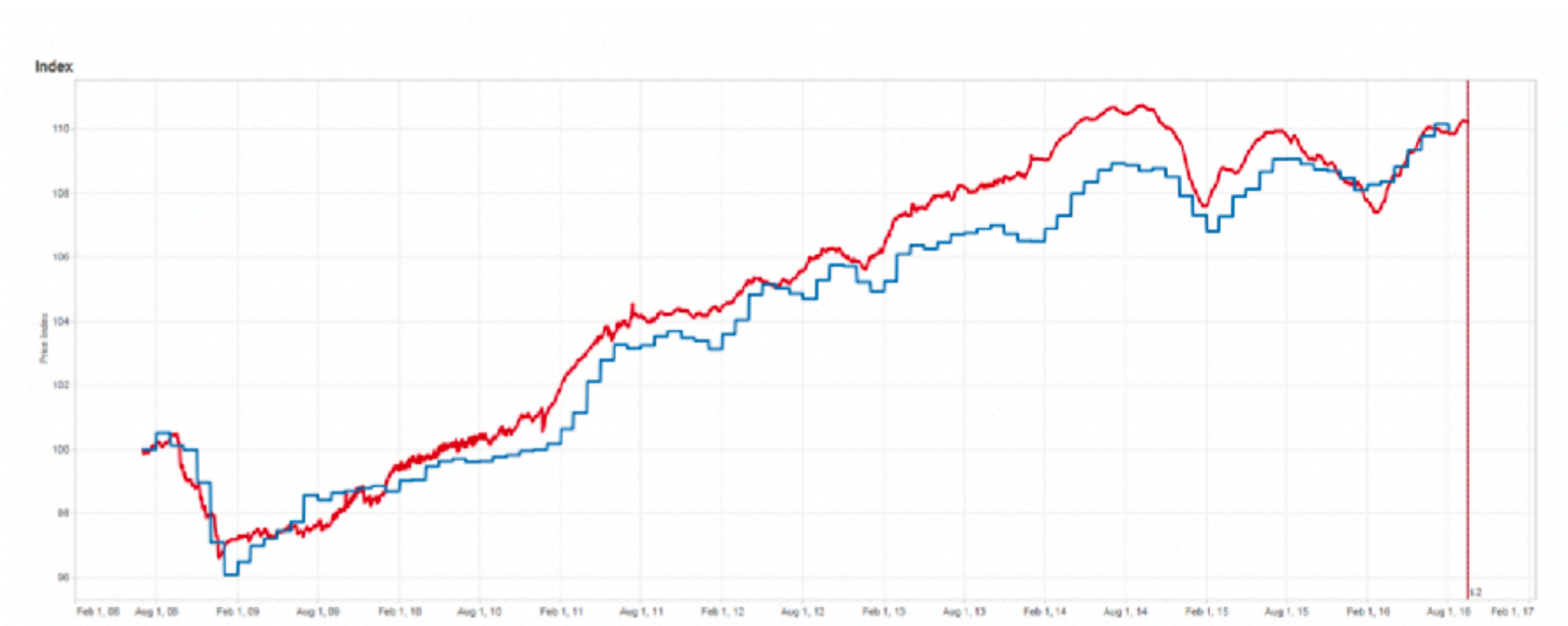
Exchange Rate GBP to $NZ



Air Pressure & Fish Population



Price-index USA



2008 2017

source: Roberto Rigobon

Objective of TS analysis

Time series: collection of random variables $X_0, \ldots, X_n$ indexed by time

- Descriptive analysis (visualization, components, dependencies)
- Modeling
- Forecasting
- Time Series Regression
- Control
- …

How can we model a time series?

usually: local models, e.g. Autoregressive of order p, AR(p)

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t-p} + W_t$$

"linear regression in time"

How can we model a time series?

usually: local models, e.g. Autoregressive of order p, AR(p)

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t-p} + W_t$$

noise

"linear regression in time"

How can we model a time series?

usually: local models, e.g. Autoregressive of order p, AR(p)

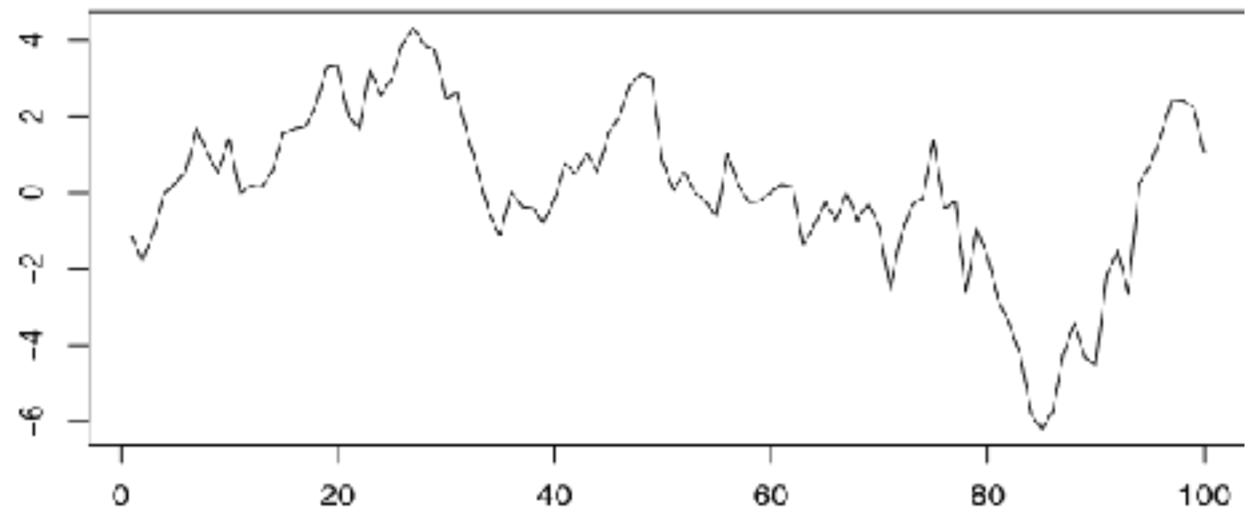
$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t-p} + W_t$$

noise

“linear regression in time”

AR(1) examples

$$X_t = 0.9 X_{t-1} + W_t$$



How can we model a time series?

usually: local models, e.g. Autoregressive of order p, AR(p)

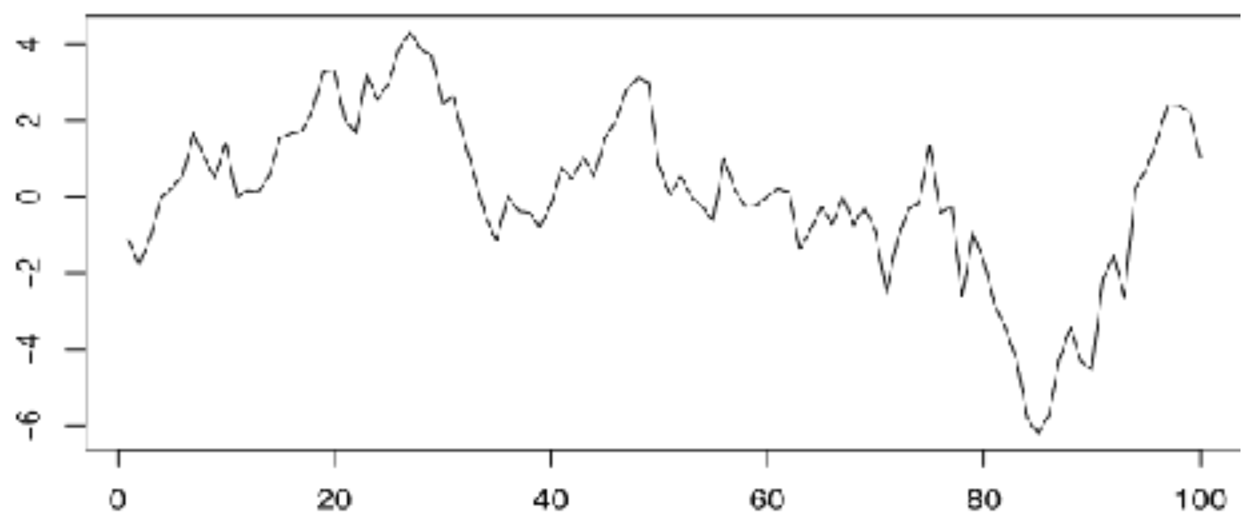
$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t-p} + W_t$$

noise

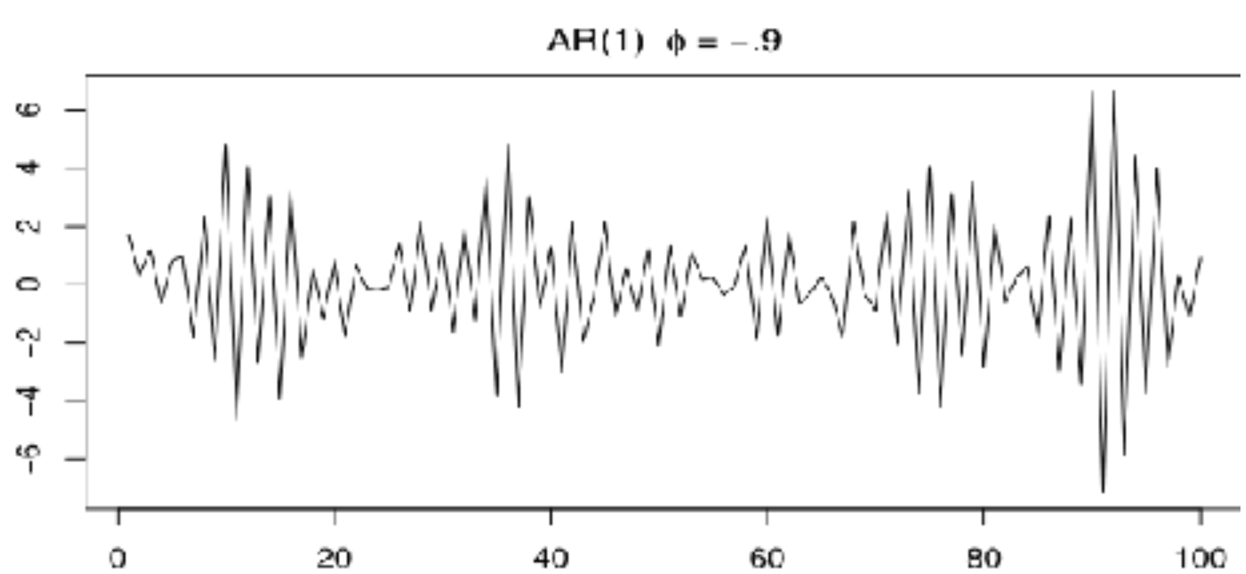
"linear regression in time"

AR(1) examples

$$X_t = 0.9X_{t-1} + W_t$$



$$X_t = -0.9X_{t-1} + W_t$$



How can we model a time series?

- usually: local models, e.g. Autoregressive of order p

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t-p} + W_t$$

noise

"linear regression in time"

- assume the same model applies everywhere in time

How can we model a time series?

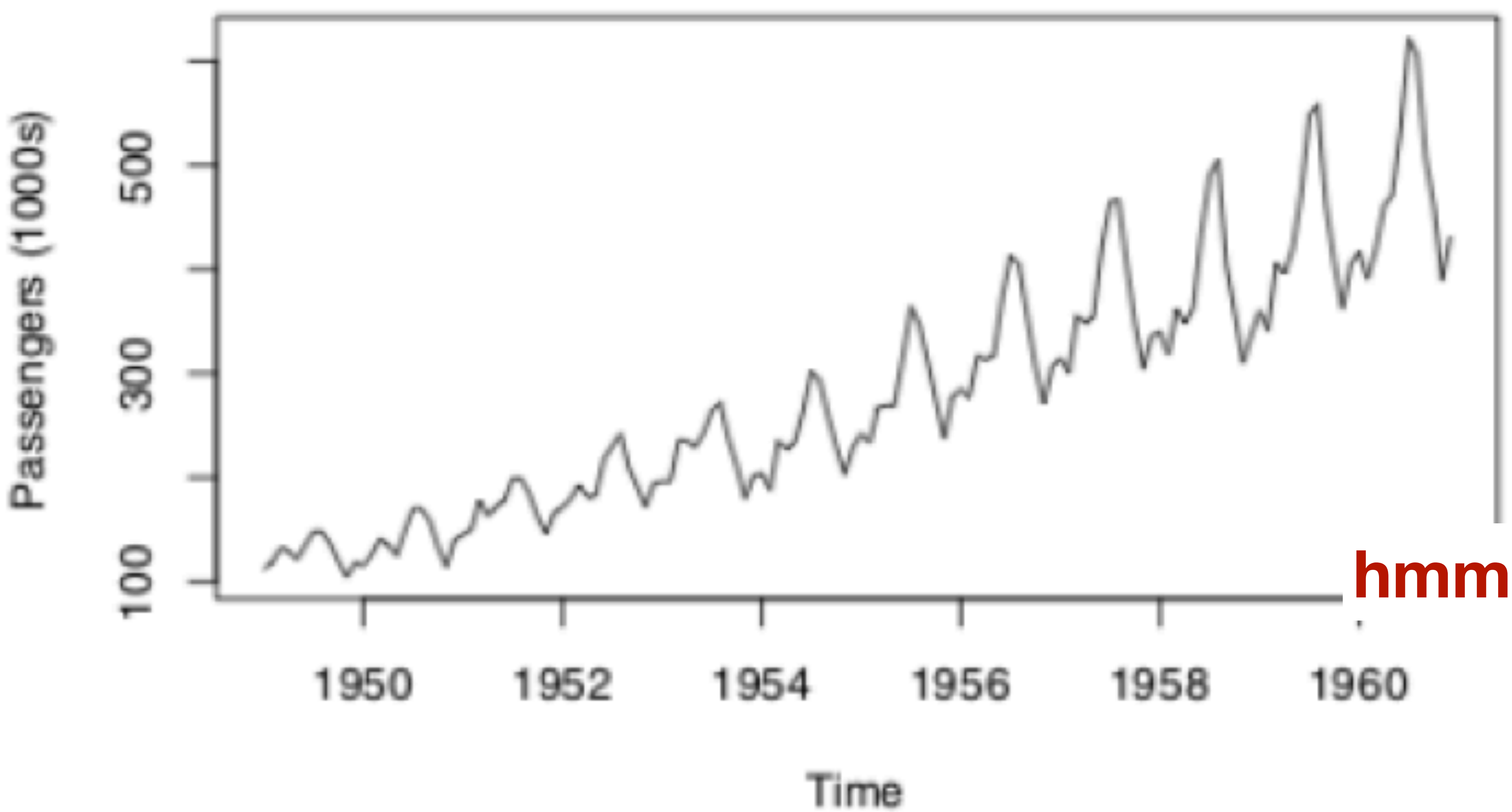
- usually: local models, e.g. Autoregressive of order p

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t-p} + W_t$$

noise

"linear regression in time"

- assume the same model applies everywhere in time



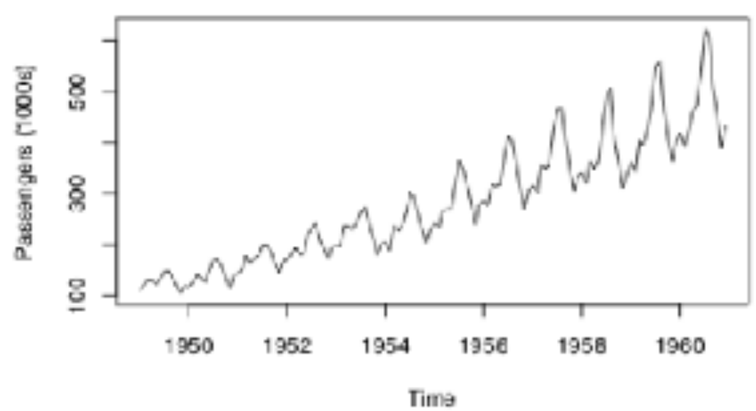
hmmmm…..?????

How can we model a time series?

- usually: local models, e.g. Autoregressive of order p:

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t-p} + W_t$$

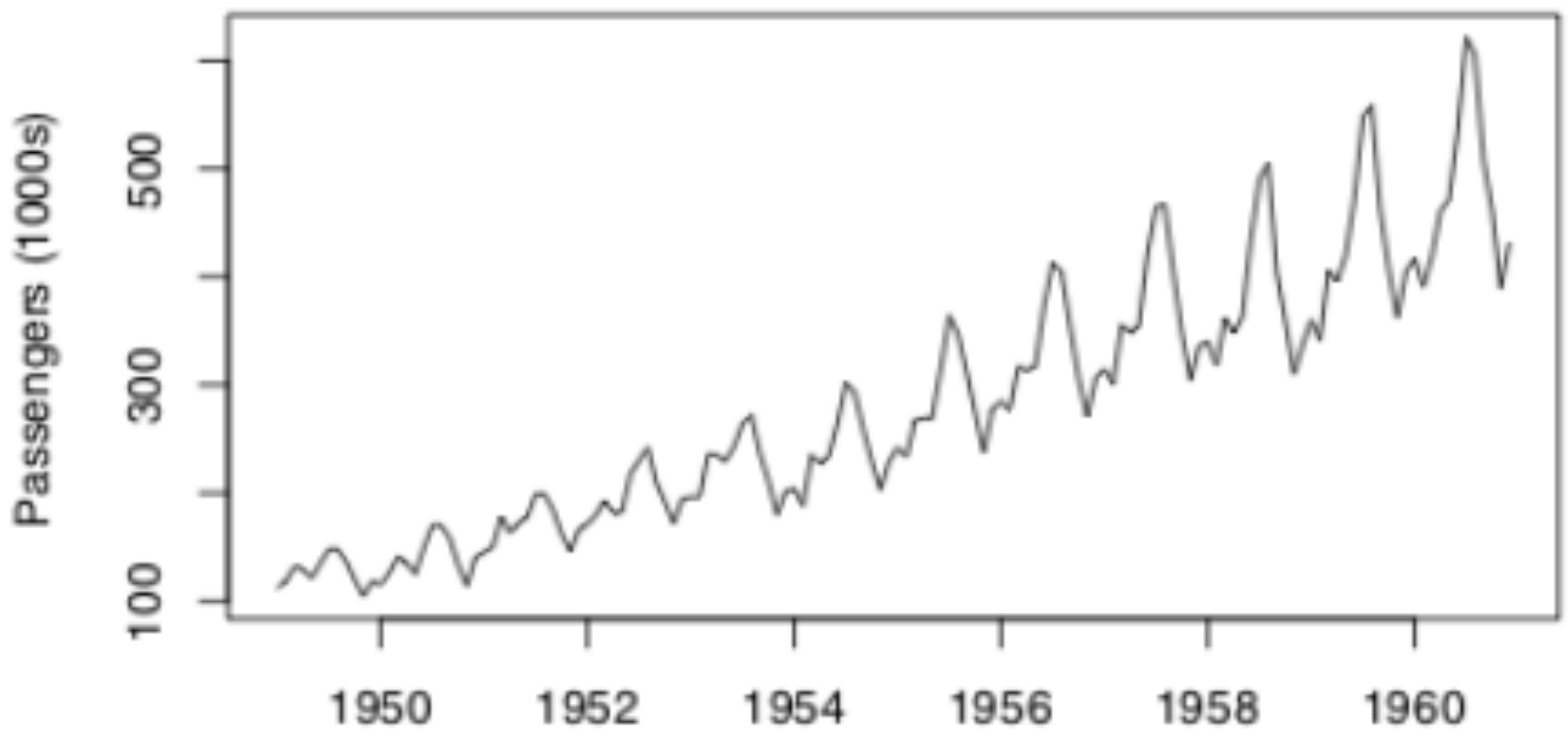
- assume the same model applies everywhere



➡need series to be "locally invariant": same average value & local relationships, independent of time
(*weak stationarity*: same mean and covariances)
first transform to be weakly stationary

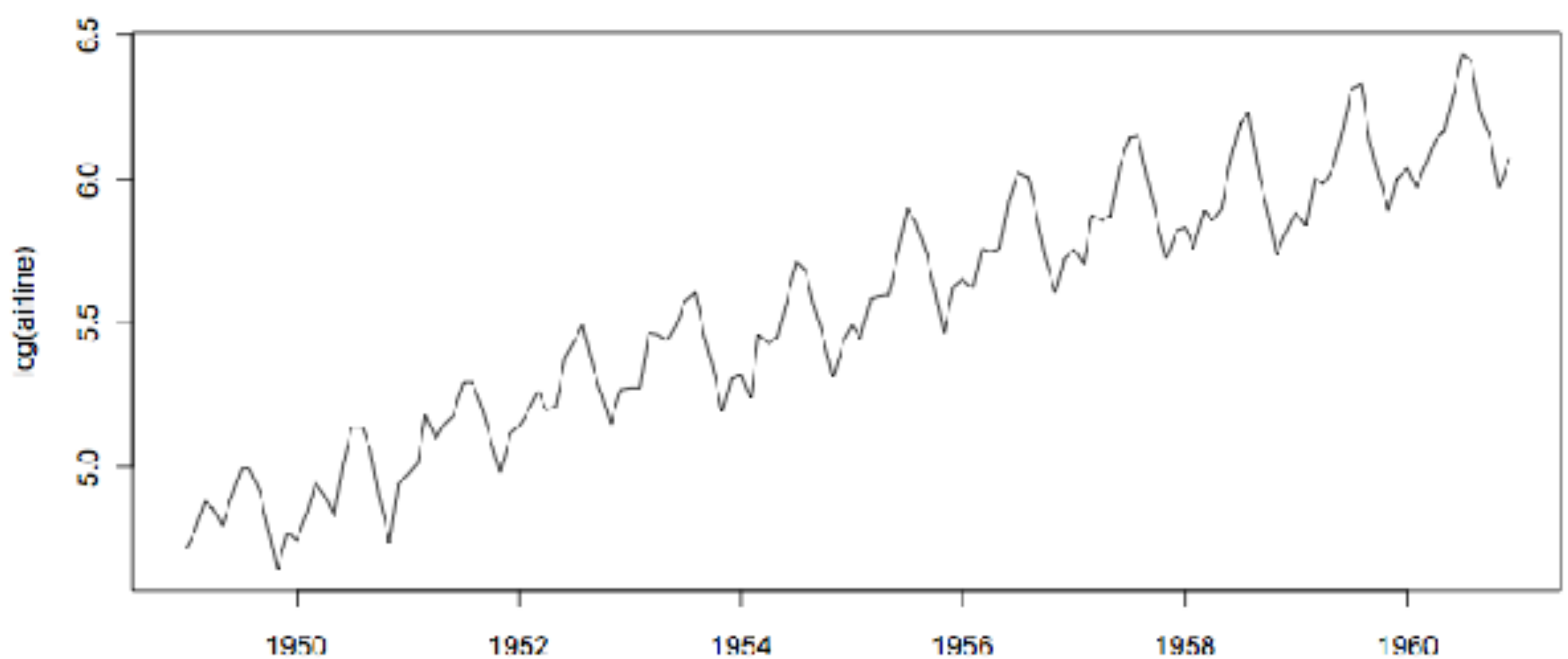
Achieving weak stationarity

- More uniform variation: nonlinear transformation, e.g. log



Log-Transformed Airline Passenger Data

log(x_t)

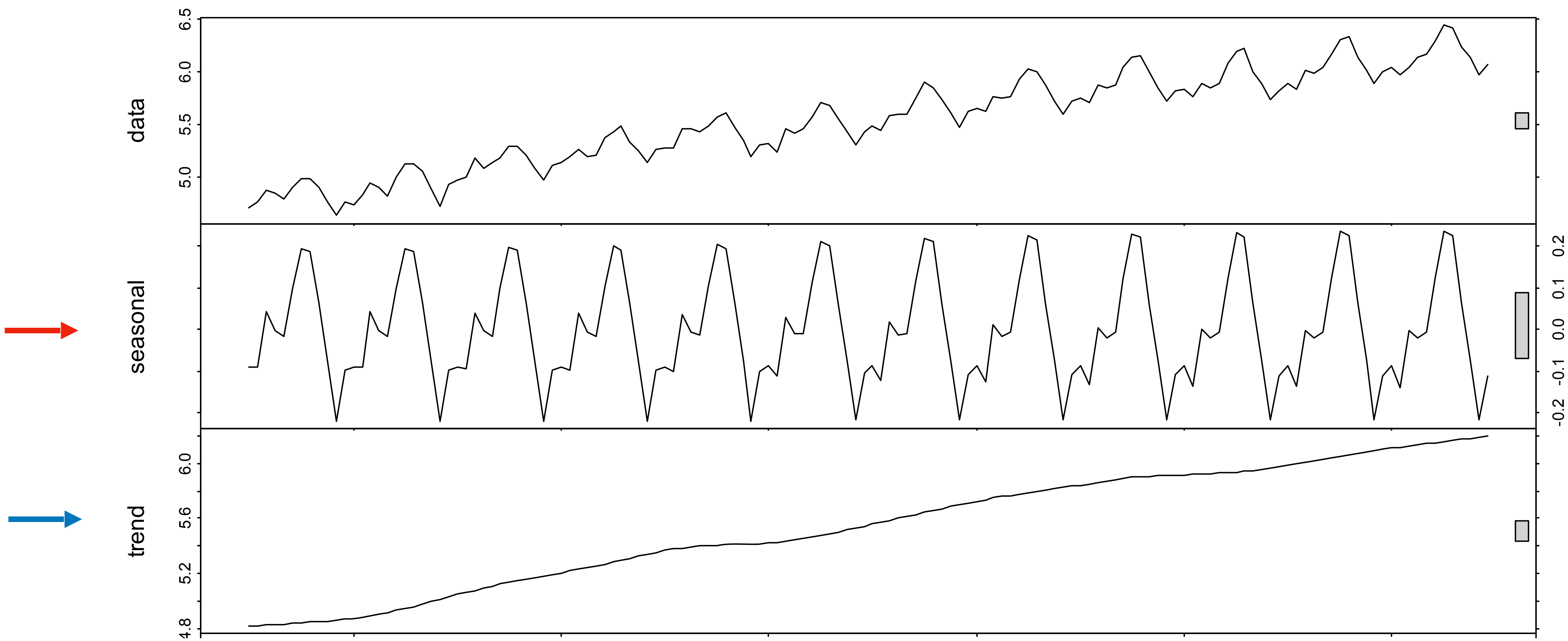


Achieving weak stationarity

- Decompose time series:

$$X_t = \color{blue}{T_t} + \color{red}{S_t} + Z_t$$

Linear Trend **Seasonal Component**

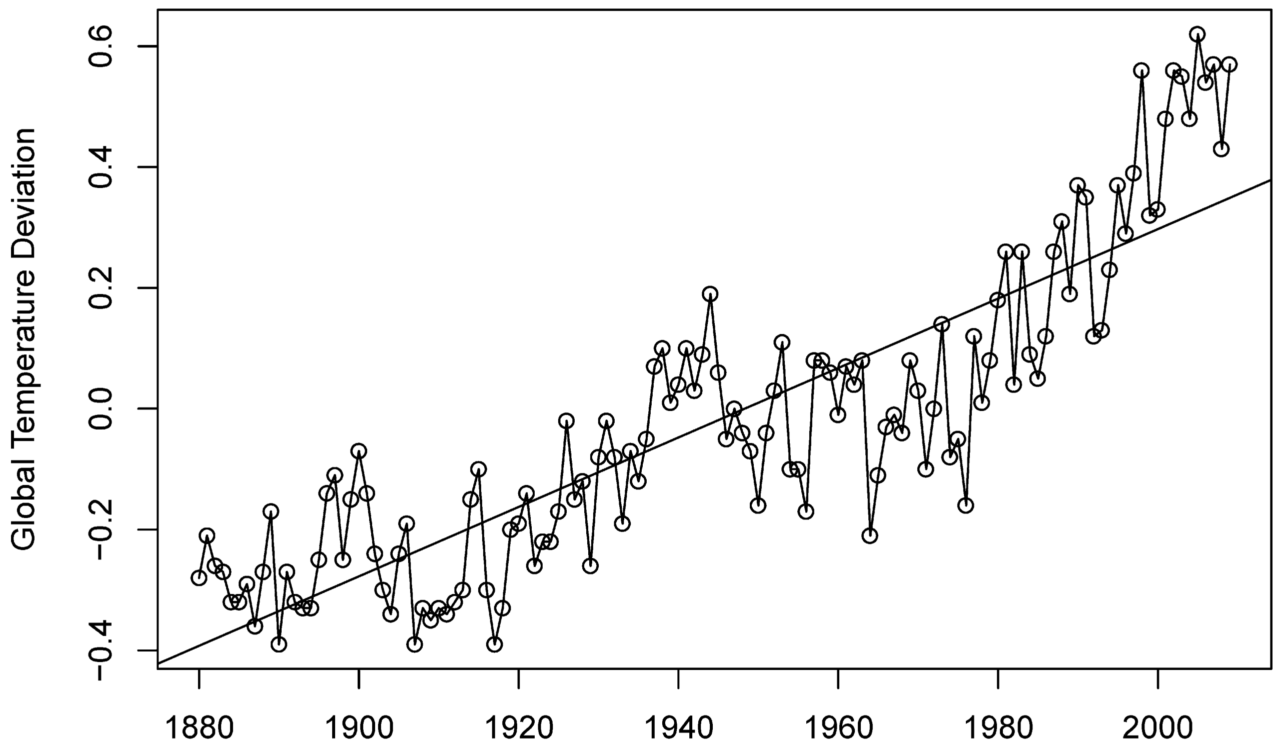


Achieving weak stationarity

- Decompose time series:

$$X_t = \color{blue}{T_t} + \color{red}{S_t} + Z_t$$

- Linear trend: linear regression with time as the feature, subtract linear trend

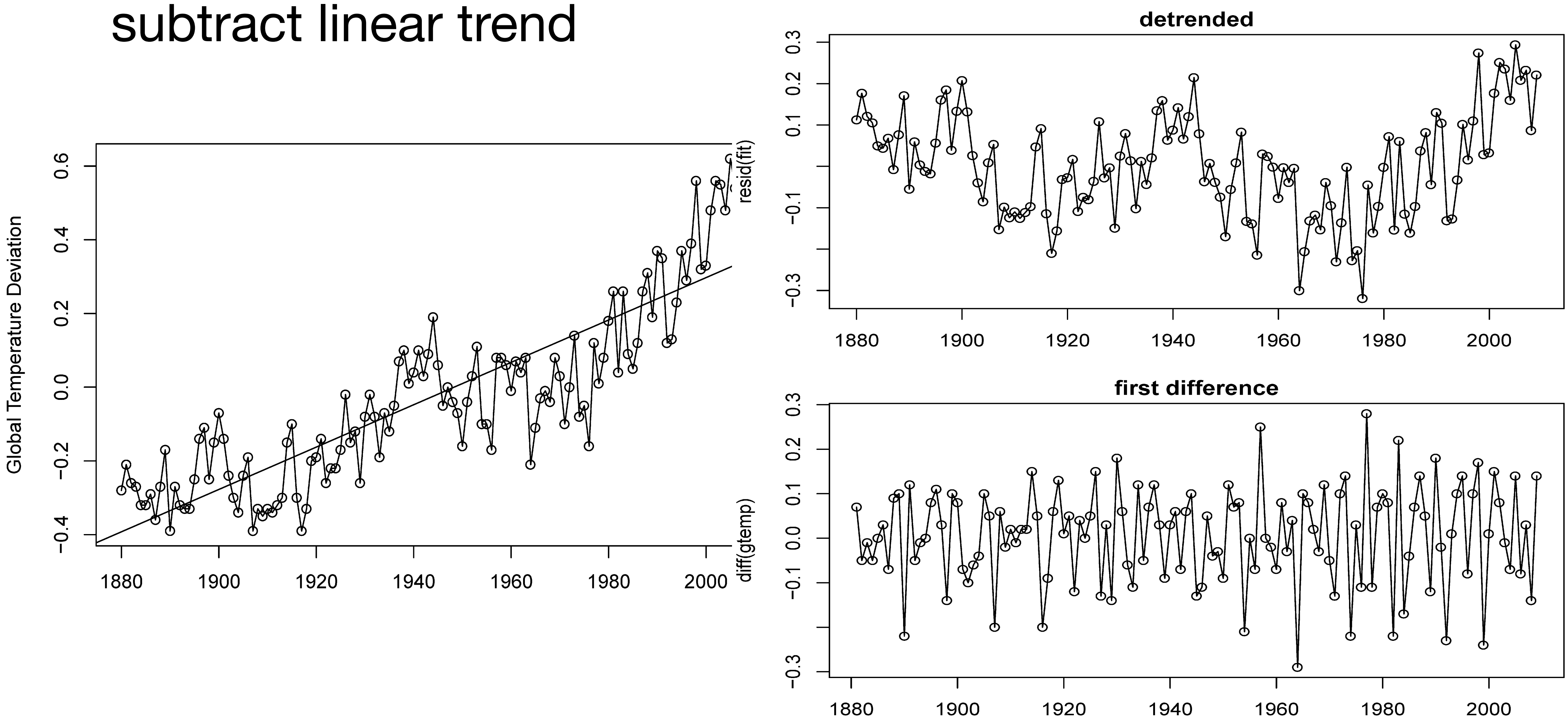


Achieving weak stationarity

- Decompose time series:

$$X_t = \color{blue}{T_t} + \color{red}{S_t} + Z_t$$

- Linear trend: linear regression with time as the feature, subtract linear trend



Stationarity remarks

- remove trends and seasonal components: $X_t = T_t + S_t + Z_t$
  - deterministic trend T_t: linear regression
  - deterministic seasonal component S_t
  - remainder: stationary, mean zero

- differentiation

$$
\begin{aligned}
Y_t = \nabla X_t &= X_t - X_{t-1} && \text{removes linear trend} \\
\nabla^2 X_t &= \nabla X_t - \nabla X_{t-1} \\
&= X_t - 2X_{t-1} + X_{t-2} && \text{removes quadratic trend}
\end{aligned}
$$

Linear Time Series Models

- Autoregressive of order p

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t-p} + W_t$$

Linear Time Series Models

- Autoregressive of order p

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t-p} + W_t$$

- Moving average of order q

$$X_t = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \ldots + \theta_q W_{t-q}$$

Linear Time Series Models

- Autoregressive of order p

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t-p} + W_t$$

- Moving average of order q

$$X_t = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \ldots + \theta_q W_{t-q}$$

- ARMA(p,q): autoregressive plus moving average

Linear Time Series Models

- Autoregressive of order p

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t-p} + W_t$$

- Moving average of order q

$$X_t = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \ldots + \theta_q W_{t-q}$$

- ARMA(p,q): autoregressive plus moving average

- good for predicting short horizons, not long ones

Time Series Models

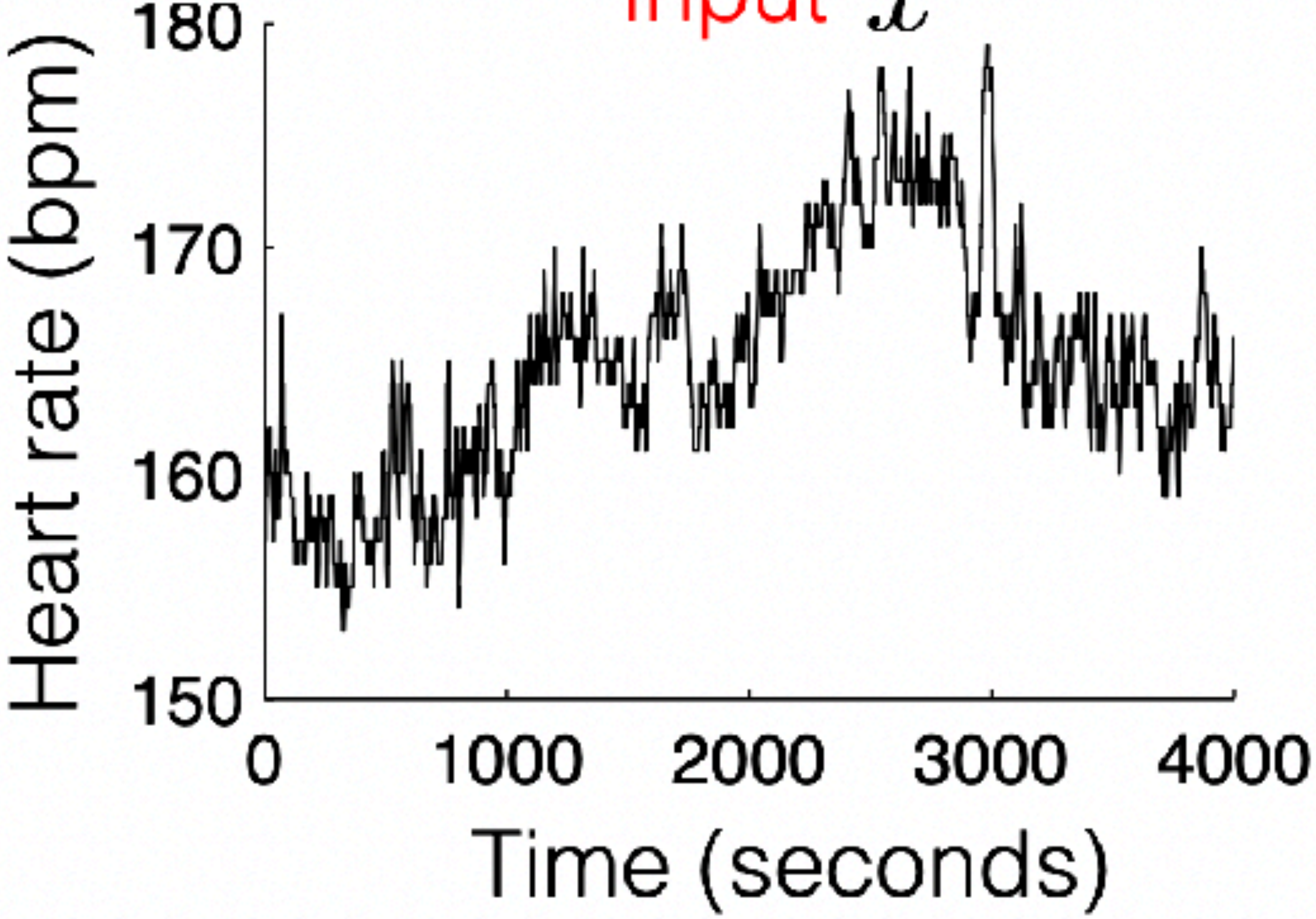
- Decompose & transform for local invariance (weak stationarity)
- Linear Models: AR, MA, ARMA: short horizons
- More advanced models: "State space models", neural network models (next)
- Caution about cross-validation in time series: data points are not independent!
  Need to separate train and test/validation

A few remarks about direct supervised learning approach….

Temporal/sequence problems

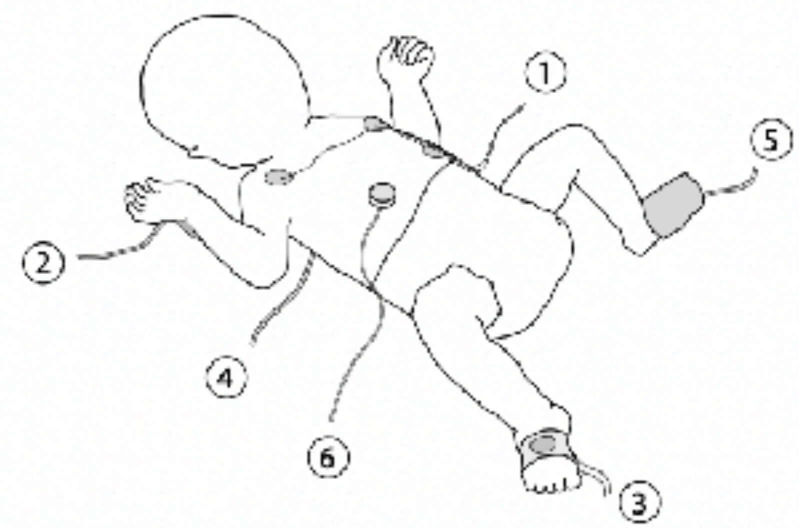
- How to cast as a supervised learning problem?

Input x



Output y

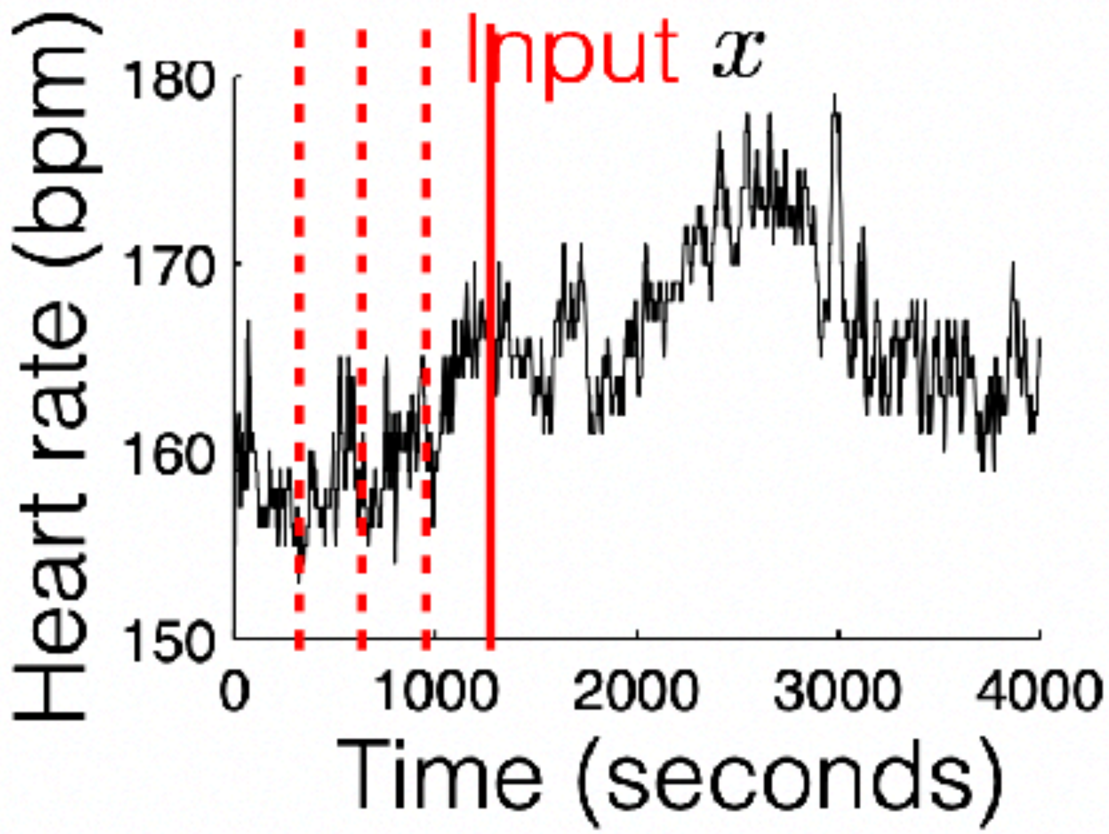
Likelihood of mortality?



Quinn et al., TPAMI 2008

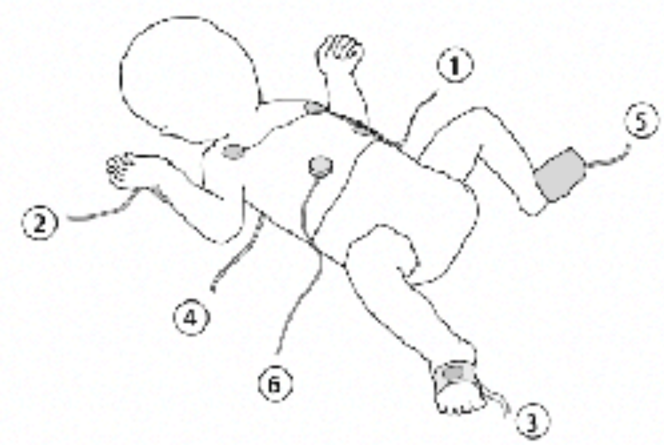
Temporal/sequence problems

- How to cast as a supervised learning problem?



Output y

Likelihood of mortality?

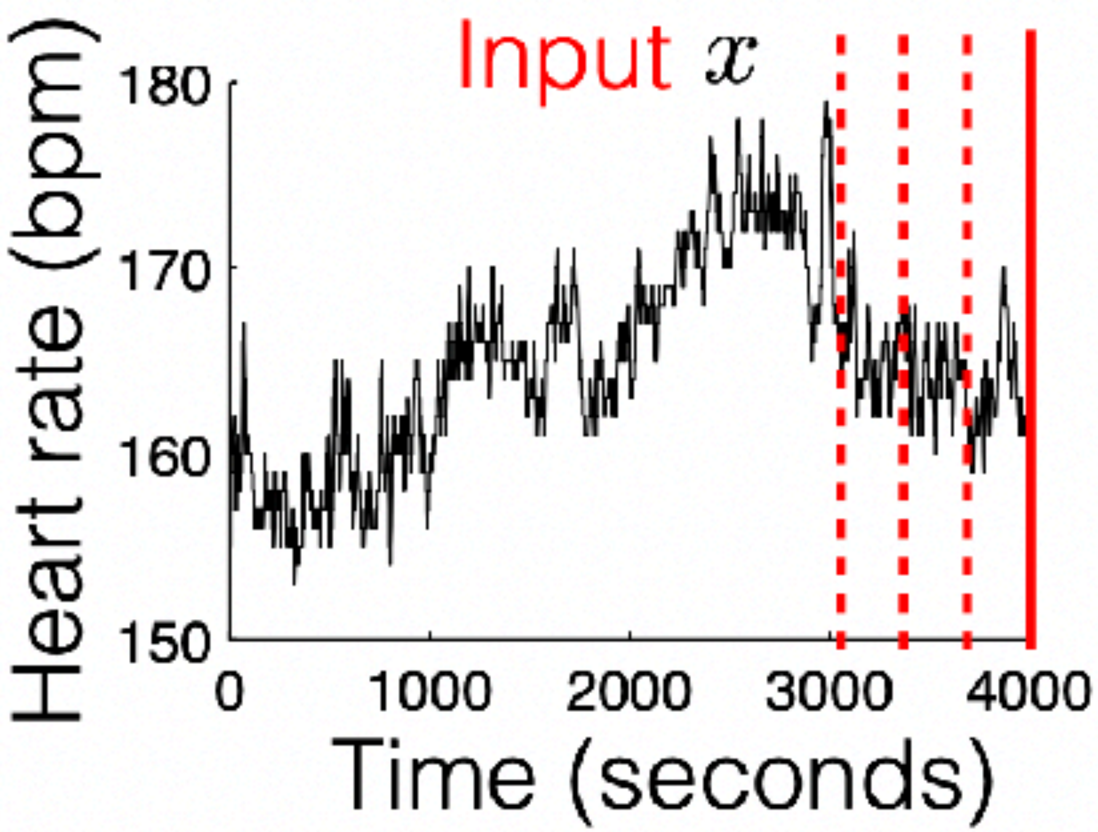


- Historical data can be broken down into feature vectors and target values (sliding window)

$$\underset{x^{(t)}}{\langle 155, 160, 165, 167 \rangle} \qquad \underset{y^{(t)}}{0.35}$$

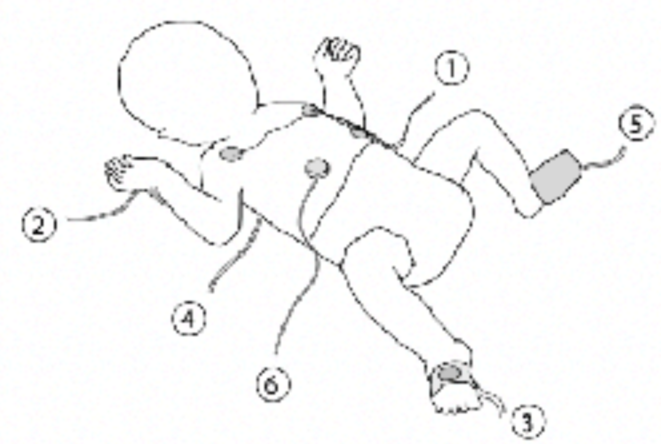
Temporal/sequence problems

- How to cast as a supervised learning problem?



Output y

Likelihood of mortality?



- Historical data can be broken down into feature vectors and target values (sliding window)

$$\underset{x^{(t)}}{\langle 170, 160, 155, 160 \rangle} \qquad \underset{y^{(t)}}{0.24}$$