Time Series Outlier Detection

Tingyi Zhu

July 28, 2016

Outline

- Time Series Basics
- Outliers Detection in Single Time Series
- Outlier Series Detection from Multiple Time Series
- Demos

Time Series Basics

First-order Autoregression

A model denoted as AR(1), in which the value of X at time t is a linear function of the value of X at time $t-1$:

$$X_t = \phi X_{t-1} + \varepsilon_t \tag{1}$$

Assumptions:

- $\varepsilon_t \overset{i.i.d}{\sim} N(0, \sigma)$, stochastic term.
- ε_t is independent of X_t.

General Autoregressive Model

AR(p):

$$
\begin{aligned}
X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t \\
&= \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t \\
&= \sum_{i=1}^{p} \phi_i B^i X_t + \varepsilon_t
\end{aligned}
$$

where we use the backshift operator B ($BX_t = X_{t-1}$, $B^k X_t = X_{t-k}$).

Alternative notation:

$$\phi(B) X_t = \varepsilon_t$$

$\phi(B)$ is a polynomial of B,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p = 1 - \sum_{i=1}^{p} \phi_i B^i$$

Moving Average

- Another approach for modeling univariate time series
- X_t depends linearly on its own current and previous stochastic terms
- MA(1):

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

- MA(q):

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

- $\theta_1, \ldots, \theta_q$: parameters of MA model
- $\varepsilon_t, \ldots, \varepsilon_{t-q}$: stochastic terms
- Using backshift operator B, model simplified as

$$
\begin{aligned}
X_t &= (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t \\
&= (1 + \sum_{i=1}^{q} \theta_i B^i)\varepsilon_t \\
&= \theta(B)\varepsilon_t
\end{aligned}
$$

ARMA Model

- A model consists of both autoregressive (AR) part and moving average (MA) part:

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{2}$$

  referred to as the ARMA(p,q) model.

  p: the order of the autoregressive part

  q: the order of the moving average part

- More concisely, using backshift operator B, (2) becomes:

$$\phi(B) X_t = \theta(B) \varepsilon_t$$

Stationarity of Time Series

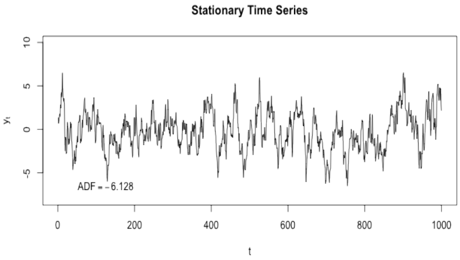
- In short, a time series is stationary if its statistical properties are all constant over time.
- To mention some properties:

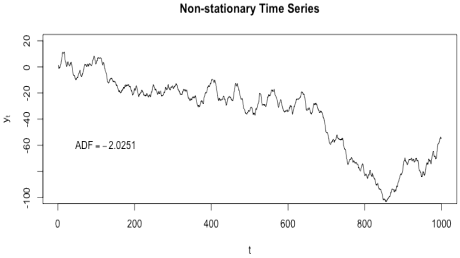
  Mean: $E[X_t] = E[X_s]$ for any $t, s \in \mathbb{Z}$,

  Variance: $Var[X_t] = Var[X_s]$ for any $t, s \in \mathbb{Z}$,

  Joint distribution:

  $$Cov(X_t, X_{t+1}) = Cov(X_s, X_{s+1}) \text{ for any } t, s \in \mathbb{Z}.$$


Stationary Time Series
ADF = −6.128
y_t
t
0
200
400
600
800
1000
10
5
0
-5


Non-stationary Time Series
ADF = −2.0251
y_t
t
0
200
400
600
800
1000
20
0
-20
-40
-60
-100

Requirements for a Stationary Time Series

- AR(1) $X_t = \phi X_{t-1} + \varepsilon_t$: $\quad |\phi| < 1$
- AR(p) $\phi(B)X_t = \varepsilon_t$:

  All the roots of $\phi(z) = 0$ are outside unit circle.
- MA models are always stationary
- ARMA(p,q) $\phi(B)X_t = \theta(B)\varepsilon_t$:

  All the roots of $\phi(z) = 0$ are outside unit circle.

Non-stationary time series

- Trend effect
- Seasonal effect

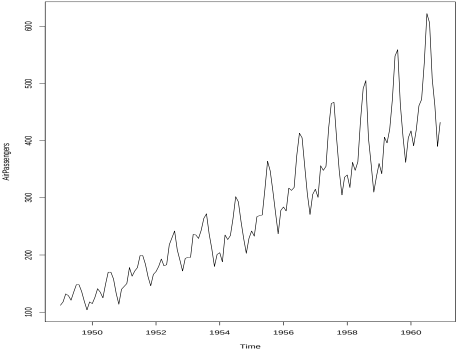


Figure: Monthly totals of international airline passengers, 1949 to 1960.

Time Series Decomposition

- Think of a more general time series formulation including both trend and seasonal effect:

$$X_t = T_t + S_t + E_t \tag{3}$$

  - X_t is data point at time t
  - T_t is the trend component at time t
  - S_t is the seasonal component at time t
  - E_t is the remainder component at time t (containing AR and MA terms)

Series with Trend, examples:

- Assuming no seasonal effect, i.e. $S_t = 0$
- Linear trend:

$$X_t = 2t + 0.5X_{t-1} + \varepsilon_t$$

- Quadratic trend:

$$X_t = 2t + t^2 + 0.5X_{t-1} + \varepsilon_t$$

- Goal: remove the trend, to transform the series to be stationary
- Solution: lag-1 differencing

Differencing and Trend

Define the lag-1 difference operator,

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t,$$

where B is the backshift operator.

- If $X_t = \beta_0 + \beta_1 t + E_t$, then

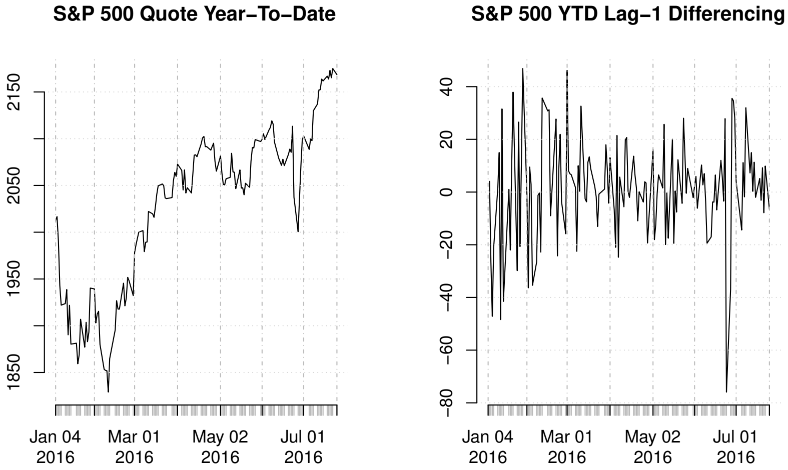
$$\nabla X_t = \beta_1 + \nabla E_t.$$

- If $X_t = \sum_{i=0}^{k} \beta_i t^i + E_t$, then

$$\nabla^k X_t = (1 - B)^k X_t = k!\beta_k + \nabla^k E_t.$$

we call ∇^k kth lag-1 difference operator.

Lag-1 Differencing



Series with Seasonal Effect, example:

- For quarterly data, with possible seasonal (quarterly) effects, we can define indicator function S_j. For $j = 1, 2, 3, 4$,

$$S_j = \begin{cases} 1 & \textit{if observation is in quarter } j \textit{ of a year}, \\ 0 & \textit{otherwise}. \end{cases}$$

- A model with seasonal effects could be written as

$$X_t = \alpha_1 S_1 + \alpha_2 S_2 + \alpha_3 S_3 + \alpha_4 S_4 + \varepsilon_t$$

- Goal: remove the seasonal effects
- Solution: lag-s differencing, where s is the number of seasons

Differencing and Seasonal Effects

Define the lag-s difference operator,

$$\nabla_s X_t = X_t - X_{t-s} = (1 - B^s)X_t,$$

where B is the backshift operator.

If $X_t = T_t + S_t + E_t$, and S_t has period s (i.e. $S_t = S_{t-s}$ for all t), then

$$\nabla_s X_t = (1 - B^s)X_t = T_t - T_{t-s} + \nabla_s E_t.$$

Non-seasonal ARIMA

- $S_t = 0$
- ARIMA stands for Auto-Regressive Integrated Moving Average, ARMA integrated with differencing.
- A nonseasonal ARIMA model is classified as ARIMA(p,d,q), where
  - p is the order of AR terms,
  - d is the number of nonseasonal differences needed for stationarity,
  - q is the order of MA terms.

Non-seasonal ARIMA, Cont.

- Recall ARMA(p,q):

$$\phi(B)X_t = \theta(B)\varepsilon_t,$$

  - $\phi(B)$ and $\theta(B)$ are polynomials of B of order p and q.
  - Stationary requirement: all roots of $\phi(z) = 0$ outside unit circle.

- ARIMA(p,d,q):

$$\phi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t,$$

  - X_t is not stationary. Why?
  - $Z_t = (1 - B)^d X_t$ is ARMA(p,q), is stationary.

Seasonal ARIMA

- A seasonal ARIMA model is classified as

$$ARIMA(p, d, q) \times (P, D, Q)_m$$

  - p is the order of AR terms,
  - d is the number of nonseasonal differences,
  - q is the order of MA terms.
  - P is the order of seasonal AR terms,
  - D is the number of seasonal differences,
  - Q is the order of seasonal MA terms.
  - m is the number of seasons.

Example: $ARIMA(1,1,1) \times (1,1,1)_4$

$$\underset{\text{Non-seasonal AR(1)}}{(1-\phi_1 B)}\;\underset{\text{Seasonal AR(1)}}{(1-\Phi_1 B^4)}\;\underset{\text{Non-seasonal difference}}{(1-B)}\;\underset{\text{Seasonal difference}}{(1-B^4)}y_t = \underset{\text{Non-seasonal MA(1)}}{(1+\theta_1 B)}\;\underset{\text{Seasonal MA(1)}}{(1+\Theta_1 B^4)}e_t.$$

General ARIMA

- The ARIMA model can be generalized as follow:

$$\phi(B)\alpha(B)X_t = \theta(B)\varepsilon_t,$$

  - $\phi(B)$: autoregressive polynomial, all roots outside unit circle
  - $\alpha(B)$: differencing filter renders the data stationary, all roots on the unit circle
  - $\theta(B)$: moving average polynomial, all roots outside unit circle (to assure $\theta(B)$ is invertible.

- Alternatively,

$$X_t = \frac{\theta(B)}{\phi(B)\alpha(B)}\varepsilon_t.$$

Outliers Detection in Single Time Series

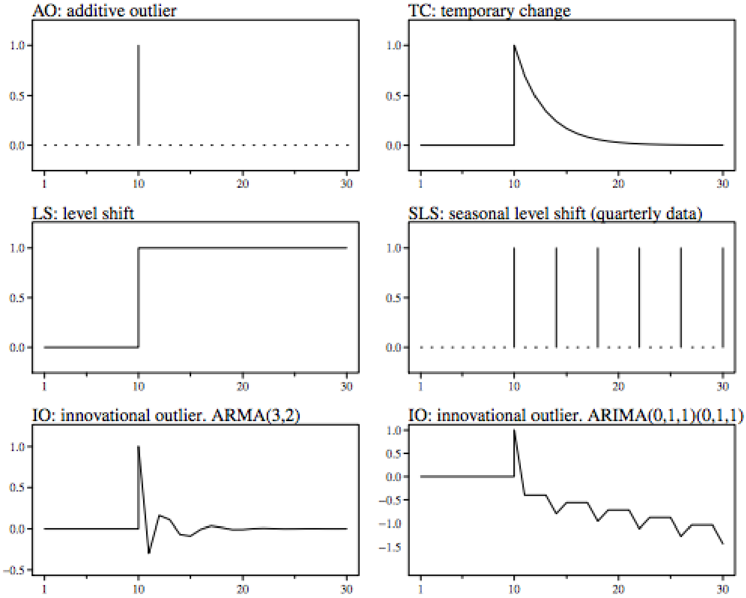
Automatic Detection Procedure

- Described in Chung Chen, Lon-Mu Liu. Joint Estimation of Model Parameters and Outlier Effects in Time Series,*JASA*, 1993
- Based on the framework of ARIMA models
- R package `tsoutlier` written by YAHOO in 2014

Types of Outliers

- General representation: $L(B)I_t(t_j)$
  - $L(B)$: a polynomial of lag operator B
  - $I_t(t_j) = 1$ there's outlier at time $t = t_j$, and 0 otherwise.
- Types of outliers:
  - Additive Outliers (AO): $L(B) = 1$;
  - Level Shift (LS): $L(B) = \frac{1}{1-B}$;
  - Temporary Change (TC): $L(B) = \frac{1}{1-\delta B}$;
  - Seasonal Level Shift (SLS): $L(B) = \frac{1}{1-B^s}$;
  - Innovational Outliers (IO): $L(B) = \frac{\theta(B)}{\phi(B)\alpha(B)}$.

Types of Outliers



Formulation

- ARIMA model:

$$X_t = \frac{\theta(B)}{\phi(B)\alpha(B)}\varepsilon_t.$$

- Model with outliers at time $t_1, t_2, \ldots, t_m$:

$$X_t^* = \sum_{j=1}^{m} \omega_j L_j(B) I_t(t_j) + \frac{\theta(B)}{\phi(B)\alpha(B)}\varepsilon_t.$$

  - $L_j(B)$ depends on pattern of the jth outlier
  - $I_t(t_j) = 1$ there's outlier at time $t = t_j$, and 0 otherwise.
  - ω_j denotes the magnitude of the jth outlier effect

Effect of One Outlier

- Assume the time series parameters are known, we examine the effect of one outlier:

$$X_t^* = \omega L(B) I_t(t_1) + \frac{\theta(B)}{\phi(B)\alpha(B)}\varepsilon_t$$

- Define polynomial $\pi(B)$ as:

$$\pi(B) = \frac{\phi(B)\alpha(B)}{\theta(B)} = 1 - \pi_1 B - \pi_2 B - \cdots ,$$

- Contaminated by the outlier, the estimated residual $\hat{e}_t$ becomes

$$\hat{e}_t = \pi(B) X_t^*$$

(Without outlier, $\hat{e}_t = \pi(B) X_t$.)

For the four types of outliers,

- IO: $\hat{e}_t = \omega I_t(t_1) + \varepsilon_t$,
- AO: $\hat{e}_t = \omega \pi(B) I_t(t_1) + \varepsilon_t$,
- LS: $\hat{e}_t = \omega \frac{\pi(B)}{1-B} I_t(t_1) + \varepsilon_t$,
- TC: $\hat{e}_t = \omega \frac{\pi(B)}{1-\delta B} I_t(t_1) + \varepsilon_t$.

Alternatively,

$$\hat{e}_t = \omega x_{i,t} + \varepsilon_t, \quad t = t_1, t_1 + 1, \ldots \ and \ i = 1, 2, 3, 4$$

- $x_{i,t} = 0$ for all i and $t < t_1$,
- $x_{i,t} = 1$ for all i,
- $x_{1,t_1+k} = 0, \qquad x_{2,t_1+k} = -\pi_k,$
  $x_{3,t_1+k} = 1 - \sum_{j=1}^{k} \pi_j, \qquad x_{4,t_1+k} = \delta^k - \sum_{j=1}^{k-1} \delta^{k-j} \pi_j - \pi_k.$

A simple linear regression!

Estimate of ω

The least square estimate doe the effect of a single outlier at $t = t_1$ can be expressed as

$$\hat{\omega}_{\mathrm{IO}}(t_1) = \hat{e}_{t_1}$$

$$\hat{\omega}_{\mathrm{AO}}(t_1) = \frac{\sum_{t=t_1}^{n} \hat{e}_t x_{2t}}{\sum_{t=t_1}^{n} x_{2t}^2}$$

$$\hat{\omega}_{\mathrm{LS}}(t_1) = \frac{\sum_{t=t_1}^{n} \hat{e}_t x_{3t}}{\sum_{t=t_1}^{n} x_{3t}^2}$$

$$\hat{\omega}_{\mathrm{TC}}(t_1) = \frac{\sum_{t=t_1}^{n} \hat{e}_t x_{4t}}{\sum_{t=t_1}^{n} x_{4t}^2} .$$

Test Statistics τ

From regression analysis, we have

$$\frac{\hat{\omega} - \omega}{\hat{\sigma}_a} (\sum_{t=t_1}^{n} x_{i,t}^2)^{1/2} \sim N(0,1),$$

where $\hat{\sigma}_a$ is the estimation of residual standard deviation.

We want to test whether $\omega = 0$, then the following statistics are approximately $N(0,1)$:

$$\hat{\tau}_{\mathrm{IO}}(t_1) = \hat{\omega}_{\mathrm{IO}}(t_1)/\hat{\sigma}_a$$

$$\hat{\tau}_{\mathrm{AO}}(t_1) = \{\hat{\omega}_{\mathrm{AO}}(t_1)/\hat{\sigma}_a\} \left(\sum_{t=t_1}^{n} x_{2t}^2 \right)^{1/2}$$

$$\hat{\tau}_{\mathrm{LS}}(t_1) = \{\hat{\omega}_{\mathrm{LS}}(t_1)/\hat{\sigma}_a\} \left(\sum_{t=t_1}^{n} x_{3t}^2 \right)^{1/2}$$

$$\hat{\tau}_{\mathrm{TC}}(t_1) = \{\hat{\omega}_{\mathrm{TC}}(t_1)/\hat{\sigma}_a\} \left(\sum_{t=t_1}^{n} x_{4t}^2 \right)^{1/2}.$$

Procedure in the Presence of Multiple Ouliers

In the presence of multiple outliers, recall the model

$$X_t^* = \sum_{j=1}^{m} \omega_j L_j(B) I_t(t_j) + \frac{\theta(B)}{\phi(B)\alpha(B)} \varepsilon_t.$$

where $\hat{\sigma}_a$ is the estimation of residual standard deviation.

The estimated residual becomes

$$\hat{e}_t = \sum_{j=1}^{m} \omega_j \pi(B) L_j(B) I_t(t_j) + \varepsilon_t$$

Stage 1: Joint Estimation of Outlier Effect and Model Parameters

- Fitting the series by an ARIMA model (`forecast` package in R), obtain initial parameter $(\phi(B), \theta(B), \alpha(B))$ estimation of the model.
- Detect outliers one by one sequentially

I.2. For $t = 1, \ldots, n$, compute $\hat{\tau}_{\mathrm{IO}}(t)$, $\hat{\tau}_{\mathrm{AO}}(t)$, $\hat{\tau}_{\mathrm{LS}}(t)$, and $\hat{\tau}_{\mathrm{TC}}(t)$ in (14) using the residuals obtained from I.1, and let $\eta_t = \max\{|\hat{\tau}_{\mathrm{IO}}(t)|, |\hat{\tau}_{\mathrm{AO}}(t)|, |\hat{\tau}_{\mathrm{LS}}(t)|, |\hat{\tau}_{\mathrm{TC}}(t)|\}$. If $\max_t \eta_t = |\hat{\tau}_{tp}(t_1)| > C$, where C is predetermined critical value, then there is a possibility of a type tp outlier at t_1; tp may be IO, AO, LS, or TC.

Stage 2: Initial Parameter Estimation and Outlier Detection

II.1. Suppose that m time points $t_1, t_2, \ldots, t_m$ are identified as possible outliers of various types. The outlier effects ω_j's can be estimated jointly using the multiple regression model described in (20), where $\{\hat{e}_t\}$ is regarded as the output variable and $\{L_j(B)I_t(t_j)\}$ are the input variables.

II.2. Compute the $\hat{\tau}$ statistics of the estimated ω_j's, where $\hat{\tau}_j = \hat{\omega}_j/\mathrm{std}(\hat{\omega}_j)$, $j = 1, \ldots, m$. If $\min_j |\hat{\tau}_j| = \hat{\tau}_\nu \leq C$, where C is the same critical value used in step I.2, then delete the outlier at time point t_ν from the set of the identified outliers and go to step II.1 with the remaining $m - 1$ outliers. Otherwise, go to step II.3.

II.3. Obtain the adjusted series by removing the outlier effects, using the most recent estimates of ω_j's at step II.1. In other words, remove only the outlier effects that are significant based on the iterations of steps II.1 and II.2.

II.4. Compute the maximum likelihood estimates of the model parameters based on the adjusted series obtained at step II.3. If the relative change of the residual standard error from the previous estimate is greater than ε, go to step II.1 for further iterations;

Outlier Series Detection from Multiple Time Series

Detect Anomalous Series

- Goal: efficiently find the least similar time series in a large set
- Motivation: Internet companies monitoring the servers(CPU, Memory), find unusual behaviors

Detect Anomalous Series

- Described in Rob J Hyndman et al. Large-Scale Unusual Time Series Detection, *ICDM*, 2015
- Approach: Extract features from time series, PCA
- R package `anomalous`
- Test on real data from YAHOO email server,
  80% accuracy compared to 40% from previous methods

Step 1: Extract Features from Time Series

- 15 features selected, each captures the global information of time series

Feature	Description
Mean	Mean.
Var	Variance.
ACF1	First order of autocorrelation.
Trend	Strength of trend.
Linearity	Strength of linearity.
Curvature	Strength of curvature
Season	Strength of seasonality.
Peak	Strength of peaks.
Trough	Strength of trough.
Entropy	Spectral entropy.
Lumpiness	Changing variance in remainder.
Spikiness	Strength of spikiness
Lshift	Level shift using rolling window.
Vchange	Variance change.
Fspots	Flat spots using disretization.
Cpoints	The number of crossing points.
KLscore	Kullback-Leibler score.
Change.idx	Index of the maximum KL score.

Table 1: Summary of features used for detecting unusual time series.

- Step2: PCA to reduce dimension
  - dim=15 initially, correlation existing between features
  - The first 2 PCs are sufficient, capturing most of the variance
- Step 3: Implement multi-dimentional outlier detection algorithm to find outlier series
  - Density based
  - α-hull

Demo