Bootstrapping Stationary Sequences by the Nadaraya-Watson Regression Estimator

Cheolyong Park and Tae Yoon Kim

Keimyung University

June 21, 2002

1. Introduction

- Classical bootstrap: $X_1, X_2, \ldots, X_n \sim iid\ F$

  Suppose we like to estimate $Var(\overline{X_n})$.

  step 1) Calculate F_n, the empirical cdf of $X_1, X_2, \ldots, X_n$.

  step 2) Generate $X_1^*, X_2^*, \ldots, X_n^* \sim iid\ F_n$.

  Calculate $\overline{X_n^*}$.

  step 3) Repeat step 2 a large number, say B, of times.

  Based on bootstrapped copies $\overline{X_n^*}^{(1)}, \overline{X_n^*}^{(2)}, \ldots, \overline{X_n^*}^{(B)}$,

  we calculate a bootstrap estimator of $Var(\overline{X_n})$ by

$$\frac{\sum_{i=1}^{B}\left\{\overline{X_n^*}^{(i)} - \overline{\overline{X_n^*}}\right\}^2}{B-1}, \quad \text{where } \overline{\overline{X_n^*}} = \frac{\sum_{i=1}^{B}\overline{X_n^*}^{(i)}}{B}.$$

- The classical bootstrap fails if $X_1, X_2, \ldots, X_n$ are dependent.

  Clear since bootstrap copies $X_1^*, X_2^*, \ldots, X_n^*$ are independent.

- Two modified versions for stationary time series:
  1) Block bootstrap
  2) Parametric model fitting

- Block bootstrap

Consider $n-l+1$ blocks of l consecutive observations

$B_1=(X_1,\dots,X_l)$, $B_2=(X_2,\dots,X_{l+1}),\dots,$ $B_{n-l+1}=(X_{n-l+1},\dots,X_n)$.

We apply the classical bootstrap to these blocks.

Suppose $k=n/l$ is a positive integer.

We choose k blocks $B_{i_1},B_{i_2},\dots,B_{i_k}$ by the classical bootstrap.

A bootstrap copy is obtained by $(B_{i_1},B_{i_2},\dots,B_{i_k})$.

If k is not an integer, we took $k=[n/l]+1$ and delete redundant elements over n.

It is critical to choose the right block size l.

If l is large, we can maintain dependence structure but we make only a few possible copies.

If l is small, we can make various copies but we can not maintain dependence structure.

- Parametric model fitting

We assume that the series follows a parametric model.

We fit the model and obtain the residuals $e_1, e_2, \ldots, e_n$.

The classical bootstrap is applied to these residuals.

A bootstrap copy is "fitted values+bootstrapped residuals".

It is critical to know the right parametric model.

True model is hardly known in advance.

e.g.) AR(1) model

$X_t = \widehat{\phi}_1 X_{t-1} + e_t$ is fitted to the series.

A bootstrapped residuals $e_1^*, e_2^*, \ldots, e_n^*$ are obtained.

A bootstrap copy is obtained by $X_t^* = \widehat{\phi}_1 X_{t-1}^* + e_t^*$.

- Idea of our method: nonparametric model fitting

Our method is the same as parametric model fitting except that we fit a nonparametric model by a kernel regression estimator.

Here we use the Nadaraya-Watson regression estimator.

2. Our Method

- Suppose the data follows a p-th order autoregressive model

$$X_t = f(X_{t-1}, \ldots, X_{t-p}) + \varepsilon_t$$

where $\{\varepsilon_t\}$ is a white noise process. Then the conditional mean

$$E(X_t | X_{t-1} = x_1, \ldots, X_{t-p} = x_p) = f(x_1, \ldots, x_p)$$

can be estimated by the Nadaraya-Watson regression estimator

$$\hat{f}(x_1, \ldots, x_p) = \frac{\sum_{t=p+1}^{n} X_t \prod_{i=1}^{p} K\left(\frac{x_i - X_{t-i}}{h}\right)}{\sum_{t=p+1}^{n} \prod_{i=1}^{p} K\left(\frac{x_i - X_{t-i}}{h}\right)}$$

where K is a kernel function and h is a smoothing parameter.

- Procedures

1) Estimate the conditional mean function by the N-W estimator and then we obtain the residuals $e_t = X_t - \hat{f}(X_{t-1}, \ldots, X_{t-p})$.

2) We obtain a sequence of bootstrapped residuals $\{e_t^*\}$ by the classical bootstrap.

3) A bootstrapped copy of the time series data is obtained by the formula

$$X_t^* = \hat{f}(X_{t-1}^*, \ldots, X_{t-p}^*) + e_t^*, \quad (X_0^* = x_0, \ldots, X_{-p+1}^* = x_{-p+1}), \quad t = 1, 2, \ldots, n.$$

- The choice of the order p is critical.

  But the order misspecification is not so problematic compared to the dependence corruption of the block bootstrap.

  We mainly focus on the case where $p = 1$.

- If the order is correct, the choice of smoothing parameter h is critical. It is somewhat related to the choice of block size l.

  If h is large, then $\hat{f}(x) = \overline{X}$ with $e_t = X_t - \overline{X}$.

  Corresponds to the block bootstrap with $l = 1$.

  If h is small, then $\hat{f}(X_{t-1}) = X_t$ with $e_t = 0$.

  Corresponds to the block bootstrap with $l = n$.

3. Monte Carlo Simulation Study

- Our method with $p=1$ is compared to the block bootstrap.
- Simulation models

\- the first order models

1) Linear autoregressive model of order 1, denoted by AR(1):

$X_t = 0.8X_{t-1} + \varepsilon_t$ with ε_t i.i.d. $N(0, 0.36)$.

2) Threshold autoregressive model, denoted by TAR;

$X_t = 0.528Z_t$

where $Z_t = (0.9Z_{t-1} + \varepsilon_t)I(Z_{t-1} \leq -2.5) + (0.8Z_{t-1} + \varepsilon_t)I(Z_{t-1} > -2.5)$

with ε_t i.i.d. $N(0,1)$.

3) Bilinear model, denoted by BL;

$X_t = 0.254Z_t$ where $Z_t = 0.4Z_{t-1} + 0.8Z_{t-1}\varepsilon_{t-1} + \varepsilon_t$

with ε_t i.i.d $N(0,1)$.

\- models of misspecified orders

1) Linear autoregressive model of order 2, denoted by AR(2);

$X_t = 1.372X_{t-1} - 0.677X_{t-2} + \varepsilon_t$ with ε_t i.i.d. $N(0, 0.179)$.

2) Moving average model of order 1, denoted by MA(1);

$X_t = \varepsilon_t + 0.5\varepsilon_{t-1}$ with ε_t i.i.d. $N(0, 0.8)$.

- Simulation setup

1) We choose the sample size $n = 128$ and all our results are based on 200 simulations and 500 bootstrap copies.

2) For each simulation, we generate the copies using the N-W bootstrap with bandwidth $h = 0.1, 0.2, \ldots, 1.0$ as well as using the block bootstrap with block length $l = 1, 2, \ldots, 20$.

3) Measure of performance is evaluated by RMSE(relative MSE)

$$E\left(\sqrt{n\,Var(\overline{X}_n^*)} - \sigma_n\right)^2 / \sigma_n^2$$

where $\sigma_n = \sqrt{n\,Var(\overline{X}_n)}$.

- Simulation results for the first order models are shown in Tables 1 and 2.

  - Our method is better in AR(1), TAR and worse in BL.

- Simulation results for the models of misspecified orders are shown in Table 3.

  - Our method is better in MA(1) and worse in AR(2).
  - Optimal l for AR(2) by Bühlman and Künsch (1999) is 9.

    Our method is as good as the block bootstrap near $l = 9$.

Table 3.1: Relative mean-square errors of our method for the bandwidth $h = 0.1, 0.2, \ldots, 1.0$

h	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
AR(1)	.0742	.0895	.1100	.1604	.2120	.2613	.3028	.3327	.3592	.3770
TAR	.1491	.1665	.2130	.2774	.3388	.3901	.4268	.4543	.4735	.4869
BL	.2688	.4599	.8053	1.189	1.122	1.087	1.067	1.096	1.045	1.096

Table 3.2: Relative mean square errors of the block bootstrap for the block size $l = 1, 2, \ldots, 20$.

l	1	2	3	4	5	6	7	8	9	10
AR(1)	.4514	.3182	.2427	.1925	.1595	.1381	.1224	.1108	.1031	.0969
TAR	.5399	.4120	.3381	.2852	.2552	.2289	.2081	.1928	.1830	.1764
BL	.3195	.2423	.2235	.2171	.2187	.2202	.2242	.2253	.2277	.2335
l	11	12	13	14	15	16	17	18	19	20
AR(1)	.0946	.0908	.0909	.0907	.0905	.0902	.0925	.0948	.0949	.0988
TAR	.1717	.1675	.1646	.1653	.1655	.1633	.1656	.1699	.1696	.1726
BL	.2341	.2390	.2436	.2462	.2463	.2512	.2498	.2547	.2567	.2556

Table 3.3: Relative mean square errors of our method for $h = 0.1, 0.2, \ldots, 1.0$ and the block bootstrap for $l = 1, 2, \ldots, 20$.

h	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$AR(2)$	1.087	.7419	.4310	.2060	.0971	.0494	.0312	.0309	.0364	.0423
$MA(1)$	.0303	.0274	.0216	.0163	.0145	.0143	.0151	.0182	.0209	.0245
l	1	2	3	4	5	6	7	8	9	10
$AR(2)$	.0944	.0159	.0247	.0453	.0548	.0466	.0397	.0341	.0324	.0304
$MA(1)$	.0677	.0214	.0156	.0162	.0176	.0191	.0229	.0240	.0266	.0300
l	11	12	13	14	15	16	17	18	19	20
$AR(2)$	.0329	.0336	.0377	.0380	.0390	.0424	.0436	.0491	.0531	.0523
$MA(1)$	.0324	.0363	.0401	.0431	.0449	.0513	.0501	.0566	.0560	.0564

4. Conclusion

- Our method might be a better bootstrap scheme for time series data in most cases.
  - It might be useful to provide some theoretical justification.

- For a practical purpose, we need to develop an automatic bandwidth selection method.
  - A method based on cross validation is available but performs poorly in a simulation study we had conducted.
  - A bandwidth selector achieving the independence of residuals efficiently might be a good alternative.