CHANGE DETECTION IN TIME SERIES

Edit Gombay

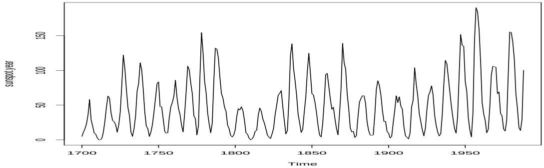
TIES - 2008
University of British Columbia, Kelowna

June 8 - 13, 2008

Outline

- Introduction
- Results
- Examples
- References

Introduction



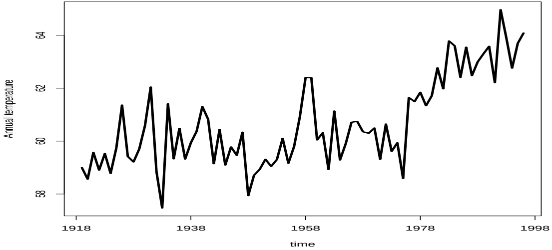
Wölfer sunspot numbers 1700 and 1998, $n = 289$.

Consider the Wölfer sunspot numbers. As an example, statistical package IMSL fits an ARMA(2,1) model given by

$$X_k = \mu + Z_k, Z_k = \phi_1 Z_{k-1} + \phi_2 Z_{k-2} + \theta_1 \epsilon_{k-1} + \epsilon_k$$

to the yearly Wölfer sunspot numbers for years 1749–1924 (n=176), and their method of moments estimators for this sequence are $\hat{\phi}_1 = 1.24, \hat{\phi}_2 = -.58$ and $\hat{\theta}_1 = -.12$.

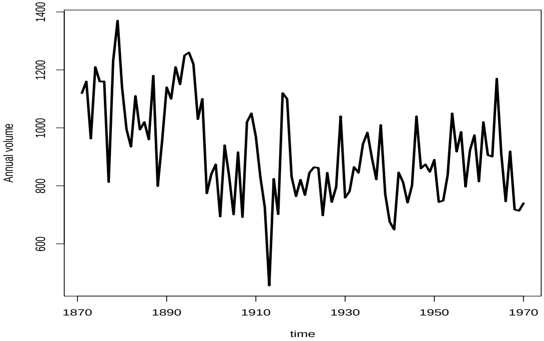
Introduction



Chula Vista, California annual average temperatures 1919 - 1996, $n = 78$.

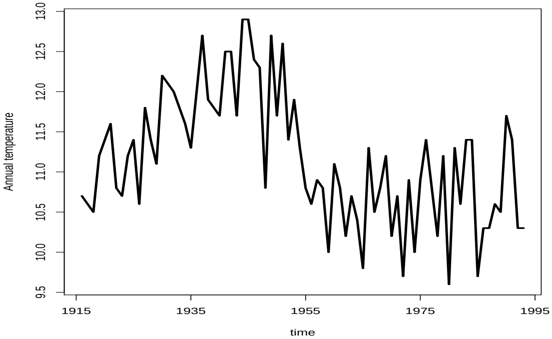
Lund and Reeves (2002) examined the data on annual mean temperatures at Chula Vista, California during 1919-1996, inclusive ($n = 78$). They used F-statistics in their test for change in linear regression, and assumed independence in the error process.

Introduction



Nile annual discharge at Aswan. 1871-1970, n=100.

Introduction



Collegeville,NS, annual temperatures. 1916-1995, n=80.

The annual mean maximum temperatures of Collegeville, Nova Scotia, Canada, (1916-1995) were tested for inhomogeneity by Vincent (1998), $n = 80$.

Introduction

- Consider the model

$$X_k = \mu + \sum_{0 \leq i < \infty} a_i \epsilon_{k-i}. \qquad (1)$$

Assume that

$$\epsilon_i, \ -\infty < i < \infty,$$

are independent, identically distributed r.v.

$$E\epsilon_i = 0, \ 0 < \sigma^2 = E\epsilon_i^2 < \infty, \ E|\epsilon_0|^\kappa < \infty \quad (2)$$

with some $\kappa > 2$.

- Instability in the value of any parameter will lead to wrong forecasts and data analysis, so detecting change in them is a statistical problem of great importance. Once change has been detected, the time of change has to be estimated.

Introduction

- Consider the model

$$X_k = \mu + \sum_{0 \le i < \infty} a_i \epsilon_{k-i}. \qquad (1)$$

  Assume that

$$\epsilon_i, \ -\infty < i < \infty,$$

  are independent, identically distributed r.v.

$$E\epsilon_i = 0, \ 0 < \sigma^2 = E\epsilon_i^2 < \infty, \ E|\epsilon_0|^\kappa < \infty \quad (2)$$

  with some $\kappa > 2$.

- Instability in the value of any parameter will lead to wrong forecasts and data analysis, so detecting change in them is a statistical problem of great importance. Once change has been detected, the time of change has to be estimated.

Introduction

- There is an extensive literature on detecting change in the parameters of autoregressive time series. (See e.g. Picard (1985), Davis et al. (1995), Gombay (2008), and references therein). The AR(p) model allows to write the likelihood function in a simple form, and from this the maximum likelihood estimators are readily derived.
- This is not the case for moving average models, or for the more general model of linear processes.
- There is a group of papers on how to detect multiple structural changes in linear processes using dynamic programming techniques. (e.g. Bai and Perron (1998), Davies et al. (2006)). In those papers the tests performance under the null hypothesis of no change is not known in most cases, and the critical values are obtained by simulation.

Introduction

- There is an extensive literature on detecting change in the parameters of autoregressive time series. (See e.g. Picard (1985), Davis et al. (1995), Gombay (2008), and references therein). The AR(p) model allows to write the likelihood function in a simple form, and from this the maximum likelihood estimators are readily derived.
- This is not the case for moving average models, or for the more general model of linear processes.
- There is a group of papers on how to detect multiple structural changes in linear processes using dynamic programming techniques. (e.g. Bai and Perron (1998), Davies et al. (2006)). In those papers the tests performance under the null hypothesis of no change is not known in most cases, and the critical values are obtained by simulation.

Introduction

- There is an extensive literature on detecting change in the parameters of autoregressive time series. (See e.g. Picard (1985), Davis et al. (1995), Gombay (2008), and references therein). The AR(p) model allows to write the likelihood function in a simple form, and from this the maximum likelihood estimators are readily derived.
- This is not the case for moving average models, or for the more general model of linear processes.
- There is a group of papers on how to detect multiple structural changes in linear processes using dynamic programming techniques. (e.g. Bai and Perron (1998), Davies et al. (2006)). In those papers the tests performance under the null hypothesis of no change is not known in most cases, and the critical values are obtained by simulation.

- Our approach is analytical, it belongs to the family of AMOC procedures, multiple changes can be treated by the usual segmentation technique.
- Detecting change in the auto-covariance structure for model (1) was considered by Lee et al. (2003). Their test statistic is based on a quadratic form constructed with the $m+1$ estimators of $\gamma(0), \ldots, \gamma(m)$, where $\gamma(r) = E\{(X_1 X_{1+r}) - [E(X_1]^2\}$ $r \geq 0$.
- In their algorithms the fourth moment of innovations $\{\epsilon_t\}$ has to be estimated. They suggest fitting a long $AR(p)$, $p = p(n)$, model to the observations, estimating the parameters of this model, then the residuals, and finally the fourth moment of the residuals.

- Our approach is analytical, it belongs to the family of AMOC procedures, multiple changes can be treated by the usual segmentation technique.
- Detecting change in the auto-covariance structure for model (1) was considered by Lee et al. (2003). Their test statistic is based on a quadratic form constructed with the $m+1$ estimators of $\gamma(0), \dots, \gamma(m)$, where $\gamma(r) = E\{(X_1 X_{1+r}) - [E(X_1]^2\}\ r \geq 0$.
- In their algorithms the fourth moment of innovations $\{\epsilon_t\}$ has to be estimated. They suggest fitting a long $AR(p)$, $p = p(n)$, model to the observations, estimating the parameters of this model, then the residuals, and finally the fourth moment of the residuals.

- Our approach is analytical, it belongs to the family of AMOC procedures, multiple changes can be treated by the usual segmentation technique.
- Detecting change in the auto-covariance structure for model (1) was considered by Lee et al. (2003). Their test statistic is based on a quadratic form constructed with the $m+1$ estimators of $\gamma(0), \ldots, \gamma(m)$, where $\gamma(r) = E\{(X_1 X_{1+r}) - [E(X_1]^2\} \; r \geq 0$.
- In their algorithms the fourth moment of innovations $\{\epsilon_t\}$ has to be estimated. They suggest fitting a long $AR(p)$, $p = p(n)$, model to the observations, estimating the parameters of this model, then the residuals, and finally the fourth moment of the residuals.

Introduction

Testing for change in the mean of the process uses CUSUM statistics

- $$M_n(t) = n^{-1/2}\Big(\sum_{1\le i\le (n+1)t} X_i - t\sum_{1\le i\le n} X_i\Big),\quad 0 \le t < 1$$

  in the case of independent observations.
- Now we can extend the use of the CUSUM statistics above to observation as in (1) by the virtue of the next theorems. (I will present a simplified form of the theorems as there is no time for both theory and applications in a short talk. This talk is meant to be one concentrating on applications.)

Introduction

Testing for change in the mean of the process uses CUSUM statistics

- $$M_n(t) = n^{-1/2}\Big(\sum_{1\le i\le (n+1)t} X_i - t\sum_{1\le i\le n} X_i\Big), \quad 0 \le t < 1$$

  in the case of independent observations.
- Now we can extend the use of the CUSUM statistics above to observation as in (1) by the virtue of the next theorems. (I will present a simplified form of the theorems as there is no time for both theory and applications in a short talk. This talk is meant to be one concentrating on applications.)

Introduction

Theorem

Assume the sequence of observations $\{X_i\}$ *satisfy relation (1) and (2) and*

$$\sum_{1 \leq j < \infty} j|a_j| < \infty, \quad \sum_{1 \leq j < \infty} a_j \neq 0.$$

Then with $\tau^2 = \sigma^2 (\sum_{1 \leq j < \infty} j|a_j|)^2 < \infty$

$$\tau^{-1} \sup_{0<t<1} |M_n(t)| \to^D \sup_{0<t<1} |B(t)|,$$

where $\{B(t),\ 0 \leq t \leq 1\}$ *denotes a Brownian bridge.*

Results

- The CUSUM process $M_n^{(r)}(t)$ adapted for testing for change in the covariance of the process (1), that is in $\gamma(r) = E\{X_i X_{i-r} - [E(X_i)]^2\}$, ($r \geq 0$ integer), is

$$M_n^{(r)}(t) = n^{-1/2}\Big(\sum_{1 \leq i \leq (n+1)t} X_i X_{i-r} - t \sum_{1 \leq i \leq n} X_i X_{i-r} \Big),$$

  for $0 \leq t < 1$.

- The following theorem gives the theoretical justification for using asymptotic critical values based on the distribution of the Brownian bridge.

Results

- The CUSUM process $M_n^{(r)}(t)$ adapted for testing for change in the covariance of the process (1), that is in $\gamma(r) = E\{X_i X_{i-r} - [E(X_i)]^2\}$, ($r \geq 0$ integer), is

$$M_n^{(r)}(t) = n^{-1/2}\Big(\sum_{1 \leq i \leq (n+1)t} X_i X_{i-r} - t \sum_{1 \leq i \leq n} X_i X_{i-r} \Big),$$

  for $0 \leq t < 1$.
- The following theorem gives the theoretical justification for using asymptotic critical values based on the distribution of the Brownian bridge.

Results

Theorem

Assume the sequence of observations $\{X_i\}$ satisfy relation (1) and (2)with some $\kappa > 8$, $a_j = O(\rho^j)$, $0 < \rho < 1$, as $j \to \infty$,

$$\sum_{1 \le j < \infty} a_j \neq 0.$$

Then with

$$\tau^2(r) = \lim_{n \to \infty} n^{-1} Var\Big(\sum_{1 \le i \le n} X_i X_{i-r} \Big) < \infty$$

$$\tau(r)^{-1} \sup_{0<t<1} |M_n^{(r)}(t)| \to^D \sup_{0<t<1} |B(t)|,$$

where $\{B(t),\ 0 \le t \le 1\}$ denotes a Brownian bridge.

Results

Consider the model of simple linear regression

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \ldots, n,$$

with $x_i = t_i$, the time parameter. We assume $\{\epsilon_i\}$ are dependent variables described by the linear relationship

$$\epsilon_i = \sum_{j=0}^{\infty} a_i \eta_{i-j}, \qquad i = 1, 2, \ldots, n,$$

where η_j are i.i.d.r.v.'s with mean zero, variance σ_η^2, and for some $\nu > 2$

$$E|\eta|^\nu < \infty,$$

and for the $\{a_j\}$ sequence of constants

$$a_j = O(\gamma^j), \quad j \to \infty,$$

for some γ, $0 < \gamma < 1$.

Results

We can extend the results of Gombay and Horváth (1994) for time series errors and use $Z_n(1,n) = \max_{1<k\leq n}|U_n(k)|$ for testing, where

$$U_n(k) = w_n(k)R_n(k),$$

$$R_n(k) = \left(\frac{n}{k(n-k)}\right)^{1/2}\sum_{i=1}^{k}(y_i - \overline{y}_n - \hat{\beta}_n(x_i - \overline{x}_n)),$$

$$w_n(k) = \{1 - k[\overline{x}_k - \overline{x}_n]^2/[\sum_{i=1}^{n}(x_i - \overline{x}_n)(1 - k/n)]\}^{-1/2}.$$

Results

Theorem

Under certain general conditions

$$\lim_{n\to\infty} P\Big\{\frac{1}{\sigma\, n} a(\log n) Z_n(1,n) \leq x + b(\log n)\Big\} = \exp(-2e^{-x}), \tag{5}$$

for some $\sigma_n > 0$, *where* $a(x) = (2\log x)^{1/2}$, $b(x) = 2\log x + 1/2\log\log x - 1/2\log\pi$. *Based on this theorem the algorithm is defined as follows.*

TEST. Reject H_0 at α level of significance if

$$a(\log n)\sigma_n Z(1,n) - b(\log n) \geq -\log(-1/2\log(-(1-\alpha)),$$

otherwise do not reject H_0.

Results

Theorem

Under certain general conditions

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma_n} a(\log n) Z_n(1,n) \leq x + b(\log n)\right\} = \exp(-2e^{-x}), \tag{5}$$

for some $\sigma_n > 0$, *where* $a(x) = (2\log x)^{1/2}$, $b(x) = 2\log x + 1/2\log\log x - 1/2\log\pi$. *Based on this theorem the algorithm is defined as follows.*

TEST. Reject H_0 at α level of significance if

$$a(\log n)\sigma_n Z(1,n) - b(\log n) \geq -\log(-1/2\log(-(1-\alpha))),$$

otherwise do not reject H_0.

Results

- The applicability of the results in Theorems 1-3 requires the estimation of the standardizing constants τ, $\tau(r)$, and σ_n.
- **Testing for change in the $MA(q)$ process**
  When testing for change in moving average models our new algorithm will estimate the standardizing constants with the help of the strong law of large numbers.
  Assume that process is $X_t = \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$
- To test for change in the autocovariance functions $\gamma(r) = E\{X_i X_{i-r} - [E(X_i]^2)\}$, $r = 0, \ldots, q$, given data $X_1, \ldots, X_n$, consider the process

$$\sum_{t=1}^{k} [X_t X_{t-r} - E(X_t X_{t-r})], \qquad k = 1, \ldots, n,$$

  where we use $X_i = 0$ if $i \leq 0$.

Results

- The applicability of the results in Theorems 1-3 requires the estimation of the standardizing constants τ, $\tau(r)$, and σ_n.
- **Testing for change in the $MA(q)$ process**
  When testing for change in moving average models our new algorithm will estimate the standardizing constants with the help of the strong law of large numbers.
  Assume that process is $X_t = \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$
- To test for change in the autocovariance functions $\gamma(r) = E\{X_i X_{i-r} - [E(X_i]^2)\}$, $r = 0, \ldots, q$, given data $X_1, \ldots, X_n$, consider the process

$$\sum_{t=1}^{k} [X_t X_{t-r} - E(X_t X_{t-r})], \qquad k = 1, \ldots, n,$$

where we use $X_i = 0$ if $i \leq 0$.

Results

- The applicability of the results in Theorems 1-3 requires the estimation of the standardizing constants τ, $\tau(r)$, and σ_n.
- **Testing for change in the $MA(q)$ process**
  When testing for change in moving average models our new algorithm will estimate the standardizing constants with the help of the strong law of large numbers.
  Assume that process is $X_t = \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$
- To test for change in the autocovariance functions $\gamma(r) = E\{X_i X_{i-r} - [E(X_i]^2)\}$, $r = 0, \ldots, q$, given data $X_1, \ldots, X_n$, consider the process

$$\sum_{t=1}^{k} [X_t X_{t-r} - E(X_t X_{t-r})], \qquad k = 1, \ldots, n,$$

where we use $X_i = 0$ if $i \leq 0$.

Results

As $X_t X_{t-r}$ is a function of $\epsilon_t, \dots, \epsilon_{t-r}, \epsilon_{t-r-1}, \dots, \epsilon_{t-r-q}$, the sequence $\{X_t X_{t-r}\}$ is $m = 1 + r + q$-dependent. It is a simple exercise to show that the strong law of large numbers holds for m-dependent sequences, so

$$\frac{1}{n} \sum_{k=1}^{n} X_k X_{k-r} \to^{a.s.} E(X_t X_{t-r}),$$

and for the same reason,

$$\frac{1}{n} \sum_{k=1}^{n} X_k X_{k-r} X_{k+u} X_{k+u-l} \to^{a.s.} E(X_t X_{t-r} X_{t+u} X_{t+u-l}).$$

Results

By stationarity

$$Var(\sum_{t=1}^{n} X_t X_{t-r}) =$$

$$\sum_{t=1}^{n} Var(X_t X_{t-r}) + 2\sum_{t=1}^{n} \sum_{s=t+1}^{n} Cov(X_t X_{t-r}, Y_s Y_{s-r})$$

$$= nVar(X_t X_{t-r}) + 2\sum_{t=1}^{n} \sum_{u=1}^{q+r} Cov(X_t X_{t-r}, X_{t+u} X_{t+u-r}).$$

Results

By the strong law of large numbers

$$\frac{1}{n}\sum_{k=1}^{n} X_k^2 X_{k-r}^2 - (\frac{1}{n}\sum_{k=1}^{n} X_k X_{k-r})^2 \rightarrow^{a.s.} Var(X_t X_{t-r}),$$

and

$$\frac{1}{n}\sum_{k=1}^{n} X_k Y_{k-r} X_{k+u} X_{k+u-r} - (\frac{1}{n}\sum_{k=1}^{n} X_k X_{k-r})^2 \rightarrow^{a.s.}$$

$$Cov(X_t X_{t-r}, X_s X_{s-r}).$$

Putting these estimators into expression above we get an estimate of $\tau^2(r)$.

Examples

Example 1: MA(1) process is $X_t = \epsilon_t + \theta\epsilon_{t-1}$, so $q = 1$ and the covariance function of the process is $\gamma(0) = \sigma_\epsilon^2(1+\theta^2)$, $\gamma(1) = \theta\sigma_\epsilon^2$, and $\gamma(h) = 0$ for $|h| > 1$. The standardizing constants are estimated by

$$\hat{\tau}(0) = \sum_{k=1}^{n} X_k^4 + 2\sum_{k=1}^{n} X_k^2 X_{k+1}^2 - 3\frac{1}{n}(\sum_{k=1}^{n} X_k^2)^2,$$

$$\hat{\tau}(1) = \sum_{k=1}^{n} X_k^2 X_{k-1}^2 +$$

$$+2\{(\sum_{k=1}^{n} X_k^2 X_{k-1} X_{k+1} + \sum_{k=1}^{n} X_k X_{k-1} X_{k+1} X_{k+2})\}$$

$$-5\frac{1}{n}(\sum_{k=1}^{n} X_k X_{k-1})^2.$$

Examples

Example 1: MA(1) process is $X_t = \epsilon_t + \theta\epsilon_{t-1}$, so $q = 1$ and the covariance function of the process is $\gamma(0) = \sigma_\epsilon^2(1 + \theta^2)$, $\gamma(1) = \theta\sigma_\epsilon^2$, and $\gamma(h) = 0$ for $|h| > 1$. The standardizing constants are estimated by

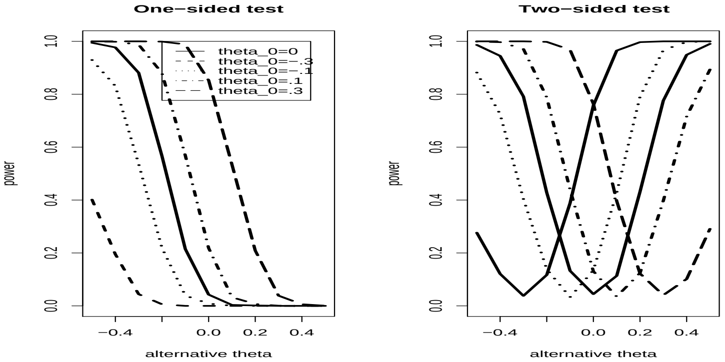
$$\hat{\tau}(0) = \sum_{k=1}^{n} X_k^4 + 2\sum_{k=1}^{n} X_k^2 X_{k+1}^2 - 3\frac{1}{n}(\sum_{k=1}^{n} X_k^2)^2,$$

$$\hat{\tau}(1) = \sum_{k=1}^{n} X_k^2 X_{k-1}^2 +$$

$$+2\{(\sum_{k=1}^{n} X_k^2 X_{k-1} X_{k+1} + \sum_{k=1}^{n} X_k X_{k-1} X_{k+1} X_{k+2})\}$$

$$-5\frac{1}{n}(\sum_{k=1}^{n} X_k X_{k-1})^2.$$

Examples



Empirical power of one-sided and two-sided tests for change in $\gamma(1)$ of MA(1) models. The sample size is $n = 500$ and the change point is at $\tau = 250$. The standard deviation is 1.0.

Examples

- **Example 2:** MA(5) process is $Y_t = \epsilon_t + \sum_{i=1}^{5} \theta_i \epsilon_{t-i}$, so $q = 5$,
  If the interest is in detecting change in the variance, then with $r = 0$ the statistic is

$$\sup_{1<k<n} \hat{\tau}^{-1}(0)(\sum_{t=1}^{k} X_t^2 - \frac{k}{n} \sum_{t=1}^{n} X_t^2), \qquad 1 \leq k \leq n.$$

- Assuming the MA(q) model is equivalent to assuming $\gamma(r) = 0$, $|r| > q$. So in that regard our approach is the same as that of Bartlett's estimator as it also gives $\gamma(r)$ value zero if $|r| > q(n) = c \log_{10} n$, (e.g. c=10 or c=15).

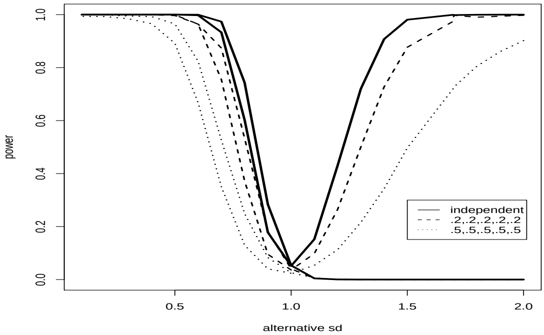
Examples

- **Example 2:** MA(5) process is $Y_t = \epsilon_t + \sum_{i=1}^{5} \theta_i \epsilon_{t-i}$, so $q = 5$,
  If the interest is in detecting change in the variance, then with $r = 0$ the statistic is

$$\sup_{1<k<n} \hat{\tau}^{-1}(0)\left(\sum_{t=1}^{k} X_t^2 - \frac{k}{n}\sum_{t=1}^{n} X_t^2\right), \qquad 1 \leq k \leq n.$$

- Assuming the MA(q) model is equivalent to assuming $\gamma(r) = 0$, $|r| > q$. So in that regard our approach is the same as that of Bartlett's estimator as it also gives $\gamma(r)$ value zero if $|r| > q(n) = c \log_{10} n$, (e.g. c=10 or c=15).

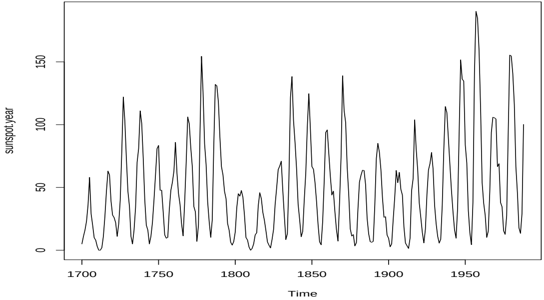
Examples



Empirical power of one and two-sided tests for change in the standard deviation of MA(5) models. The sample size is $n = 300$ and the change point is at $\tau = 150$. The initial standard deviation is 1.0, parameters θ_i are specified in the legend.

Examples

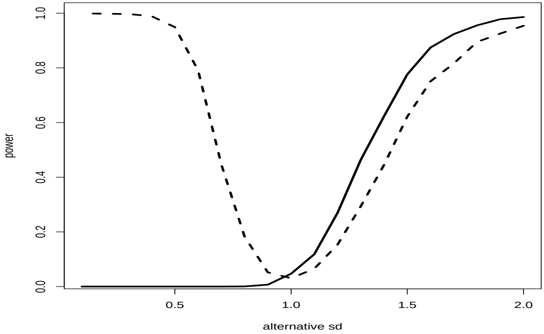
We now consider how this method works if the process is not MA(q), but a more general linear one as in (1).



Wölfer sunspot numbers 1700 - 1988, $n = 289$.

Recall section 1749–1924 (n=176) is ARMA(2,1) with $\phi_1 = 1.24$, $\phi_2 = -.58$, $\theta_1 = -.12$.

Examples



Empirical power of one and two-sided tests for change in the standard deviation of ARMA(2,1) models when MA(5) model is assumed. The sample size is $n = 300$ and the change point is $\tau = 150$. The initial standard deviation is 1.0. The parameters of the ARMA model: $\phi_1 = 1.24$, $\phi_2 = -.58$, $\theta_1 = -.12$.

Examples

For the longer series 1700 - 1988 (n=289) our algorithm rejects the null hypothesis of no change in the variance (observed value of test statistic is 1.45055):
p-value = .0298 for the two-sided test
p-value = .0149 for the one-sided test
The change is put to observation no. $k = 236$, corresponding to year 1936.
Estimates:
$\hat{\sigma}_1^2 = 1137.9$, ($\hat{\sigma}_1 = 33.7$, $n_1 = 236$)
$\hat{\sigma}_2^2 = 2591.4$, ($\hat{\sigma}_2 = 50.9$, $n_2 = 53$)

Examples

As the MA component is not very important relative to the AR component ($\theta = -.12$), we can use an AR(2) approximation to these data. The results of *"Change detection in autoregressive time series"* (G., JMVA(2008)) can be applied and test for change in the variance.

The results: observed value of the two-sided test statistic is 2.18, with change-point estimate $k = 246$ (year 1946).

p-value = .0002 for the two-sided test

Estimates:

$\hat{\sigma}_1^2 = 1155.17$, ($\hat{\sigma}_1 = 33.99$, $n_1 = 246$)

$\hat{\sigma}_2^2 = 2833.59$, ($\hat{\sigma}_2 = 53.23$, $n_2 = 43$)

$\hat{\phi}_1 = 1.34$, $\hat{\phi}_2 = -.64$.

This difference in the observed level of significance is the price for the rough fit of the MA(5) model to these data.

Examples

As the MA component is not very important relative to the AR component ($\theta = -.12$), we can use an AR(2) approximation to these data. The results of *"Change detection in autoregressive time series"* (G., JMVA(2008)) can be applied and test for change in the variance.
The results: observed value of the two-sided test statistic is 2.18, with change-point estimate $k = 246$ (year 1946).
p-value = .0002 for the two-sided test
Estimates:
$\hat{\sigma}_1^2 = 1155.17$, ($\hat{\sigma}_1 = 33.99$, $n_1 = 246$)
$\hat{\sigma}_2^2 = 2833.59$, ($\hat{\sigma}_2 = 53.23$, $n_2 = 43$)
$\hat{\phi}_1 = 1.34$, $\hat{\phi}_2 = -.64$.
This difference in the observed level of significance is the price for the rough fit of the MA(5) model to these data.

Examples

As the MA component is not very important relative to the AR component ($\theta = -.12$), we can use an AR(2) approximation to these data. The results of *"Change detection in autoregressive time series"* (G., JMVA(2008)) can be applied and test for change in the variance.
The results: observed value of the two-sided test statistic is 2.18, with change-point estimate $k = 246$ (year 1946).
p-value = .0002 for the two-sided test
Estimates:
$\hat{\sigma}_1^2 = 1155.17$, ($\hat{\sigma}_1 = 33.99$, $n_1 = 246$)
$\hat{\sigma}_2^2 = 2833.59$, ($\hat{\sigma}_2 = 53.23$, $n_2 = 43$)
$\hat{\phi}_1 = 1.34$, $\hat{\phi}_2 = -.64$.
This difference in the observed level of significance is the price for the rough fit of the MA(5) model to these data.

Examples

Example 2: Time series linear regression.
The annual volume of discharge from the Nile River at Aswan 1871-1970 is a data set that has been examined by many statistical procedures. It is reasonable to apply our test to see if one linear regression model would fit the data.
Our test indicated a change in the parameters of linear regression at $\hat{\tau} = 28$, corresponding to year 1899. Different orders q of MA(q) were used for the error process. Each lead to the rejection of H_0, with p-value 5.35×10^{-10}, 0.0019, and 0.01, respectively for $q = 0, 1, 2$.
The overall slope is estimated as $\hat{\beta} = -2.71$, the before and after change slopes are estimated as $\hat{\beta}_1 = 1.27$ and $\hat{\beta}_2 = 0.38$, respectively.

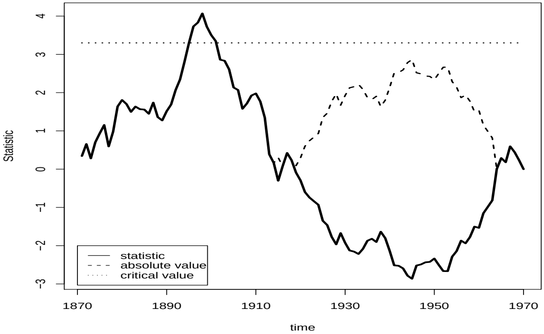
Examples

Example 2: Time series linear regression.
The annual volume of discharge from the Nile River at Aswan 1871-1970 is a data set that has been examined by many statistical procedures. It is reasonable to apply our test to see if one linear regression model would fit the data.
Our test indicated a change in the parameters of linear regression at $\hat{\tau} = 28$, corresponding to year 1899. Different orders q of MA(q) were used for the error process. Each lead to the rejection of H_0, with p-value 5.35×10^{-10}, 0.0019, and 0.01, respectively for $q = 0, 1, 2$.
The overall slope is estimated as $\hat{\beta} = -2.71$, the before and after change slopes are estimated as $\hat{\beta}_1 = 1.27$ and $\hat{\beta}_2 = 0.38$, respectively.

Examples

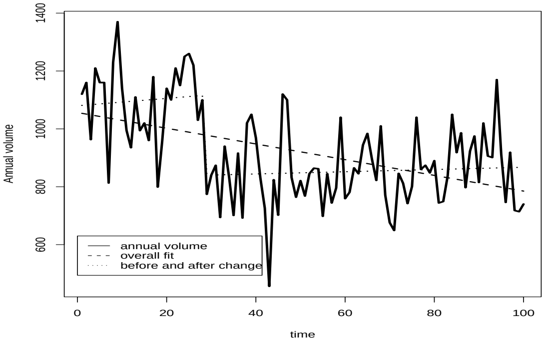
Example 2: Time series linear regression.
The annual volume of discharge from the Nile River at Aswan 1871-1970 is a data set that has been examined by many statistical procedures. It is reasonable to apply our test to see if one linear regression model would fit the data.
Our test indicated a change in the parameters of linear regression at $\hat{\tau} = 28$, corresponding to year 1899. Different orders q of MA(q) were used for the error process. Each lead to the rejection of H_0, with p-value 5.35×10^{-10}, 0.0019, and 0.01, respectively for $q = 0, 1, 2$.
The overall slope is estimated as $\hat{\beta} = -2.71$, the before and after change slopes are estimated as $\hat{\beta}_1 = 1.27$ and $\hat{\beta}_2 = 0.38$, respectively.

Examples



The statistics process and its absolute value for the Nile annual discharge at Aswan with critical level.

Examples

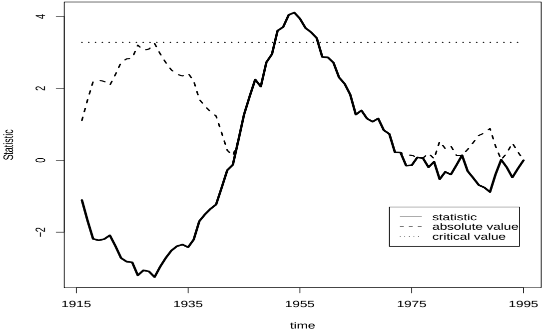


Annual discharge of Nile with regression lines. Change is detected in 1899, $\hat{\tau} = 28$.

Examples

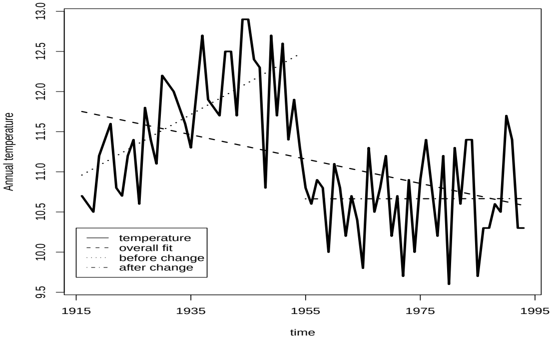
Example 3. The annual mean maximum temperatures of Collegeville, Nova Scotia, Canada, (1916-1995)were tested for inhomogeneity by Vincent (1998), $n = 80$.Our test detected change in the linear regression parameters at $\hat{\tau} = 39$ with p-value 0.020 and 2.2×10^{-10}, for MA(2) and MA(0) error approximations, respectively. The overall slope estimate is $\hat{\beta} = -0.0146$, accounting for one change we have estimators $\hat{\beta}_1 = 0.0397$ and $\hat{\beta}_2 = 0.0016$.

Examples



The statistics process and its absolute value for Collegeville annual temperatures with critical level.

Examples



Collegeville annual temperatures with regression lines. Change is detected around 1899, $\hat{\tau} = 39$.

Examples

Example 4. Lund and Reeves (2002), Berkes et al. (2008) examined the data on annual mean temperatures at Chula Vista, California during 1919-1996, inclusive ($n = 78$).

Lund and Reeves (2002) used F-statistics in their test for change in linear regression, and assumed independence in the error process, Berkes et al. (2008) tested for change in the mean, without assuming independence. It is reasonable to combine the advantages of both methods and test for change in the parameters of linear regression without assuming independence of errors.

Our test was significant with change detected at estimated time $\hat{\tau} = 62$ (1981). The three slope estimates are $\hat{\beta} = 0.02$, $\hat{\beta}_1 = 0.05$, $\hat{\beta}_2 = 0.089$.

Examples

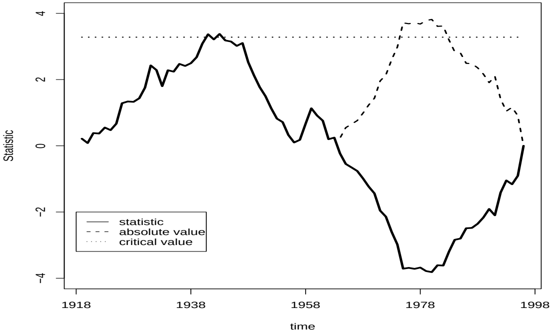
Example 4. Lund and Reeves (2002), Berkes et al. (2008) examined the data on annual mean temperatures at Chula Vista, California during 1919-1996, inclusive ($n = 78$). Lund and Reeves (2002) used F-statistics in their test for change in linear regression, and assumed independence in the error process, Berkes et al. (2008) tested for change in the mean, without assuming independence. It is reasonable to combine the advantages of both methods and test for change in the parameters of linear regression without assuming independence of errors.

Our test was significant with change detected at estimated time $\hat{\tau} = 62$ (1981). The three slope estimates are $\hat{\beta} = 0.02$, $\hat{\beta}_1 = 0.05$, $\hat{\beta}_2 = 0.089$.

Examples

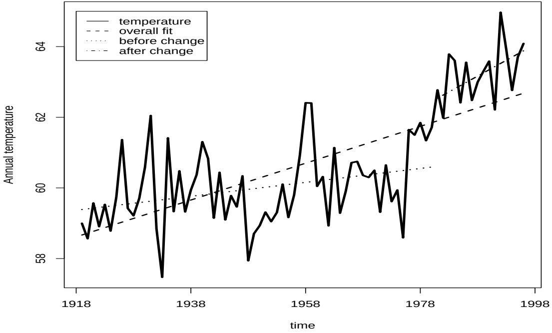
Example 4. Lund and Reeves (2002), Berkes et al. (2008) examined the data on annual mean temperatures at Chula Vista, California during 1919-1996, inclusive ($n = 78$).
Lund and Reeves (2002) used F-statistics in their test for change in linear regression, and assumed independence in the error process, Berkes et al. (2008) tested for change in the mean, without assuming independence. It is reasonable to combine the advantages of both methods and test for change in the parameters of linear regression without assuming independence of errors.
Our test was significant with change detected at estimated time $\hat{\tau} = 62$ (1981). The three slope estimates are $\hat{\beta} = 0.02$, $\hat{\beta}_1 = 0.05$, $\hat{\beta}_2 = 0.089$.

Examples



Chula Vista annual temperatures statistics process and its absolute value with $\alpha = 0.05$ critical level.

Examples



Chula Vista annual temperatures with regression lines. Change is detected in around 1998, $\hat{\tau} = 62$.

References

- Berkes, I., Gombay, E. and Horvath, L. (2008). Testing for changes in the covariance structure of linear processes. to appear
- Gombay, E. (2008). Change Detection in Autoregressive Time Series, *J. Multivariate Analysis*, **99** 451-464.
- Gombay, E. and D. Serban (2005). Monitoring Parameter Change in AR(p) Time Series Models, *Statistics Centre Technical Reports* **05.03**, The University of Alberta, Edmonton, Canada