Time Series

Karanveer Mohan   Keegan Go   Stephen Boyd

EE103
Stanford University

November 15, 2016

Outline

Time series data

- represent time series $x_1, \ldots, x_T$ as T-vector x
- x_t is value of some quantity at time (period, epoch) t, $t = 1, \ldots, T$
- examples:
  - average temperature at some location on day t
  - closing price of some stock on (trading) day t
  - hourly number of users on a website
  - altitude of an airplane every 10 seconds
  - enrollment in a class every quarter
- vector time series: x_t is an n-vector; can represent as $T \times n$ matrix

Types of time series

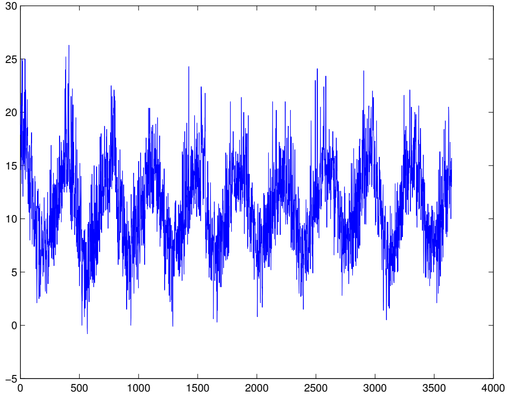
time series can be

- smoothly varying or more wiggly and random
- roughly periodic (*e.g.*, hourly temperature)
- growing or shrinking (or both)
- random but roughly continuous

(these are vague labels)

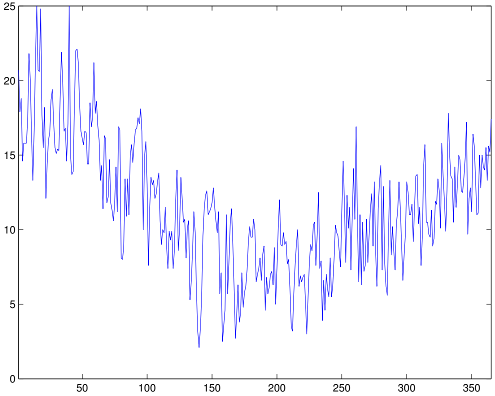
Melbourne temperature

- daily measurements, for 10 years
- you can see seasonal (yearly) periodicity



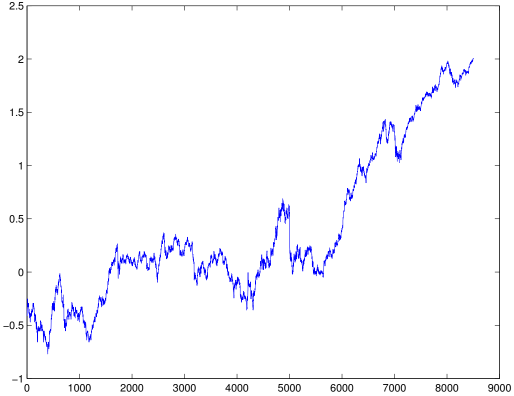
Melbourne temperature

- zoomed to one year



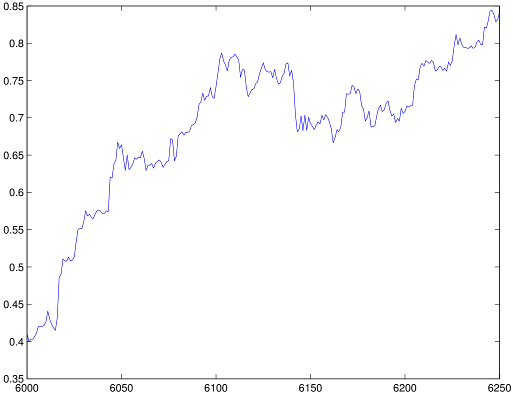
Apple stock price

- $\log_{10}$ of Apple daily share price, over 30 years, 250 trading days/year
- you can see (not steady) growth



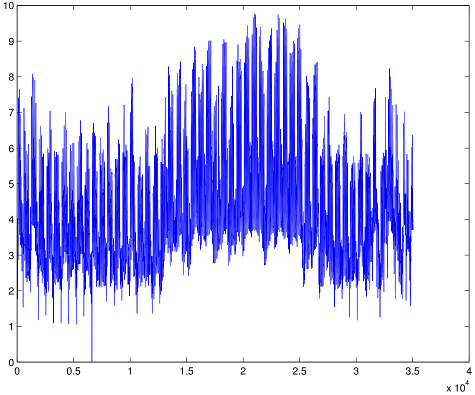
Log price of Apple

- zoomed to one year



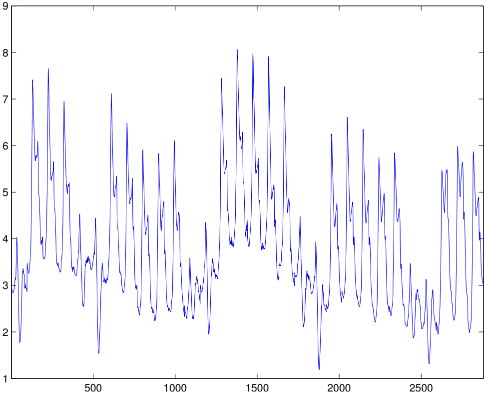
Electricity usage in (one region of) Texas

- total in 15 minute intervals, over 1 year
- you can see variation over year



Electricity usage in (one region of) Texas

- zoomed to 1 month
- you can see daily periodicity and weekend/weekday variation



Outline

Down-sampling

- $k\times$ *down-sampled* time series selects every kth entry of x
- can be written as $y = Ax$
- for $2\times$ down-sampling, T even,

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 \end{bmatrix}$$

- alternative: average consecutive k-long blocks of x

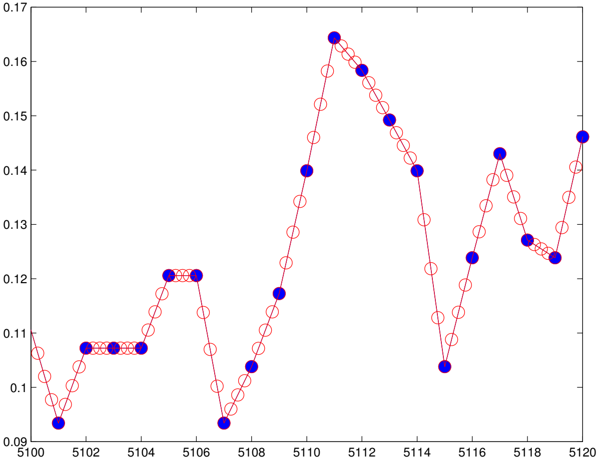
Up-sampling

- $k\times$ (linear) *up-sampling* interpolates between entries of x
- can be written as $y = Ax$
- for $2\times$ up-sampling

$$A = \begin{bmatrix} 1 & & & & & \\ 1/2 & 1/2 & & & & \\ & 1 & & & & \\ & 1/2 & 1/2 & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & & & 1 & \\ & & & & 1/2 & 1/2 \\ & & & & & 1 \end{bmatrix}$$

Up-sampling on Apple log price

$4\times$ up-sample



Smoothing

- k-long *moving average* y of x is given by

$$y_i = \frac{1}{k}(x_i + x_{i+1} + \cdots + x_{i+k-1}), \quad i = 1, \ldots, T-k+1$$

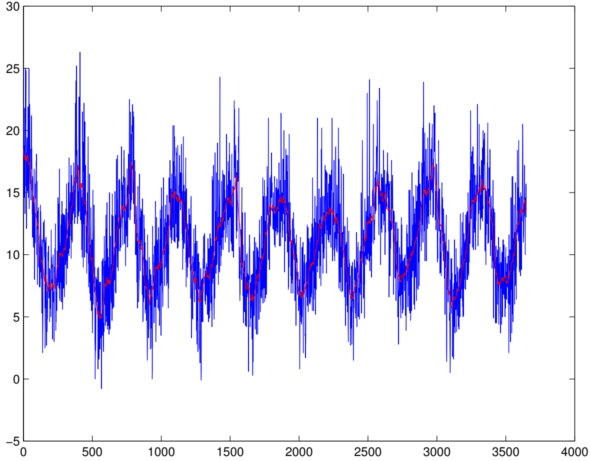
- can express as $y = Ax$, *e.g.*, for $k = 3$,

$$A = \begin{bmatrix} 1/3 & 1/3 & 1/3 & & & & & \\ & 1/3 & 1/3 & 1/3 & & & & \\ & & 1/3 & 1/3 & 1/3 & & & \\ & & & & \ddots & & & \\ & & & & & 1/3 & 1/3 & 1/3 \end{bmatrix}$$

- can also have trailing or centered smoothing

Melbourne daily temperature smoothed

- centered smoothing with window size 41



First-order differences

- (first-order) difference between adjacent entries
- discrete analog of derivative
- express as $y = Dx$, D is the $(T-1) \times T$ difference matrix

$$D = \begin{bmatrix} -1 & 1 & & \dots & & \\ & -1 & 1 & \dots & & \\ & & \ddots & \ddots & & \\ & & & \dots & -1 & 1 \end{bmatrix}$$

- $\|Dx\|^2$ (Laplacian) is a measure of the wiggliness of x

$$\|Dx\|^2 = (x_2 - x_1)^2 + \cdots + (x_T - x_{T-1})^2$$

Outline

Introduction

Linear operations

Least-squares

Prediction

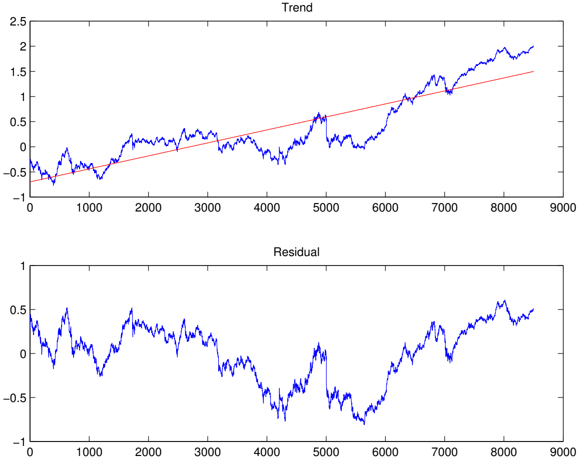
De-meaning

- *de-meaning* a time series means subtracting its mean: $\tilde{x} = x - \mathbf{avg}(x)$
- $\mathbf{rms}(\tilde{x}) = \mathbf{std}(x)$
- this is the least-squares fit with a constant

Straight-line fit and de-trending

- fit data $(1, x_1), \ldots, (T, x_T)$ with affine model $x_t \approx a + bt$ (also called *straight-line fit*)
- b is called the *trend*
- $a + bt$ is called the *trend line*
- *de-trending* a time series means subtracting its straight-line fit
- de-trended time series shows variations above and below the straight-line fit

Straight-line fit on Apple log price



Periodic time series

- let P-vector z be one period of periodic time series

$$x^{\text{per}} = (z, z, \ldots, z)$$

  (we assume T is a multiple of P)
- express as $x^{\text{per}} = Az$ with

$$A = \begin{bmatrix} I_P \\ \vdots \\ I_P \end{bmatrix}$$

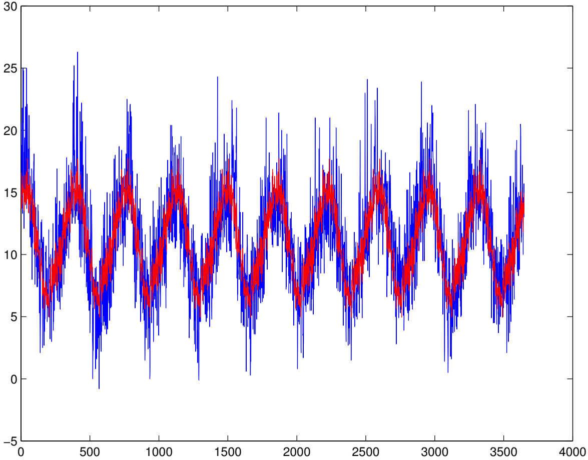
Extracting a periodic component

- given (non-periodic) time series x, choose z to minimize $\|x - Az\|^2$
- gives best least-squares fit with periodic time series
- simple solution: average periods of original:

$$\hat{z} = (1/k)A^T x, \quad k = T/P$$

- *e.g.*, to get $\hat{z}$ for January 9, average all x_i's with date January 9

Periodic component of Melbourne temperature



Extracting a periodic component with smoothing

- can add *smoothing* to periodic fit by minimizing

$$\|x - Az\|^2 + \lambda\|Dz\|^2$$

- $\lambda > 0$ is smoothing parameter
- D is $P \times P$ *circular difference matrix*

$$D = \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & & -1 & 1 \\ 1 & & & & & -1 \end{bmatrix}$$

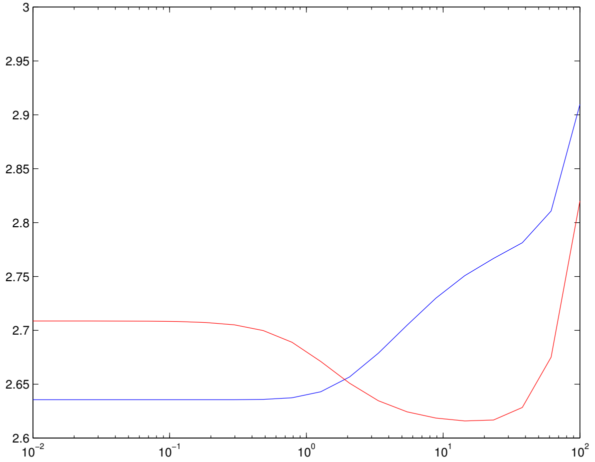
- λ is chosen visually or by validation

Choosing smoothing via validation

- split data into train and test sets, *e.g.*, test set is last period (P entries)
- train model on train set, and test on the test set
- choose λ to (approximately) minimize error on the test set

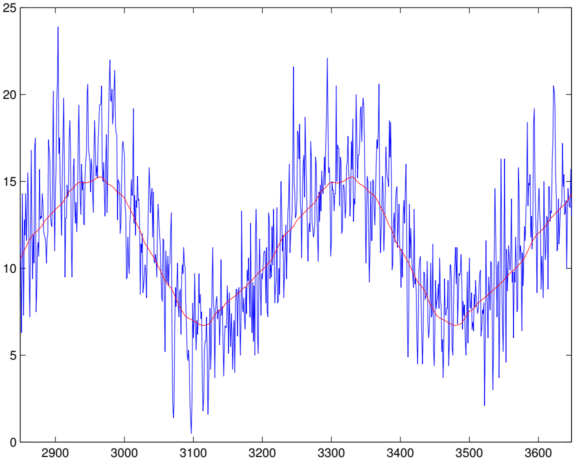
Validation of smoothing for Melbourne temperature

trained on first 8 years; tested on last two years



Periodic component of temperature with smoothing

- zoomed on test set, using $\lambda = 30$



Outline

Prediction

- goal: *predict* or *guess* x_{t+K} given $x_1, \ldots, x_t$
- $K = 1$ is *one-step-ahead prediction*
- prediction is often denoted $\hat{x}_{t+K}$, or more explicitly $\hat{x}_{(t+K|t)}$ (estimate of x_{t+K} at time t)
- $\hat{x}_{t+K} - x_{t+K}$ is *prediction error*
- applications: predict
  - asset price
  - product demand
  - electricity usage
  - economic activity
  - position of vehicle

Some simple predictors

- constant: $\hat{x}_{t+K} = a$
- current value: $\hat{x}_{t+K} = x_t$
- linear (affine) extrapolation from last two values:

$$\hat{x}_{t+K} = x_t + K(x_t - x_{t-1})$$

- average to date: $\hat{x}_{t+K} = \mathbf{avg}(x_{1:t})$
- $(M+1)$-period rolling average: $\hat{x}_{t+K} = \mathbf{avg}(x_{(t-M):t})$
- straight-line fit to date (*i.e.*, based on $x_{1:t}$)

Auto-regressive predictor

- *auto-regressive* predictor:

$$\hat{x}_{t+K} = (x_t, x_{t-1}, \ldots, x_{t-M})^T \beta$$

  - M is *memory length*
  - $(M+1)$-vector β gives predictor weights
  - can add offset v to $\hat{x}_{t+K}$
- prediction $\hat{x}_{t+K}$ is linear function of *past window* $x_{t-M:t}$
- (which of the simple predictors above have this form?)

Least squares fitting of auto-regressive models

- choose coefficients β via least squares (regression)
- regressors are $(M+1)$-vectors

$$x_{1:(M+1)}, \ldots, x_{(N-M):N}$$

- outcomes are numbers

$$\hat{x}_{M+K+1}, \ldots, \hat{x}_{N+K}$$

- can add regularization on β

Evaluating predictions with validation

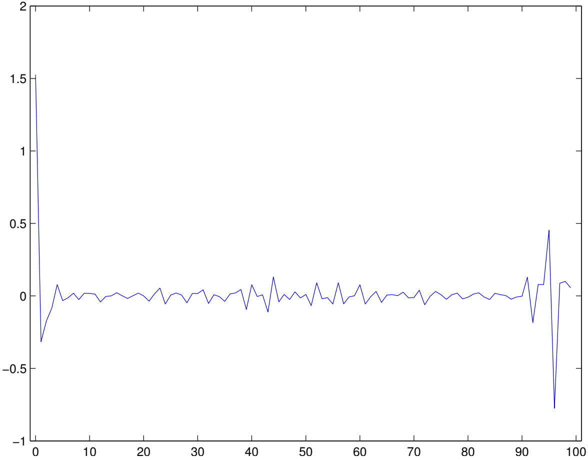
- for simple methods: evaluate RMS prediction error
- for more sophisticated methods:
  - split data into a training set and a test set (usually sequential)
  - train prediction on training data
  - test on test data

Example

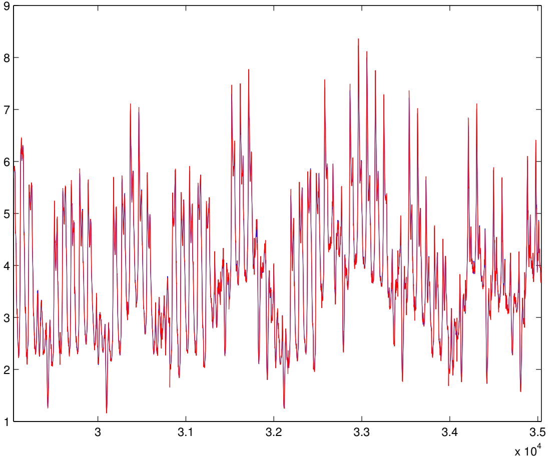
- predict Texas energy usage one step ahead ($K = 1$)
- train on first 10 months, test on last 2

Coefficients

- using $M = 100$
- 0 is the coefficient for today

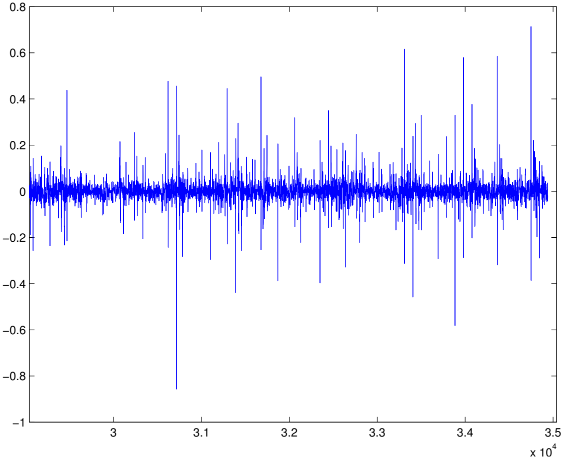


Auto-regressive prediction results



Auto-regressive prediction results

showing the residual



Auto-regressive prediction results

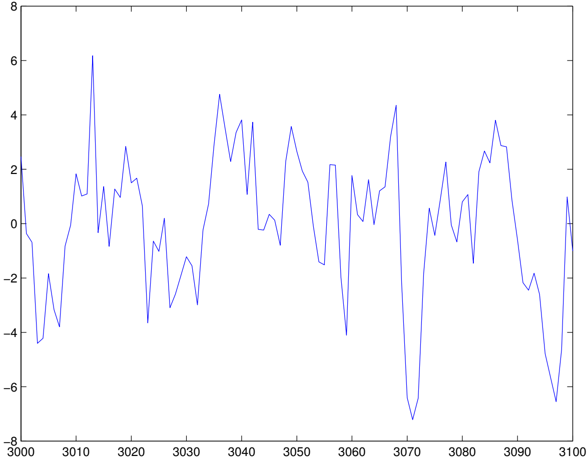
predictor	RMS error
average (constant)	1.20
current value	0.119
auto-regressive $(M = 10)$	0.073
auto-regressive $(M = 100)$	0.051

Autoregressive model on residuals

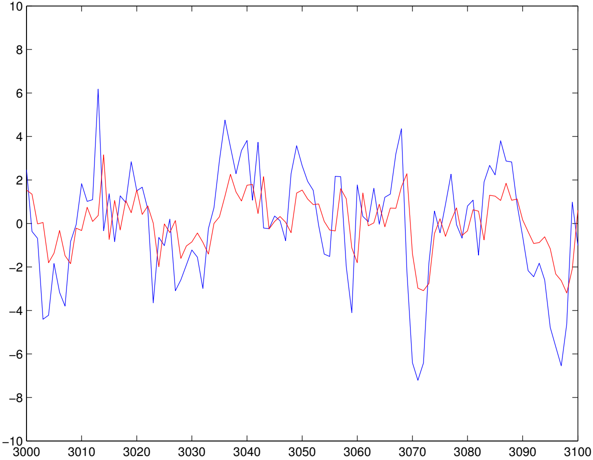
- fit a model to the time series, *e.g.*, linear or periodic
- subtract this model from the original signal to compute residuals
- apply auto-regressive model to predict residuals
- can add predicted residuals back to model to obtain predictions

Example

- Melbourne temperature data residuals
- zoomed on 100 days in test set



Auto-regressive prediction of residuals



Prediction results for Melbourne temperature

- tested on last two years

predictor	RMS error
average	4.12
current value	2.57
periodic (no smoothing)	2.71
periodic (smoothing, $\lambda = 30$)	2.62
auto-regressive $(M = 3)$	2.44
auto-regressive $(M = 20)$	2.27
auto-regressive on residual $(M = 20)$	2.22