Some real life data analysis on stationary and non-stationary Time Series

Ajay Shankar Bidyarthy
b.ajay@iitg.ernet.in

Abhishek Kumar
kumar.ak@iitg.ernet

Mohit Yadav
y.mohit@iitg.ernet.in

April 3, 2012

Abstract

Now days there are several applications of block bootstrap method in stationary and non stationary time series models. One of the well known application of this method is in case of IGARCH time series models. Bootstrap confidence intervals also play mazor role in finding out confidence intervals of given any time series data. First we have presented how block bootstrap method works in stationary and non stationary time series models. Then we have presented the two bootstrap confidence interval methods one is bootstrap percentile confidence interval method and second is bootstrap-t confidence interval method.

For any time series data using Autoregressinve AR(1) and AR(2) time series model to analyze the data set and then find out the bootstrap confidence interval using percentile confidence interval method and bootstrap-t confidence interval method.

1 Block bootstrap method

The bootstrap is a simulation approach to estimate the distribution of test statistics. The original method is to create bootstrap samples by resampling the data randomly, and then constructs the associated empirical distribution function. Often, the original bootstrap method provides improvements to the poor asymptotic approximations when data are independently and identically distributed. However, the performance of the original procedure can be far from satisfactory for time series data with serial correlation and heteroscedasticity of unknown form.

The block bootstrap is the most general method to improve the accuracy of bootstrap for time series data. By dividing the data into several blocks, it can preserve the original time series structure within a block. However, the accuracy of the block bootstrap is sensitive to the choice of block length, and the optimal block length depends on the sample size, the data generating process, and the statistic considered. To date, there is no proper diagnostic tool to choose the optimal block lengths and it still remains as an unsolved question for future study.

The detail of the block bootstrap procedure in our Monte Carlo experiment takes the following steps:

Step 1. Choose the block length which increases with the sample size. In our block bootstrap procedure, we choose the block length (l) by the criterion l = $T^{1/3}$, where T is the sample size. Hall and Horowitz (1996) use two block lengths l = 5 and l = 10 for both two sample sizes (T = 50, 100) and Inoue and Shintani (2001) select the block length by an automatic procedures, which result in an average block length of 3.5 for sample size 64 and 6 for sample size 128. Here we use a simple rule, which the block length in our simulations is similar to the average block length of Inoue and Shintani (2001).

Step 2. Resample the blocks and generate the bootstrap sample. The blocks may be overlapping or non-overlapping. According to Lahiri (1999), and Andrews (2002), there is little difference in performance for these two methods. For the overlapping method, we divide the data into $T-l+1$ blocks, which block 1 being $\{y_1, y_2, ..., y_l\}$, block 2 being $\{y_2, y_3, ..., y_{l+1}\}$, , etc. For the non-overlapping method, we divide the data into T/l blocks, which block 1 being $\{y_1, y_2, ..., y_l\}$, block 2 being $\{y_{l+1}, y_{l+2}, ..., y_{l+l}\}$, , etc. In our Monte Carlo experiments, we adopt the overlapping method and resample y_t , x_t and z_t together which is called the pairs bootstrap. The block bootstrap sample can be generated as follows:

$$(y_t^*, x_t^*, z_t^*) = (y_{i+j}, x_{i+j}, z_{i+j})$$

where t = 1,2,3,...T, i is iid uniform random variable on $\{1, 2, 3, ...T-l+1\}$ and j = 1,2,3,...l.

Step 3. Calculate the efficient bootstrap GMM estimator and the test statistic. First, estimate the bootstrap TSLS estimator

$$\hat{\beta}^*_{TSLS} = (X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}X^*)^{-1}(X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}y^*)$$

and use the residual $\hat{e}_t^* = y_t^* - x_t^*\hat{\beta}^*_{TSLS}$ to construct the efficient weighting matrix

$$W_t^* = (\tfrac{1}{T}\sum_{t=1}^{T}\hat{g}_t^*\hat{g_t^*}')^{-1}$$

where $\hat{g}_t^* = z_t^*\hat{e}_t^*$. Second, calculate the efficient bootstrap GMM estimator and the bootstrap variance

$$\hat{\beta}^*_{GMM} = (X^{*\prime}Z^*W_T^*Z^{*\prime}X^*)^{-1}(X^{*\prime}Z * W_T^*Z^{*\prime}y^*)$$
$$(\hat{\sigma}^*)^2 = (X^{*\prime}Z^*W_T^*Z^{*\prime}X^*)^{-1}$$

In the sense, the bootstrap GMM estimator $\hat{\beta}^*_{GMM}$ is a consistent estimator of β. However, because the bootstrap sample can not satisfy the same moment condiation as the population distribution, it fails to achieve an asymptotic refinement. A correlation suggested by Hall and Horowitz (1996) is to re-center the bootstrap version of the moment functions. Therefore, in the linear model, the revised bootstrap GMM estimator derived by Hansen (2004) is

$$\hat{\beta}^*_{GMM} = (X^{*\prime}Z^*W^*_TZ^{*\prime}X^*)^{-1}((X^{*\prime}Z * W^*_TZ^{*\prime}y^* - Z'\hat{u}))$$

where $\hat{u}$ are residuals of GMM estimation from original data. Finally, calculate the bootstrap test statistic

$$t^* = \frac{\hat{\beta}^*_{GMM} - \beta}{\hat{\sigma}^*}.$$

and the final step is :

Step 4. Calculate the bootstrap critical values of the test statistic and test the null hypothesis. First, construct the empirical distribution of the test statistic by repeating step2 and step3 for sufficient amount of times and sorting the bootstrap test statistic t^* from the smallest to the largest. Usually, we set the $(1-\frac{\alpha}{2})$ quantile and the $(\frac{\alpha}{2})$ quantile of the distribution of t^* for our bootstrap critical values, where is significance level. However, when t^* does not have a symmetric distribution, the bootstrap critical values outlined above may perform quite poorly. This problem can be circumvented by an alternative method. For the two sided hypothesis testing, the upper bootstrap critical value $q^*_n(\alpha)$ is the $(1-\alpha)$ quantile of the distribution of $|t^*|$, and the lower bootstrap critical value is $-q^*_n(\alpha)$. Reject the null hypothesis if the test statistic of original data is between two bootstrap critical values. Otherwise, we cannot reject null hypothesis.

2 Bootstrap confidence intervals method

There are different types of bootstrap confidence intervals method available. Although we are looking for only two types:

1. Bootstrap percentile confidence interval method
2. Bootstrap-t confidence interval method

Suppose we have samples $\{X_1, X_2, ..., X_n\}$ the estimator $\hat{\theta} = \bar{X_n} = \sum_{i=1}^{n} \frac{X_i}{n}$, and we have

$$V_{\hat{F}_n}(\hat{\theta}) = \frac{\hat{\sigma}^2}{n} = \frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X_n})^2$$

let we want to generate B bootstrap samples then the bootstrap procedure follows:

1. For b = 1,2,3,...,B
2. Draw $X^*_1, ..., X^*_n \sim \hat{F}_n$
3. Compute $\hat{\theta}^*_b = T(X^*_1, ..., X^*_n)$
4. Compute $v_{boot} = \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}^*_b - \frac{1}{B}\sum_{c=1}^{B}\hat{\theta}^*_c)^2$

3 Bootstrap percentile confidence interval:

1. Draw $X^*_1, ..., X^*_n \sim \hat{F}_n$
2. Compute $\hat{^ast}_b = T(X^*_1, ..., X^*_n)$ for b = 1(1)B
3. Compute non-parametric confidence interval based on $\hat{}^*_b$ points
4. let $\hat{\theta}^*_{(r)}$ is the r-th order statistics
5. Therefore the $100(1-\alpha)\%$ confidence interval will be $[\hat{\theta}^*_k, \hat{\theta}^*_{k'}]$ where $k = [\frac{\alpha}{2}b]$ if $[\frac{\alpha}{2}b]$ is integer and $k = [\frac{\alpha}{2}b]+1$ if $[\frac{\alpha}{2}b]$ is not an integer. Similarly, $k' = [(1-\frac{\alpha}{2})b]$ if $[(1-\frac{\alpha}{2})b]$ is integer and $k' = [(1-\frac{\alpha}{2})b]+1$ if $[(1-\frac{\alpha}{2})b]$ is not an integer.

4 Bootstrap-t confidence interval

Before applying resampling to time series analysis, we look the simpler problem. Suppose we wish to construct a confidence interval for the time series data based on a given sample. One way to start with the so called t-statistic which is

$$t = \frac{\mu - \bar{X}}{\frac{s}{\sqrt{n}}}$$

where $\frac{s}{\sqrt{n}}$ is standard error of the mean.

If we are sampling from a normally distributed sample, then the probability distribution of t is known to be the t-distribution with n-1 degrees of freedom. We denote by $t_{\alpha/2}$ the $\alpha/2$ upper t-value, that is the $1-\alpha/2$ quantile of this distribution. Thus t has probability $\alpha/2$ of exceeding $t_{\alpha/2}$. Because of symmetry of the t-distribution, the probability is also $\alpha/2$ that t is less than $-t_{\alpha/2}$.

therefore for normally distributed data, the probability is $1-\alpha$ that

$$-t_{\alpha/2} \leq t \leq -t_{\alpha/2}$$

after substituting the value of t we get

$$\bar{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

which shows that

$$\bar{X} \pm \frac{s}{\sqrt{n}} t_{\alpha/2}$$

is a $1-\alpha$ confidence interval for μ, assuming normally distributed data.

What if we are not sampling from a normal distribution ?? In that case, the distribution of t not is t-distribution, but rather some other distribution that is not known to us. there are two problems. first we do not know the distribution of given sample. Second, even if sample distribution were known it is a difficult, ussually intractable, probability calculation to get the distribution of the t-statistic from the distribution of the given sample. This calculation has only been done for normal given sample. Considering the difficulty of these two problems, we can still get a confidence interval by resampling. we can take a large number, say B, of resamples from the original sample.

Let $\bar{X}_{boot,b}$ and $s_{boot,b}$ be the sample mean and standard deviation of the bth resample, b=1,2,...,B.

We define

$$t_{boot,b} = \frac{\bar{X} - \bar{X}_{boot,b}}{\frac{s_{boot,b}}{\sqrt{n}}}$$

The resamples are independent of each other, the collection $t_{boot,b}, t_{boot,b}, ...$ can be treated as a random sample from the distribution of the t-statistic. After B values of $t_{boot,b}$ have been calculated, one from each resample, we find the $100\alpha\%$ and $100(1-\alpha)\%$ percentiles of this collection of $t_{boot,b}$ values. Call these percentiles t_L and t_U. We find t_L and t_U as follows:

1. The B values of $t_{boot,b}$ are sorted from smallest to largest
2. Then we calculate $B\alpha/2$ and round to the nearest integer. suppose the result is K_L.

3. The k_Lth sorted value of $t_{boot,b}$ is t_L.
4. Similarly, let K_U be $B(1-\alpha/2)$, rounded to the nearest integer and then t_U is the K_Uth sorted value of $t_{boot,b}$.

If the original sample is skewed, then there is no reason to suspect that the $100\alpha\%$ percentile is minus the $100(1-\alpha)\%$ percentile, as happens for symmetric samples such as t-distribution. In other words, we do not necessarily expect that $t_L = t_U$. However, this fact causes us no problem, since the bootstrap allows us to estimate t_L and t_U without assuming any relationship between them. Now, we replace $-t_{\alpha/2}$ and $t_{\alpha/2}$ in the confidence interval by t_L and t_U, respectively.
Finally the bootstrap confidence interval for μ is :

$$(\bar{X} + t_L \tfrac{s}{\sqrt{n}}, \bar{X} + t_U \tfrac{s}{\sqrt{n}})$$

The bootstrap has solved both problems mentioned above. we do not need to know the give data distribution since we can estimate it by the sample. Also we do not need to calculate the distribution of the t-statistic using probability theory. Instead we simulate from this distribution.

We use the notation

$$SE = \tfrac{s}{\sqrt{n}}$$

and

$$SE_{boot} = \tfrac{s_{boot}}{\sqrt{n}}$$

****************** END *******************