INSTITUTO NACIONAL DE ESTADÍSTICA

Performance of Bootstrap Techniques with Imputed Data

Shao and Sitter's method (1)

a) Let $Y_I = Y_R \cup Y_O$ where $Y_R = \{y_k : k \in A_R (\text{respondent})\}$ and $Y_O = \{z_k : k \in A_O (\text{non-respondent})\}$. Y_O is obtained from Y_R using some imputation technique.

b) Let $Y^* = \{y_i^* : i = 1,..n\}$ a simple random sample(bootstrap sample) drawn with replacement from Y_I and $Y_I^* = Y_R^* \cup Y_O^*$ where $Y_R^* = \{y_k^* : k \in A_R^* (\text{respondent})\}$ and $Y_O^* = \{y_k^* : k \in A_O^* (\text{non-respondent})\}$. Y_O^* is obtained from Y_R^* using the same imputation technique that was used in step .

Shao and Sitter's method(2)

c) Obtain the bootstrap estimator $\hat{\theta}_I^* = \hat{\theta}(Y_I^*)$ of $\hat{\theta}_I = \hat{\theta}(Y_I)$, based on the imputed bootstrap data set Y_I^*.

d) Repeat steps b) and c) B times.The following bootstrap variance estimator for $\hat{\theta}_I$ is calculated:

$$v_B(\hat{\theta}_I) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_I^{*b} - \vec{\theta}_I^*)^2$$

where

$$\vec{\theta}_I^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_I^{*b}$$

•The percentile confidence intervals of the statistic is computed as the empirical confidence interval obtained after sorting the B bootstrap statistics.

•To build the bootstrap-t confidence intervals, we first compute the statistics $t_b^* = (\hat{\theta}_I^{*b} - \overleftarrow{\theta}_I^*) / \sigma_b^*$ where $\sigma_b^{*2} = v_B(\hat{\theta}_I^{*b})$

and then calculate $t_L^* = C\hat{D}F_t^{-1}(\alpha),\ t_U^* = C\hat{D}F_t^{-1}(1-\alpha),$

where $C\hat{D}F_t^{-1}(x) = \#\left(t_b^*;\ t_b^* \leq x,\ b = 1....B\right)/B$ The bootstrap-t confidence interval is then given by

$$\left(\hat{\theta}_I - t_U^* v_B(\hat{\theta}_I), \hat{\theta}_I - t_L^* v_B(\hat{\theta}_I)\right)$$

Montecarlo Study and Results(1)

– From our population of N=16,438 industrial businesses we draw simple random samples without replacement of sizes n=100, 500, 1000 and 5000. We simulate a loss of about 30 per cent of our data with a uniform mechanism. The number of replications is 50000 for each sample size. The number of bootstrap samples is 999 for each replication.

INSTITUTO NACIONAL DE ESTADÍSTICA

Montecarlo Study and Results(2)

- An additional simulation with 23000 replications was conducted only for small sample sizes (n=100) calculating the bootstrap-t confidence interval. We used 999 bootstrap samples in the first level and 50 bootstrap samples in the second level (variance estimation of the first level bootstrap statistics).

Montecarlo Study and Results(3)

- The analysis variable is the turnover of the businesses. We use mean imputation.
- Within each replication, we compute the percentage relative bias, the relative root mean square error and the coverage of a nominal 95% bootstrap confidence interval.

INSTITUTO NACIONAL DE ESTADÍSTICA

TABLE 1. Imputed variable: turnorver.

Imputation method: mean imputation, 95% confidence intervals

Sample size	BOOTSTRAP T			BOOTSTRAP Percentile			JACKKNIFE		
	RB(%)	MSE	COVR(%)	RB(%)	MSE	COVR(%)	RB(%)	MSE	COVR(%)
100	0.07	4.51	56.9	-0.42	4.47	60.6	0.29	4.52	56.9
500				1.00	2.00	68.3	-0.90	1.96	65.8
1000				4.24	1.45	75.6	-2.12	1.35	72.6
5000				27.15	0.76	91.9	-12.0	0.51	87.3

Conclusions(1)

– The results show that the percentile bootstrap performs better than the jackknife for coverage rate of the confidence intervals and the reverse is true for mean square errors and bias of the variance estimators.

Conclusions(2)

– There seems not to be any advantage in using the bootstrap-t confidence interval, in spite of its higher computational time, though only small sample simulations and small number of second level bootstrap samples are considered for this method.