MBB bootstrap

ShanZhong

April 29, 2019

Introduction to Bootstrap

1.The normal bootstrap methods: resample the data(individual observations) with replacement. The statistic of interest is computed from the resample, require a i.i.d sample.

Resampling residuals:

Fit a model to the data to get residuals and fitted value. Resample the residuals, add back to the fitted value to get a pseudo data(artificially generated data). Refit the model and then compute the satsitic of interest. Do not require the data to be i.i.d, but require residual to be i.i.d.

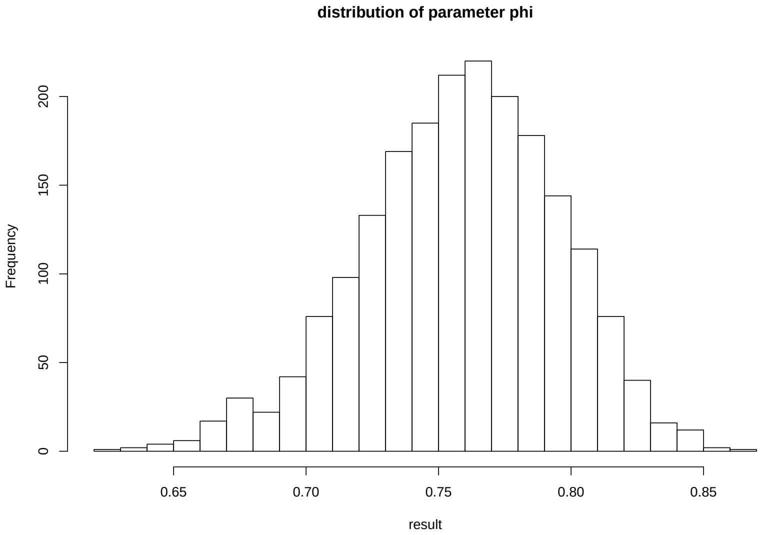
Here is an example using residual bootstrap to estimate for ar(1) parameter with true phi = 0.8

```r
# specify parameters:
phi <- c(.8); n <- 300; sigma <- 2
X <- arima.sim(n=n,list(ar=phi,sd=sigma))
fit<-arima(X,order=c(1,0,0))
fit$coef[1]
```

```
##       ar1
## 0.8559825
```

```r
# how do generate sample
pseudo.residual<-sample(fit$residuals,length(fit$residuals)
  Y<-numeric()
  Y[1]<-X[1]
  for(i in 2:length(X)){
    Y[i]<-Y[i-1]*fit$coef[1]+pseudo.residual[i]
  }
```

here is the plot



However, if the structure of serial correlation is not tractable or is misspecified, the residual based methods will give inconsistent

Block Boostrap: resample blocks, do not require residual to be i.i.d.

The block bootstrap is used when the data, or the errors in a model, are correlated. In this case, a simple case or residual resampling will fail, as it is not able to replicate the correlation in the data.

Block Boostrap:

The block bootstrap tries to replicate the correlation by resampling instead blocks of data. The block bootstrap has been used mainly with data correlated in time to maintaining the time series dependency structure within a pseudo-sample, but can also be used with data correlated in space, or among groups (so-called cluster data).

Here we introduce 3 types of block bootstrap:

Non-overlapping block:the variable of interest is split into non-overlapping blocks.

Moving-blocks bootstrap: The moving block bootstrap (MBB) is applicable to stationary time series data. Sample whole blocks and concatenate them, in contrast to a single observation at a time.

Stationary bootstrap: Block size is a random variable

MBB bootstrap example for a ar(1) model

Using block length = 25

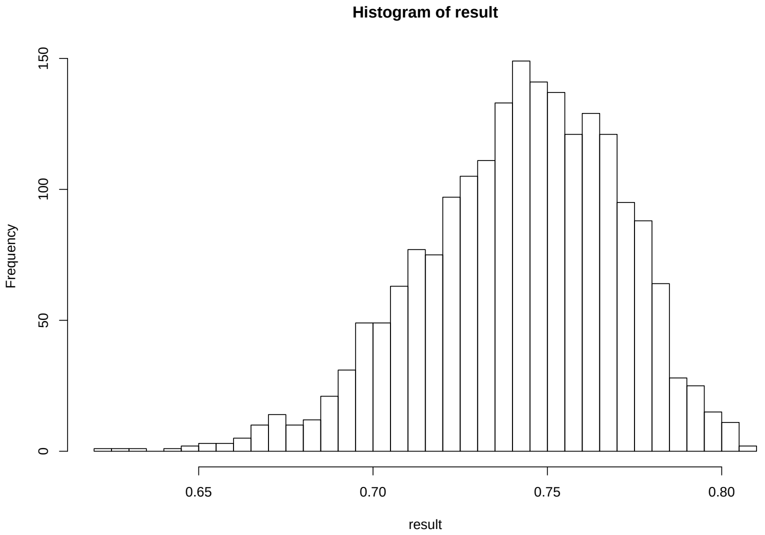
```
# specify parameters:
phi <- c(.8);n <- 300; sigma <- 2

X <- arima.sim(n=n,list(ar=phi,sd=sigma))
fit<-arima(X,order=c(1,0,0));fit$coef[1]
```

```
##       ar1
## 0.7904215
```

```
l<-25
X.blocks <- matrix(NA,n-l+1,l)
for( j in 1:(n-l+1)){
    X.blocks[j,] <- X[j:(j+l-1)]}
X.star.mat <- X.blocks[sample(1:(n-l+1),floor(n/l) + 1,repl
    X.star <- as.vector(t(X.star.mat))[1:n]
```

Plot




```
## [1] 0.7419244
```

Servel points in summary:

The mean of the moving block bootstrap is biased

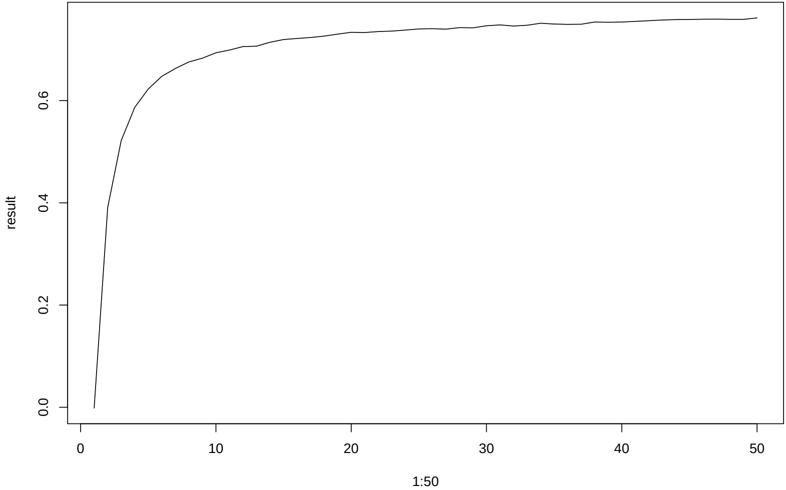
The MBB estimator of the variance of sqrt(n)*(X.bar) is also biased

MBB works with dependent data, however, the bootstrapped observations will not be stationary anymore by construction.

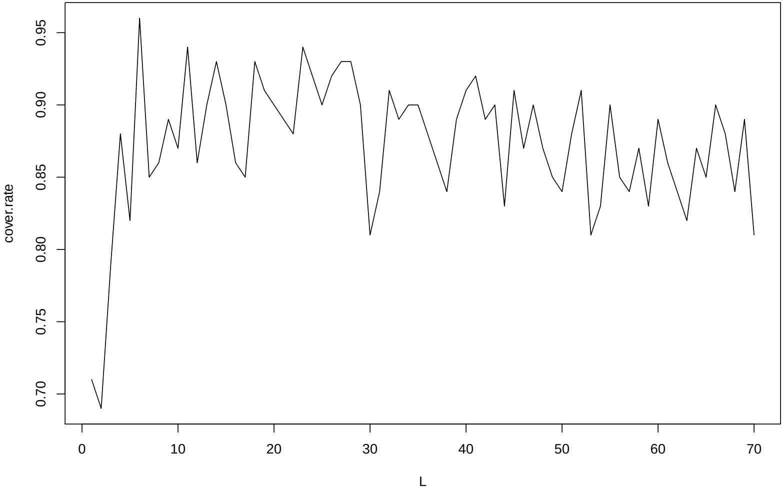
The moving bootstrap involves resampling possibly overlapping blocks. The mixed block bootstrep (MBB) does not force one to select a model and the only parameter required is the block length l.

the relation of block length l and MBB parameter

```
##       ar1
## 0.7904215
```



Here is the cover rate for different l, using a arma(2,1) model with n = 300



Stationary bootstrap:

Block size is a random variable in itself coming from a geometric distribution with some expected block size

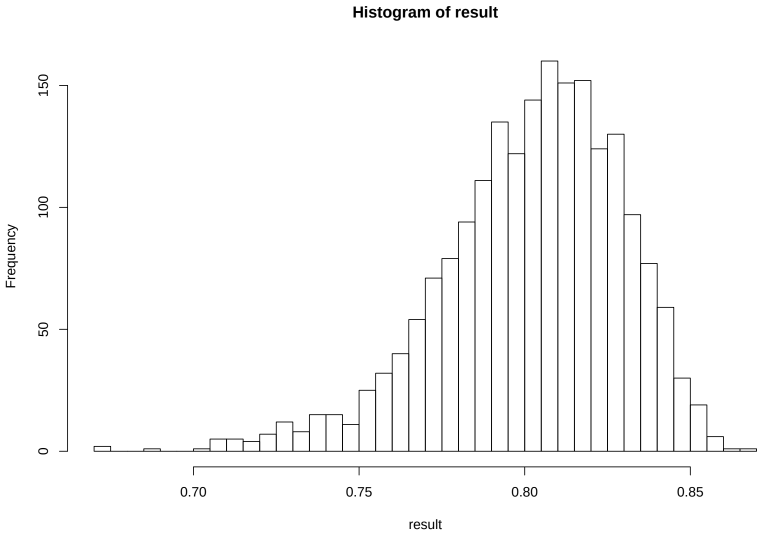
```
##       ar1
## 0.8430584
```

Here is how we generate a SB sample

```
# have a look to determine appropriate block size
expectedl<-25 ; gprob=1/expectedl
X.star<-numeric(n)

loc <- round(runif(1,min = 1, max = n)) # loc for location
  for (i in 1:n){
    # In probability gprob, we take next observation, other
    if (runif(1)>gprob){loc = loc+1} else {loc = round(runi
    if (loc>n){loc <- loc-n }# wrap the serie as a circule
    X.star[i] = X[loc]
  }
```

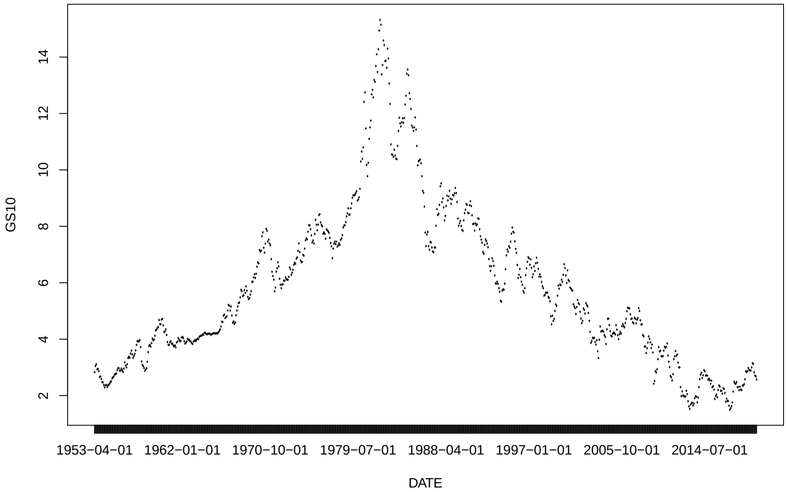
Here is the performance of SB



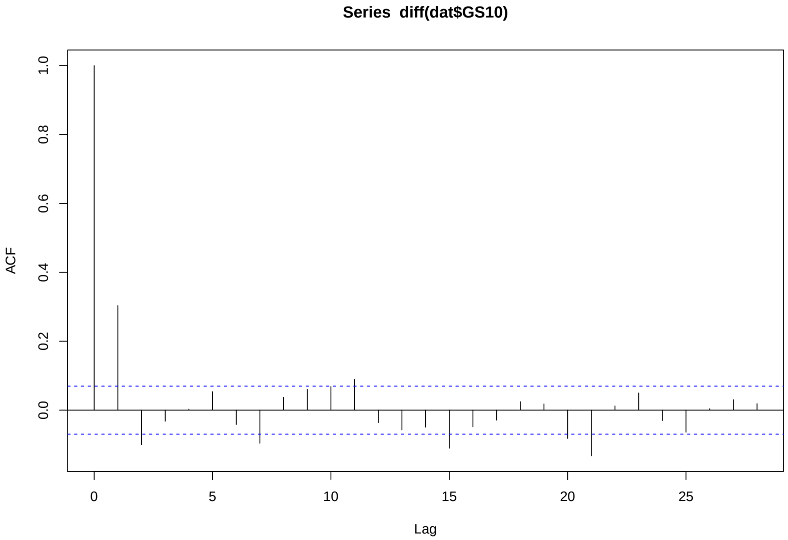

```
## [1] 0.8020774
```

Some real data exaple

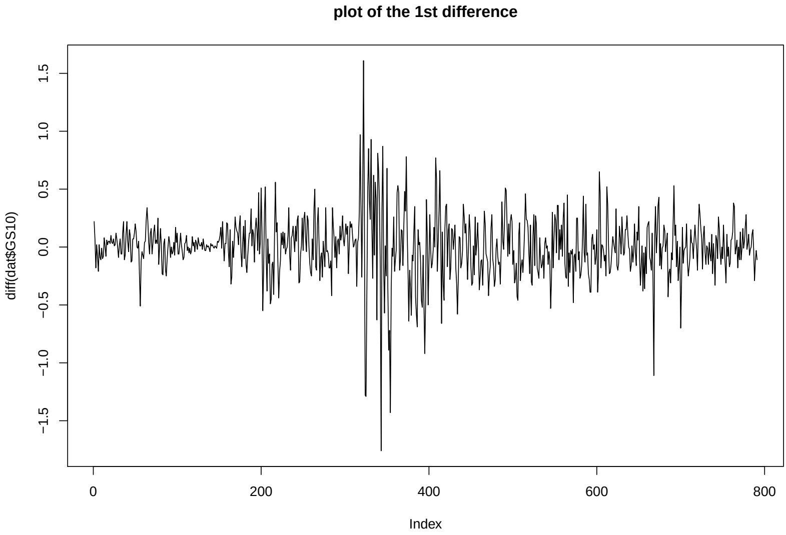
let's look at the interest rate data using 10-Year Treasury Constant Maturity Rate



plot

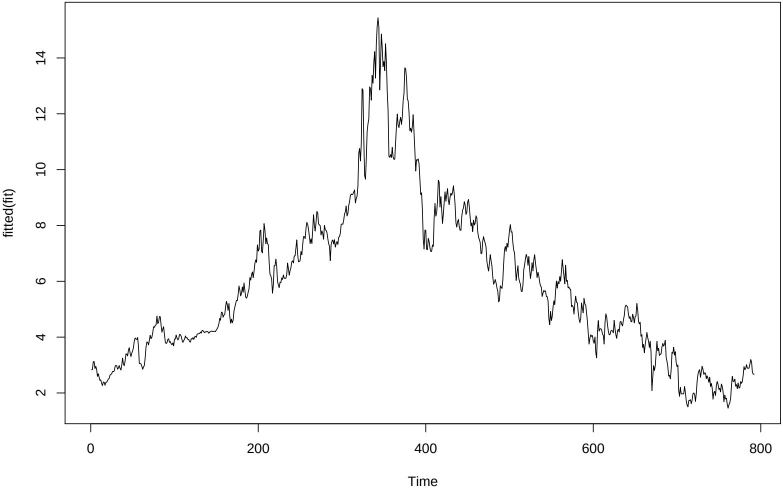


plot

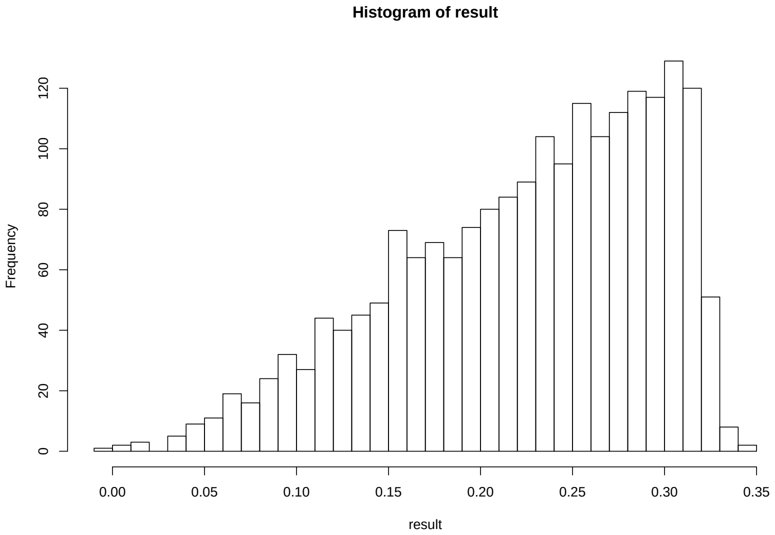

plot of the 1st difference
diff(dat$GS10)
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
200
400
600
800
Index

Fit with a arima(1,1,0) model

```
##       ar1
## 0.3035761
```

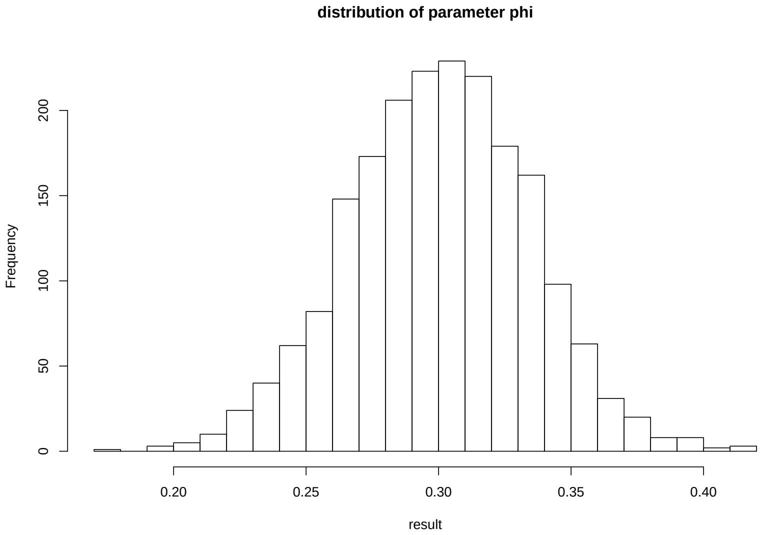


Moving block use l = 350




```
## [1] 0.2252656
```

use residual based bootstrap




```
## [1] 0.3005251
```