

BOOTSTRAP ALGORITHMS FOR RISK MODELS WITH AUXILIARY VARIABLE AND COMPLEX SAMPLES

Algoritmi Bootstrap per la Modellizzazione del Rischio in Presenza di Variabile Ausiliaria e Campionamento Complesso

Giancarlo Manzi

Dip.di Scienze Economiche, Aziendali e Statistiche, Università di Milano

giancarlo.manzi@unimi.it

Fulvia Mecatti

Dip. di Statistica, Università di Milano-Bicocca, fulvia.mecatti@unimib.it

Key words: Probability-proportional-to-size Sampling, Ratio Estimator, Regression Estimator, Resampling Methods.

Statistical modelling of the operational risk in the financial sector as well as in other private and public sectors is receiving increasing attention (De Fontnouvelle et al. 2003; Nystrom and Skoglund, 2002; Risk Management Group, 2003). Recent financial scandals leading to huge bank losses and the case of the re-nationalisation of the British railways company *Rail-Track* in 2002 are emblematic examples often cited in literature. Since the operational risk management influences strategies, decisions and policies, even in non-banking fields the *stability* of estimators of the model parameters represents a judicious issue. On the other hand the estimation framework shows many difficulties, for instance data are often collected under complex sampling plans and estimators are usually complex so that their accuracy can not be analytically assessed. In this context, resampling methods for estimating the variance of estimators of the model parameters and for constructing confidence intervals appear for a natural solution. Suitable modifications of the naïve Efron's bootstrap are needed for addressing the *non-iid* characteristics above. In the present work, bootstrap algorithms for finite population sampling are considered in presence of auxiliary variable, *i.e.* past experience, and of complex samples. A simple ratio model, *i.e.* a linear regression without intercept, with heteroscedastic errors is adopted as generating the N population values

$$y_i = \beta x_i + \varepsilon_i \quad (1)$$

$i = 1 \dots N$, where ε_i are independent random variables with $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = \sigma_i^2$. We assume that a sample s of size n is selected under a sampling design with probability proportional to the auxiliary variable and without replacement (πPS). Note that this enables a wide and more complete use of the auxiliary variable. Two classic estimators, the ratio and the regression estimator, are considered:

$$\hat{\beta}_{RA} = \sum_{i \in s} y_i / \sum_{i \in s} x_i \quad \text{and}$$

$\hat{\beta}_{RE} = \sum_{i \in s} y_i x_i / \sum_{i \in s} x_i^2$. In order to estimate the estimators variance, the so-called *wild bootstrap* is generally proposed under simple random sampling (Helmert and Wegkamp, 1998). Under πPS we consider three bootstrap algorithms relying upon a bootstrap population to perform the resampling step. Bootstrap population is defined as a set of sampled units suitably replicated so that the features of the actual population generating data are acknowledged at a possibly large extent.

Starting from the Holmberg's bootstrap (Holmberg, 1998), for which the bootstrap population results calibrated with respect to both the size N and the total $\sum_{i=1}^N x_i$ of the auxiliary variable and which can be interpreted as *design based*, we propose two modifications:

- i) a simplified Holmberg's bootstrap, still design based, by performing the resampling according to Mecatti (2000);

- ii) a *model assisted* πPS bootstrap, by assuming $\varepsilon_i = x_i^{1/2} u_i$ in model (1) and by estimating the finite population distribution according to Rao, Kovar and Mantel (1990). Note that a ratio model is implicitly assumed when a πPS sampling design is chosen. The proposed model assisted πPS bootstrap retains the design based properties by the original

Holmberg's bootstrap and it is also calibrated with respect to the model assumption $y_i \propto x_i$.

A simulation study is conducted with the twofold purpose of investigating the empirical performance of the πPS bootstrap algorithms and of comparing them with the wild bootstrap, *i.e.* the methodology available in literature.

The simulation considers an artificial set up according to McCarthy and Snowden (1985) where estimators $\hat{\beta}_{RA}$ and $\hat{\beta}_{RE}$ are expected to have optimal analytical properties. Monte Carlo estimator distributions are provided under the Rao-Sampford rejective πPS design, for different values of finite sample size n . Empirical properties of the bootstrap variance estimators are evaluated by using the Average Relative Error (Lahiri, 2003)

$$ARE[V_*(\hat{\theta})] = 100 \times \frac{E_{mc}[V_*(\hat{\theta}^*)] - V_{mc}(\hat{\theta})}{V_{mc}(\hat{\theta})} \quad (2)$$

where $\hat{\theta}$ denotes the estimator from the original sample, * denotes bootstrap quantities, expectations and variances are provided via Monte Carlo.

Partial results from some preliminary simulations show that the *simplified* Holmberg's bootstrap performs better than the original Holmberg's bootstrap, showing an overall lower ARE which improves as the sample size increases, especially for the ratio estimator $\hat{\beta}_{RA}$. Preliminary simulation results are also encouraging with respect to the performance of the *model assisted* πPS bootstrap for both the estimators $\hat{\beta}_{RA}$ and $\hat{\beta}_{RE}$. Future empirical work shall consider comparisons with the *wild* bootstrap and real data sets.

References

- De Fontouville P., DeJesus-Rueff V., Jordan J. and Rosengren E. (2003), *Capital and Risk : New Evidence on Implications of Large Operational Losses*, Federal Reserve, Bank of Boston.
- Holmberg A. (1998), A Bootstrap Approach to Probability Proportional to Size Sampling, *Proceedings of Section on Survey Research Methods*, American Statistical Association, 378-383.
- Lahiri D. (2003), On the Impact of Bootstrap in Survey Sampling and Small-Area Estimation, *Statistical Science*, 18, 199-210.
- McCarthy P.J. and Snowden C.B. (1985), The Bootstrap and Finite Population Sampling, in *Vital and Health Statistics*, Series 2, 95, Public Health Service Publication, Washington DC, U.S. Government Printing Office, 1-23.
- Mecatti F. (2000), Bootstrapping Unequal Probability Samples, *Statistica Applicata*, 12, 1, 67-77.
- Nystrom K. And Skoglund J. (2002), *Qualitative Operational Risk Management*, Swedbank, Group Financial Risk Control, Working Paper.
- Rao J.N.K, Kovar J.G., Mantel H.J. (1990), On Estimating Distribution Function and Quantiles from Survey Data Using Auxiliary Information, *Biometrika*, 77, 365-375.
- Risk Management Group (2003), *The 2002 Loss Data Collection Exercise for Operational Risk: Summary of the Data Collected*, Report to the Basel Committee on Banking Supervision, Bank for International Settlement <http://www.bis.org>
- Sampford M.R. (1967), On Sampling Without Replacement with Unequal Probabilities of Selection, *Biometrika*, 54, 499-513.