

## CHAPTER 4

### RESULTS

An R (2.5.1) program was used to produce simulation results for the new step-down dependent bootstrap min P procedures for comparing several means with a control and compare the efficiency of the new step-down dependent bootstrap min P with the traditional step-down bootstrap min P and Dunnett's  $t$  two-sided test statistic. The simulated data were sampled from a homogeneous normal distribution. The simulation design considered 3 treatment groups and 1 control group. Each group had an equal sample size,  $r$ , equal to 3, 4, 5, ..., 10.

For the step-down procedures i.e., step-down independent bootstrap min P, step-down dependent bootstrap min P with 2 copies of the sample data and step-down independent bootstrap min P with 4 copies of the sample data, the number of bootstrap resamples,  $B$ , was set equal to be 100, 1,000 and 10,000.

Monte Carlo simulation with 1,000 number of repetitions were realized. The criteria for the efficiency comparisons of each procedure is determined by a pre-specified and acceptable significance level  $\alpha = 0.05$  for the Type I error and to seek the test that has maximum power of the test within the class of tests with Type I error rates at most  $\alpha$ .

The best procedure should have empirical Type I error rates within the interval  $[0.036, 0.064]$  and have the highest empirical power of the test at significance level  $\alpha$ .

To obtain the empirical Type I error rates for each procedure, we generated sample data from a  $N(100, 25)$  for each treatment group. The empirical power of the test was obtained by generating sample data from a  $N(60, 25)$  distribution for a control group and from a  $N(100, 25)$  distribution for the other treatment groups.

Simulation results in this research can be divided into 2 parts: (1) empirical type I error rates and (2) empirical power of the test.

#### 4.1 Empirical Type I Error Rates

Empirical type I error rates of all procedures are shown in table 28. The first, second and third panel of table 28 provide the results for the empirical type I error rates of all procedures based on the number of bootstrap resamples,  $B$ , equal to 100, 1,000 and 10,000 and where the sample size,  $r$ , is equal to 3, 4, 5, ..., 10.

**Table 28: Empirical Type I Error Rates of All Procedures**

<b>B</b>	<b>r</b>	<b>TYPE I ERROR RATES</b>				
		<b>Dunnett</b>	<b>IN</b>	<b>DE2</b>	<b>DE4</b>	
100	3	0.055	0.054	0.056	0.049	
	4	0.049	0.045	0.045	0.041	
	5	0.055	0.060	0.050	0.046	
	6	0.062	0.056	0.064	0.059	
	7	0.057	0.057	0.058	0.054	
	8	0.043	0.047	0.044	0.041	
	9	0.047	0.045	0.042	0.048	
	10	0.054	0.059	0.050	0.050	
	1,000	3	0.044	0.042	0.047	0.048
		4	0.048	0.050	0.046	0.048
5		0.047	0.050	0.046	0.049	
6		0.043	0.043	0.045	0.044	
7		0.047	0.048	0.047	0.050	
8		0.049	0.049	0.046	0.050	
9		0.045	0.039	0.043	0.042	
10		0.052	0.053	0.052	0.056	
10,000		3	0.049	0.048	0.046	0.048
		4	0.055	0.052	0.053	0.054
	5	0.055	0.054	0.053	0.053	
	6	0.042	0.043	0.043	0.041	
	7	0.055	0.055	0.055	0.056	
	8	0.050	0.050	0.050	0.050	
	9	0.055	0.054	0.054	0.055	
	10	0.049	0.051	0.049	0.051	

**Note:** Dunnett: Dunnett's  $t$  statistic two-sided.

IN: Step-down independent bootstrap min P.

DE2: Step-down dependent bootstrap min P with 2 copies.

DE4: Step-down dependent bootstrap min P with 4 copies.

Referring to table 28, empirical type I error rates of Dunnett, IN, DE2 and DE4 range from 0.036 to 0.064 for the number of bootstrap resamples,  $B$ , equal to 100, 1,000 and 10,000 and where the sample sizes,  $r$ , equals 3, 4, 5, ..., 10. For example, if  $r$  equals 3 and  $B$  equals 10,000 then the empirical type I error rates of Dunnett, IN, DE2 and DE4 are 0.049, 0.048, 0.046 and 0.048, respectively. From our criteria, all procedures are able to control type I error rates at the significance level 0.05 for the number of bootstrap resamples equals 100, 1,000 and 10,000 and where the sample size equals 3, 4, 5, ..., 10.

## **4.2 Empirical Power of the Test**

The empirical power of the test of Dunnett, IN, DE2 and DE4 is shown in table 29. Table 29 provides the results of empirical power of the test for all procedures based on the number of bootstrap resamples,  $B$ , equal to 100, 1,000 and 10,000 and where the sample size,  $r$ , is equal to 3, 4, 5, ..., 10.

**Table 29: Empirical Power of the Test Classified by the number of bootstrap resamples**

<b>B</b>	<b>r</b>	<b>POWER OF THE TEST</b>			
		<b>Dunnett</b>	<b>IN</b>	<b>DE2</b>	<b>DE4</b>
100	3	0.442*	0.405	0.424	0.402
	4	0.558	0.544	0.557	0.560*
	5	0.746*	0.721	0.696	0.723
	6	0.802*	0.785	0.777	0.769
	7	0.867	0.869*	0.856	0.851
	8	0.932*	0.916	0.916	0.920
	9	0.951*	0.944	0.948	0.946
	10	0.971*	0.962	0.962	0.958
1,000	3	0.407	0.401	0.395	0.408*
	4	0.613*	0.605	0.601	0.610
	5	0.732*	0.718	0.726	0.728
	6	0.793	0.785	0.794*	0.788
	7	0.888	0.888	0.892*	0.880
	8	0.914*	0.912	0.912	0.913
	9	0.939	0.941*	0.940	0.938
	10	0.970	0.972*	0.970	0.968
10,000	3	0.400	0.401	0.403*	0.401
	4	0.578	0.577	0.574	0.579*
	5	0.723	0.722	0.723	0.725*
	6	0.812	0.811	0.814*	0.811
	7	0.898*	0.898*	0.896	0.898*
	8	0.916	0.917*	0.917*	0.914
	9	0.943	0.941	0.944*	0.944*
	10	0.962	0.961	0.964	0.965*

**Note:** Dunnett: Dunnett's  $t$  statistic two-sided.

IN: Step-down independent bootstrap min P.

DE2: Step-down dependent bootstrap min P with 2 copies.

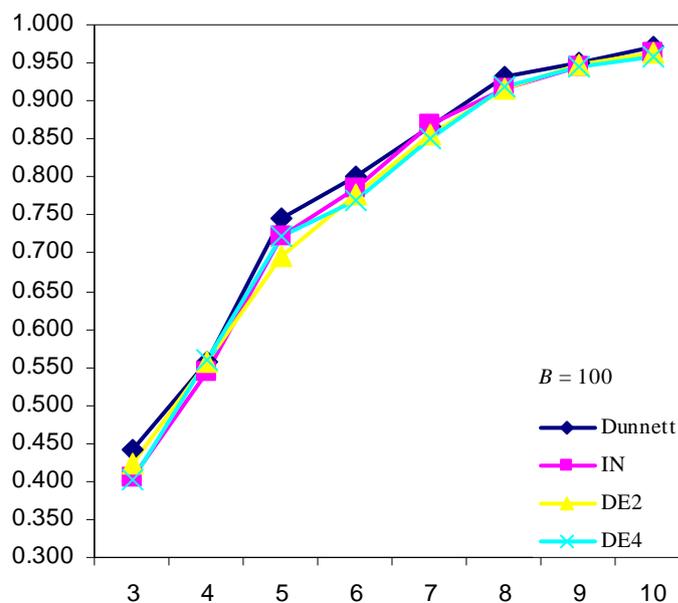
DE4: Step-down dependent bootstrap min P with 4 copies.

\*: The highest empirical power of the test.

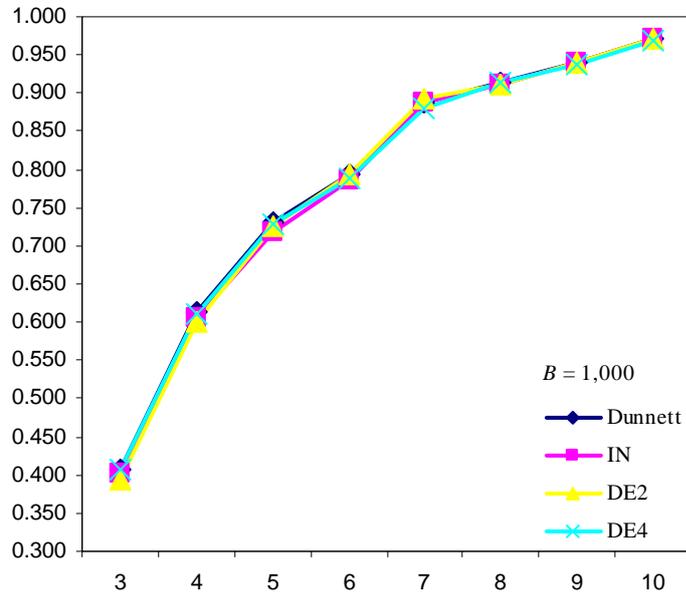
For  $B = 100$ , Dunnett procedure has the highest empirical power for a step-down procedure but not for  $r = 4$  and  $7$ .

However, for  $B = 1,000$ , Dunnett's procedure has the highest empirical power for a step-down procedure for  $r = 4, 5, 8$  while the IN procedure has the highest empirical power for  $r = 9, 10$  while the DE2 procedure has the highest empirical power for  $r = 6, 7$  and the DE4 procedure has the highest empirical power for  $r = 3$ .

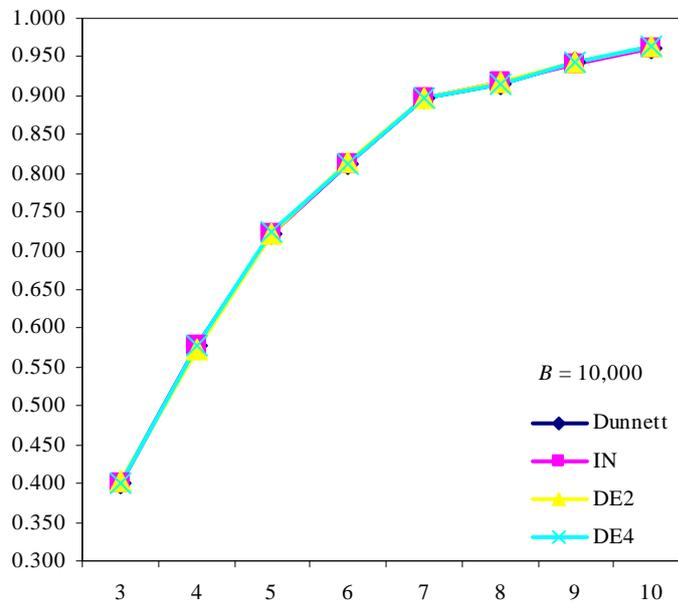
The last panel of table 29 displays the result of the empirical power of all tests, based on the number of bootstrap resamples equal to 10,000. The DE4 procedure provides the highest empirical power of the test in all cases for sample sizes equal to 4, 5, and 10. The DE2 procedure provides the highest the empirical power of the test in all cases for sample sizes equal to 3 and 6. The DE2 and DE4 procedures provide similar empirical power of the test which is equal to 0.944 when the sample size is equal to 9. The IN and DE2 procedures provide the highest empirical power of the test for sample size equal to 8. The Dunnett, IN and DE4 procedures provide the highest the empirical power of the test which is equal to 0.898 for sample size equal to 7.



**Figure 7: Empirical Power of the Test of All Procedures, Based on  $B = 100$**



**Figure 8: Empirical Power of the Test of All Procedures, Based on  $B = 1,000$**



**Figure 9: Empirical Power of the Test of All Procedures, Based on  $B = 10,000$**

Table 30 shows the empirical power of the test of the step-down procedures for the sample sizes,  $r$ , equal to 3, 4, 5, ..., 10 and where the number of bootstrap resamples equals 100, 1,000 and 10,000.

**Table 30: Empirical Power of the Test Classified by the Sample Size**

r	PROCEDURE	THE NUMBER OF RESAMPLIE		
		100	1,000	10,000
3	IN	0.405	0.401	0.401
	DE2	0.424	0.395	0.403
	DE4	0.402	0.408	0.401
4	IN	0.544	0.605	0.577
	DE2	0.557	0.601	0.574
	DE4	0.560	0.610	0.579
5	IN	0.721	0.718	0.722
	DE2	0.696	0.726	0.723
	DE4	0.723	0.728	0.725
6	IN	0.785	0.785	0.811
	DE2	0.777	0.794	0.814
	DE4	0.769	0.788	0.811
7	IN	0.869	0.888	0.898
	DE2	0.856	0.892	0.896
	DE4	0.851	0.880	0.898
8	IN	0.916	0.912	0.917
	DE2	0.916	0.912	0.917
	DE4	0.920	0.913	0.914
9	IN	0.944	0.941	0.941
	DE2	0.948	0.940	0.944
	DE4	0.946	0.938	0.944
10	IN	0.962	0.972	0.961
	DE2	0.962	0.970	0.964
	DE4	0.958	0.968	0.965

**Note:** IN: Step-down independent bootstrap min P.

DE2: Step-down dependent bootstrap min P with 2 copies.

DE4: Step-down dependent bootstrap min P with 4 copies.

The number of bootstrap resamples does not obviously affect the empirical power of the test of the step-down procedures as shown in table 30.

When  $r = 3$ , the IN and DE2 procedures have the highest empirical power for  $B = 100$  while the DE4 procedure has the highest empirical power for  $B = 1,000$ .

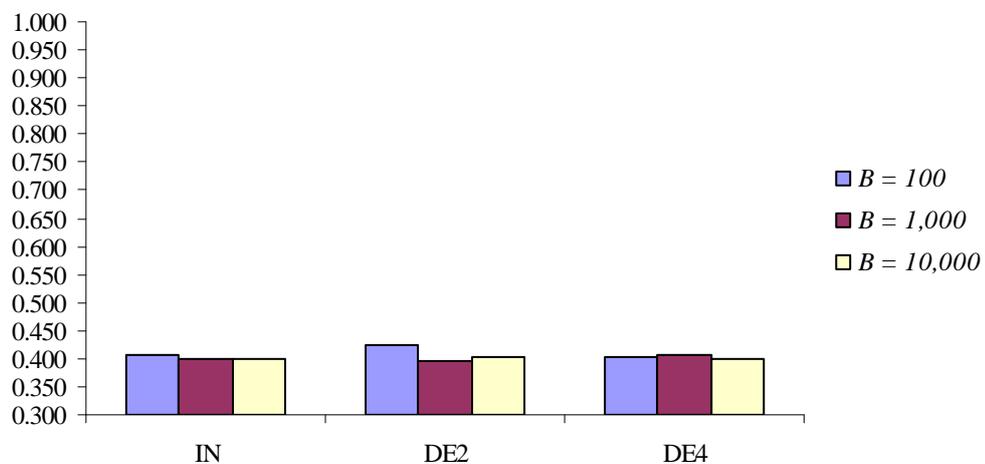
When  $r = 4, 10$  with all methods, have the highest empirical power occurs for  $B = 1,000$ .

When  $r = 5$ , the IN procedure has the highest empirical power for  $B = 10,000$  while the DE2 and DE4 procedures have the highest empirical power for  $B = 1,000$ .

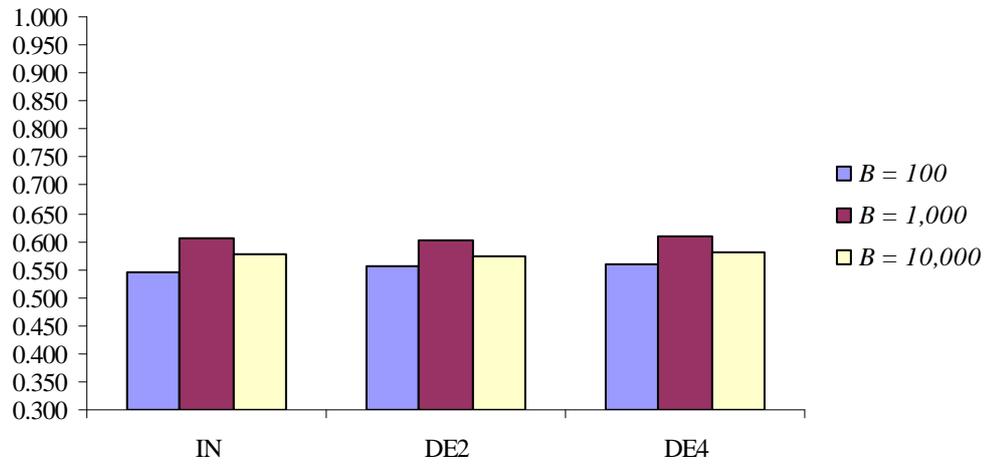
When  $r = 6, 7$  with all methods, the highest empirical power occurs for  $B = 10,000$ .

When  $r = 8$ , the DE4 procedure has the highest empirical power for  $B = 100$  while the IN and DE2 procedures have the highest empirical power for  $B = 10,000$ .

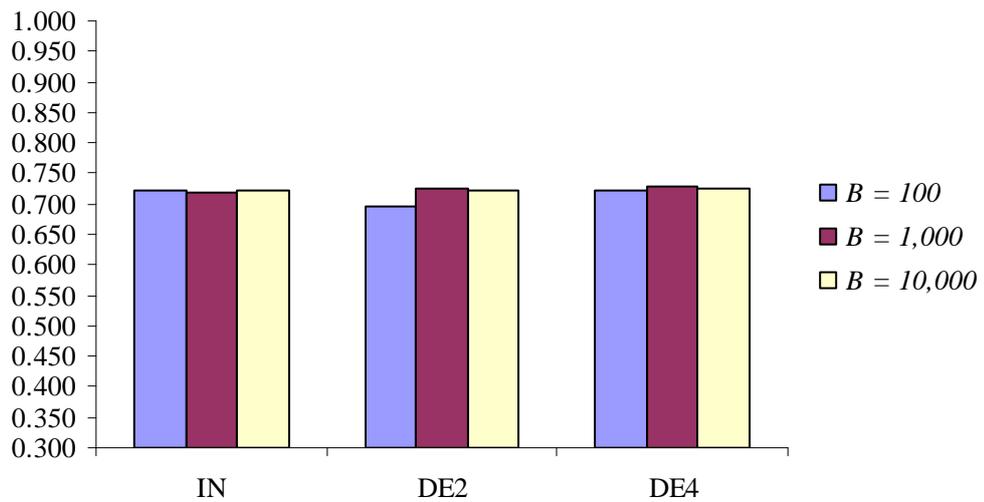
When  $r = 9$ , with all methods, have the highest empirical power occurs for  $B = 100$ .



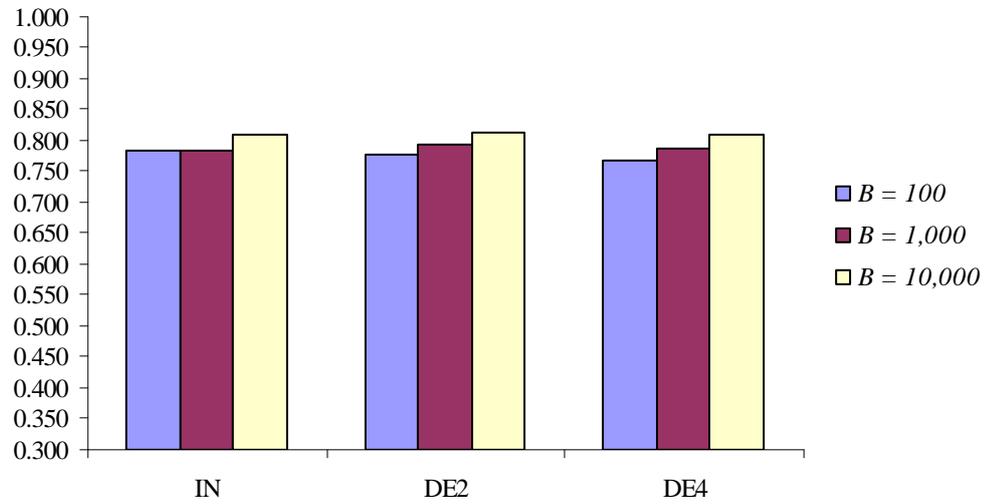
**Figure 10: Empirical Power of the Test of Step-down Procedures when the Sample Size  $r = 3$**



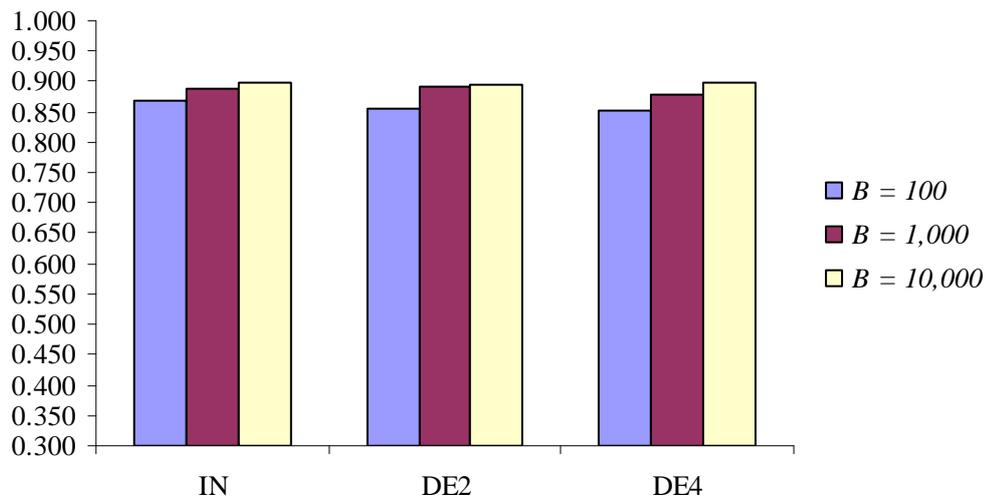
**Figure 11: Empirical Power of the Test of Step-down Procedures when the Sample Size  $r = 4$**



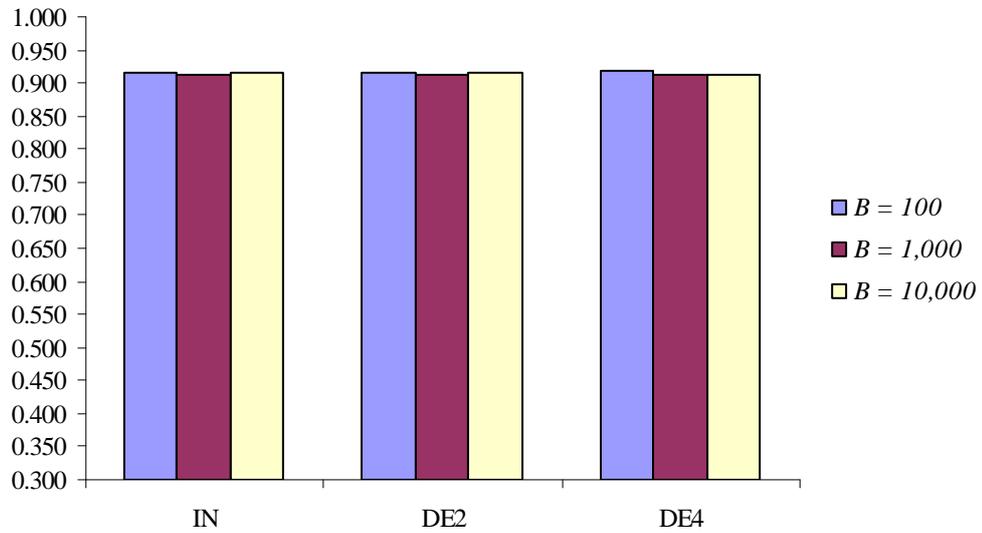
**Figure 12: Empirical Power of the Test of Step-down Procedures when the Sample Size  $r = 5$**



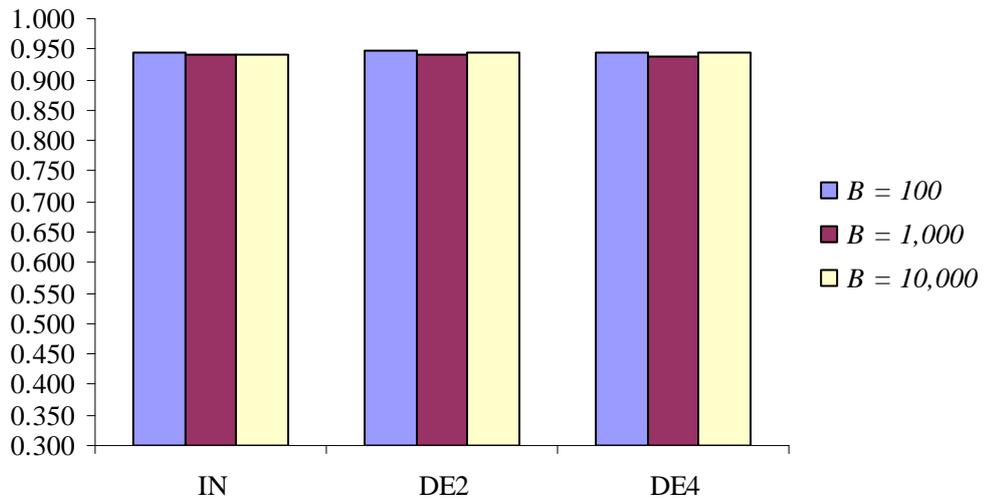
**Figure 13: Empirical Power of the Test of Step-down Procedures when the Sample Size  $r = 6$**



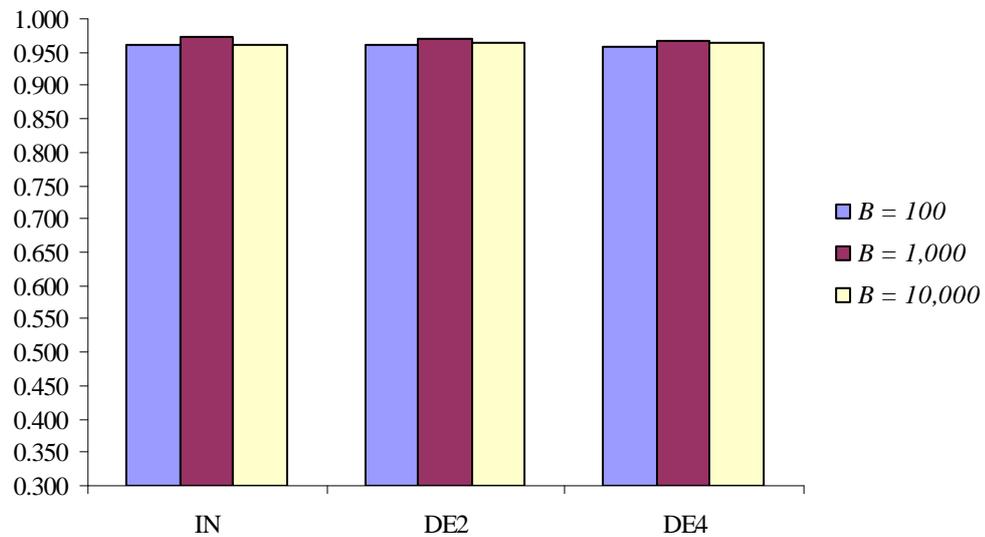
**Figure 14: Empirical Power of the Test of Step-down Procedures when the Sample Size  $r = 7$**



**Figure 15: Empirical Power of the Test of Step-down Procedures when the Sample Size  $r = 8$**



**Figure 16: Empirical Power of the Test of Step-down Procedures when the Sample Size  $r = 9$**



**Figure 17: Empirical Power of the Test of Step-down Procedures when the Sample Size  $r = 10$**

As expected, the step-down dependent bootstrap min P procedures, DE2 and DE4, are not worse than other two procedures for most simulations, and the power of all procedures also increases significantly along with the sample size. However, the simulation results show unclear evidence when the number of bootstrap resamples ( $B$ ) or the number of copy the sample data ( $c$ ) could give the highest empirical power in all cases.