III Test for randomness of a time series

Turning point test

This is a non-parametric test procedure for testing randomness of a time series

Null hypothesis

H_0 : Series is purely random (does not contain any deterministic component)

against the alternate hypothesis

H_A : not H_0 (i.e the series is not purely random)

A turning point is defined as either a 'peak' when a value is greater than its 2 neighboring values or a 'trough' when a value is less than its 2 neighboring values

i.e. y_i is a turning point if

$$y_i > y_{i-1} \ \& \ y_i > y_{i+1} \quad - \text{peak}$$

$$\text{or} \quad y_i < y_{i-1} \ \& \ y_i < y_{i+1} \quad - \text{trough}$$

All $n-2$ time points $(2, 3, \ldots, n-1)$ are checked for being declared as a turning point

Define

$$U_i = \begin{cases} 1, & \text{If } y_i \text{ is a turning pt} \\ 0, & \text{o/w} \end{cases}$$

$$P = \sum_{i=2}^{n-1} U_i \quad : \text{total \# of turning points}$$

To see the expected value of U_i and hence the expected value of P when the series is purely random, let us consider 3 values (y_{i-1}, y_i, y_{i+1}) leading to one such U_i

Let $(y_{(1)}, y_{(2)}, y_{(3)})$ denote the ordered values derived from (y_{i-1}, y_i, y_{i+1}). $\begin{pmatrix} y_{(1)} : \text{smallest} \\ y_{(3)} : \text{largest} \end{pmatrix}$

Now (y_{i-1}, y_i, y_{i+1}) can be any of the following 6 possible orders

$(y_{(1)}, y_{(2)}, y_{(3)})$, $(y_{(1)}, y_{(3)}, y_{(2)})$, $(y_{(2)}, y_{(1)}, y_{(3)})$, $(y_{(2)}, y_{(3)}, y_{(1)})$, $(y_{(3)}, y_{(1)}, y_{(2)})$ & $(y_{(3)}, y_{(2)}, y_{(1)})$.

Under the assumption that the series is purely random (i.e. H_0), all 6 possible outcomes are equally likely

$\Rightarrow$ in such a case $E(U_i) = 1 \times \frac{4}{6} + 0 \times \frac{2}{6} = \frac{2}{3}$

(Note that 4 out of the above 6 possible outcomes have turning points) under the null hypothesis that the series is purely random

$$E(P) = \frac{2}{3}(n-2) \quad \& \quad V(P) = \frac{16n-29}{90}$$

(Ref: Kendall, Stuart & Ord; Adv Th of Statistics)

Asymptotic test for H_0 is based on the statistic

$$Z = \frac{P - E(P)}{\sqrt{V(P)}} = \frac{P - \frac{2}{3}(n-2)}{\sqrt{\frac{16n-29}{90}}}$$

$$Z \overset{asym}{\sim} N(0,1) \text{ under } H_0.$$

We would reject H_0 at level of significance α

If observed $|Z| > \tau_{\alpha/2}$

($\tau_{\alpha/2}$: upper $\alpha/2$ cutoff point of $N(0,1)$)

Estimation / elimination of trend & seasonality

(I) Estimation of trend in the absence of seasonality

Consider a time series model

$$y_t = m_t + e_t; \quad e_t \text{ is } \ni \; E(e_t) = 0$$

$$Cov(e_t, e_s) = \begin{cases} \sigma^2, & t = s \\ 0, & o/w \end{cases}$$

Method 1 : Least squares estimation of m_t

We assume that trend is polynomial trend of a particular order

e.g. $m_t = a_0 + a_1 t$ — linear time trend

$m_t = a_0 + a_1 t + a_2 t^2$ — quadratic time trend

⋮

$m_t = a_0 + a_1 t + \cdots + a_k t^k$; k^{th} order polynomial trend

We obtain estimates of $(a_0, a_1, \ldots, a_k)$ by minimizing the f^n

$$g(a_0, a_1, \ldots, a_k) = \sum_{t=1}^{n} \Big(y_t - \sum_{i=0}^{k} a_i t^i\Big)^2$$

$$\hat{\underset{\sim}{a}}_{LS} = \underset{\underset{\sim}{a}}{\arg\min} \sum_{t=1}^{n} \Big(y_t - \sum_{i=0}^{k} a_i t^i\Big)^2$$

Note that the above is a simple linear model LS estimation problem and hence

$$\hat{\underset{\sim}{a}}_{LS} = (X'X)^{-1} X' \underset{\sim}{y}$$

with the model written as

$$\underset{\sim}{Y} = X \underset{\sim}{\beta} + \underset{\sim}{e}$$

$$\underset{\sim}{Y} = (Y_1, \ldots, Y_n)' ; \underset{\sim}{e} = (e_1, \ldots, e_n)'$$

$$\underset{\sim}{\beta} = \underset{\sim}{a} = (a_0, a_1, \ldots, a_K)'$$

$$\underset{n \times \overline{K+1}}{X} = \begin{pmatrix} 1 & 1 & 1^2 & \cdots & 1^K \\ 1 & 2 & 2^2 & \cdots & 2^K \\ \vdots & & & & \\ 1 & n & n^2 & \cdots & n^K \end{pmatrix}$$

$$\underset{\sim}{\hat{a}}_{LS} = \underset{\underset{\sim}{a}}{\arg\min} \, (\underset{\sim}{Y} - X\underset{\sim}{\beta})'(\underset{\sim}{Y} - X\underset{\sim}{\beta})$$

& $\underset{\sim}{\hat{a}}_{LS} = (X'X)^{-1} X' \underset{\sim}{Y}$

<u>Remark</u>: We can arrive at the appropriate order trend in the following manner

(i) estimate linear trend

$$\hat{m}_t^{(1)} = \hat{a}_0 + \hat{a}_1 t$$

(ii) detrend the data

$$\tilde{Y}_{1_t} = Y_t - \hat{m}_t^{(1)}$$

(iii) apply relative ordering test on $\{\tilde{Y}_{1t}\}$ values; If you observe that H_0 of no trend is rejected fit

$$\hat{m}_t^{(2)} = \hat{a}_0 + \hat{a}_1 t + \hat{a}_2 t^2$$

(iv) detrend as $y_t - \hat{m}_t^{(2)}$ and apply relative ordering test to check existence of trend in the detrended series

(v) continue till the detrended series show no significant trend.

Method 2: Trend estimation using moving average

Data: $(y_1, \ldots, y_n)$

Let q be a non-negative integer

Moving average trend estimate at time point t is given by

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} y_{t+j} ; \quad q+1 \le t \le n-q$$

$\hat{m}_t$: moving average trend estimate with a window length of $2q+1$

The observations within the moving average window

$$\left(y_{t-q}, \ldots y_{t-1}, \underset{\uparrow}{y_t}, y_{t+1}, \ldots ; y_{t+q} \right)$$

Note : The above is referred to as equal weighted moving average with weights as $\frac{1}{2q+1}$

Note : We do not get trend values at end points i.e no trend values $\forall t < q+1$ & $\forall t > n-q$

It is so as we do not have y_t for $t<1$ & $t>n$

In such cases, we can use a symmetric padding or end point padding to get rough estimates of trend.

Note: For even order window length moving average, a simple mean of adjacent trend values is computed so as to have trend value correspond to time points. e.g. a 4pt ma

y_1

y_2

y_3 $\quad (y_1+y_2+y_3+y_4)/4, \quad (y_2+y_3+y_4+y_5)/4 \rightarrow \hat{m}_3 = \frac{1}{2}\left\{\frac{y_1+y_2+y_3+y_4}{4} + \frac{y_2+y_3+y_4+y_5}{4}\right\}$,

y_4

y_5

y_6

$\vdots$

In general, for even order window length

$$\hat{m}_t = \frac{1}{2q}\left(\frac{1}{2}y_{t-q} + y_{t-q+1} + \cdots + y_t + \cdots + y_{t+q-1} + \frac{1}{2}y_{t+q}\right)$$

$$q+1 \leq t \leq n-q$$