3 The Block Bootstrap

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 $\{y_1, y_2, \ldots, y_l\}$, block 2 being $\{y_2, y_3, \ldots, y_{l+1}\}, \cdots, etc.$ For the non-overlapping method, we divide the data into T/l blocks, which block 1 being $\{y_1, y_2, \ldots, y_l\}$, block 2 being $\{y_{l+1}, y_{l+2}, \ldots, y_{l+l}\}, \cdots, etc.$ 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 $t = 1, 2, 3, \ldots, T$, i is iid uniform random variable on $\{1, 2, 3, \ldots, T-l+1\}$ and $j = 1, 2, 3, \ldots, l$.

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

$$\widehat{\beta}^*_{TSLS} = (X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}X^*)^{-1}(X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}y^*), \tag{11}$$

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

$$W_T^* = (\frac{1}{T} \sum_{t=1}^{T} \widehat{g}_t^* \widehat{g}_t^{*\prime})^{-1}, \tag{12}$$

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

$$\widehat{\beta}_{GMM}^* = (X^{*\prime} Z^* W_T^* Z^{*\prime} X^*)^{-1} (X^{*\prime} Z^* W_T^* Z^{*\prime} y^*), \tag{13}$$

$$(\widehat{\sigma}^*)^2 = (X^{*\prime} Z^* W_T^* Z^{*\prime} X^*)^{-1}. \tag{14}$$

In the sense, the bootstrap GMM estimator $\widehat{\beta}_{GMM}^*$ is a consistent estimator of β. However, because the bootstrap sample can not satisfy the same moment condition as the population distribution, it fails to achieve an asymptotic refinement. A correction 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

$$\widehat{\beta}_{GMM}^* = (X^{*\prime} Z^* W_T^* Z^{*\prime} X^*)^{-1} (X^{*\prime} Z^* W_T^* Z^{*\prime} Y^* - Z' \widehat{u})), \tag{15}$$

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

$$t^* = \frac{\widehat{\beta}_{GMM}^* - \beta}{\widehat{\sigma}^*}. \tag{16}$$

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 $(1-\alpha/2)$ quantile and the $(\alpha/2)$ quantile of the distribution of t^* for our bootstrap critical values, where α 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 $-q_n^*(\alpha)$ (Hansen, 2004). Reject the null hypothesis if the test statistic of original data is between two bootstrap critical values. Otherwise, we cannot reject null hypothesis.