BOOTSTRAP INTENSIVE METHODS FOR UNDERSTANDING EMPIRICAL LOSS DATA AND DYNAMIC FINANCIAL ANALYSIS

ALBEANU, Grigore
Spiru Haret University, Bucharest, Romania
g.albeanu.mi@spiruharet.ro

MADSEN, Henrik
DTU Informatics, Denmark
hm@imm.dtu.dk

GHICA, Manuela
Spiru Haret University, Bucharest, Romania
m.ghica.mi@spiruharet.ro

POPENȚIU-VLĂDICESCU, Florin
University of Oradea, UNESCO IT Chair
popentiu@imm.dtu.dk

Abstract

Bootstrap Intensive Methods for estimation assessment provide valuable information concerning the adequacy of applied probabilistic models. The bootstrap method is an extensive computational approach for understanding empirical data and is based on resampling and statistical estimation. It is a powerful tool, especially when only a small data set is used to predict the behaviour of systems or processes. This paper describes some case studies based on the Efron type bootstrap approaches for modelling loss distributions and for general dynamic financial analysis. The case studies are inspired from risk management field. The research is based on theoretical previous developments in accuracy assessment, reliability estimation and risk exchange modelling.

Keywords: *bootstrap, loss distributions, dynamic functional analysis*
AMS classification: 62P20

1. **Introduction**

Data analysis in finance, banking and insurance is an important topic, not only for academic area, but also for business. Investigating the requirements for building software to assess the accuracy for business statistical applications we found that bootstrap is a powerful technique.

Starting with some previous experience [1, 2, 3, 10, 11] and the available scientific literature, this paper describes the basics of bootstrap methodology, the principle of the dynamic financial analysis, and identifies some applications of bootstrapping for risk management in business.

The remaining part of this paper is organized as follows. Section 2, firstly, introduces elements of bootstrap methodology to obtain the distribution of a random variable, to provide a functional relation, and to estimate the accuracy of a statistics. A general framework for obtaining simultaneous confidence bands for non-linear explicit regression models using resampling techniques is described, next. Loss distribution analysis is presented in order to show the importance of the loss distributions for the dynamic functional analysis (DFA) treated in the last part of the section. The third section covers an experiment in bootstrapping loss models and identifies some application for controlling uncertainties in the case of DFA applied in insurance. Lastly, the fourth section gives a short conclusion and provides a view for future investigations.

2. **The bootstrap principle**

Bootstrap is a simple but powerful Monte Carlo method to assess statistical accuracy or to estimate a distribution from sample's statistics. The name may come from phrase "Pull yourself up by your bootstraps" having the following meaning 'Improve your situation by your own efforts' [21].

The methods are suitable for any level of modelling being useful for fully parametric, semi-parametric, and completely nonparametric analysis. These approaches are not only in use by statisticians, but also are applied anywhere statistics can be used: life sciences, business, social sciences, econometrics, reliability etc. For the aim of this paper we outline the basic bootstrap principle, the application of bootstrap sampling for accuracy estimation and the method of simultaneous confidence bands for uncertainty management.

Let X be a random variable and F the cumulative distribution function of the variable X. The Bootstrap method, proposed by Efron [7, 8], it is useful, at least, for the estimation of: the distribution function of a random variable R(X, F); a functional relation V(F), or the accuracy of a statistics s obtained from a sample $(X_1, X_2, ..., X_n)$ of size n from X. During this presentation, the accuracy, describes the variability of s when independent estimations s(1), s(2), ..., of the statistics s, are obtained by resampling.

The bootstrap technique uses the sample $(X_1, X_2, ..., X_n)$ to obtain the sampling cumulative distribution function $F_n(x)$ in order to replace the true cumulative distribution function F: $F_n(x) = (1/n)$ cardinal $\{ x_i \le x; 1 \le i \le n \}$. To repeatedly simulate bootstrap samples $X^*:=(X_1^*, X_2^*, ..., X_n^*)$ from F_n, random number generators should be used according to the Monte Carlo approaches. Then, for each bootstrap sample, it is recalculated: the distribution function of the random variable $R(X^*, F_n)$; the functional relation $V(F_n)$ or $V(F_n^*)$; the statistics s*( ). The accuracy of the statistics s can be derived under an appropriate statistical inference study on the sequence s*().

The bootstrap resampling can be realised in various ways. Uniform resampling and the importance resampling are the mostly used [6, 9, 12]. Uniform resampling assumes that the measurement (observed) values are uniformly sampled from some process. The importance resampling algorithm assumes the generation of sampling values according to a probability distribution $\{(x_i, p_i): i = 1, 2, ..., n\}$ such as every p_i is a nonnegative real number, and $p_1 + p_2 + ... + p_n = 1$ [1, 2, 3].

As a common example of the usage of the uniform resampling, we refer to the bootstrap algorithm for estimating standard errors. However, when some observations are more important than others, the importance resampling can provide close to real conclusions. If resampling is based on importance resampling weights, then the bootstrap estimates are reweighted as if uniform resampling is done.

3. **Loss distribution models**

Measuring the operational risk of a financial institution requires the Loss Distribution Approach (LDA), which makes use of the exact operational loss frequency and severity distributions [19]. Mainly, the LDA has three essential components: a distribution of the annual number of losses (the frequency), a distribution of the monetary losses (the severity), and an aggregate loss distribution that combines the two. The first step in applying LDA method is dedicated to the understanding of the structure and characteristics of the data. According to [20], a careful Exploratory Data Analysis (EDA) is necessary to be performed before modelling the shape of the data by statistical tools.

The most used approach in building loss distribution models consists in fitting the parametric distributions to the loss severity and the frequency data [13, 15]. For modelling there are available both simple parametric distributions having one to three parameters: Exponential, Weibull, Gamma, Truncated Lognormal, Loglogistic and Generalized Pareto, and generalized parametric distributions having three or more parameters: the generalized beta distribution of the second kind and the G-and-H distribution.

The estimation of the parameters is based on some well-known techniques as maximum likelihood, method of moments, or quantile estimation. Other techniques in modelling severity distributions are based on Extreme Value Theory (EVT) and non-parametric empirical sampling. EVT is a branch of statistics to study the extreme phenomena (large operational losses, for example).

Empirical sampling simulates losses from the empirical distribution in the following manner. For each simulated period:

1) Draw a frequency N from the Poisson distribution
2) Generate N loss severities (with replacement) using the original set of loss severity data: L_i ($i = 1, 2, \ldots, N$)
3) Sum the N losses to get the total annual loss: $S = \sum \{L_i \mid i = 1, 2, \ldots, N\}$.
4) Repeat steps 2 and 3 for many times (usually, one million).

The distribution of S (according to LDA) is called the aggregate loss distribution, and the risk exposure can be measured as a quantile of S. It is clear that bootstrap approach can be used for different tasks when LDA is performed.

4. **Dynamic financial analysis**

Dynamic Financial Analysis (DFA) is an important development based on stochastic simulation (Monte Carlo methods [9]) used mostly in non-life insurance and reinsurance [4, 5]. However, for life insurance, the approach called Asset Liability Management (ALM) if using stochastic simulation becomes similar to DFA. It is accepted now that DFA is a variant of ALM, showing a greater emphasis on both economic scenario generators and the interrelationships between assets and liabilities. Applying this approach the insurance and reinsurance companies are able to investigate the potential impact during decisional process.

According to [14], the DFA borrows many concepts and methods from economics and statistics and integrates them in a powerful tool (actually implemented in software as decision support systems). DFA requires a scenario generator and a calibration procedure. Also is mandatory a multivariate company model, an analysis and presentation module together with a control and optimisation module for improving the strategy [5].

The scenario generator is a module which implements stochastic models for risk factors, affecting the company strategic decisions, like economic risks, liability risks, asset risks and business risks. On the first step in developing a DFA model it will be investigated the risks and the factors affecting the company results. Then, the objective functions and the projection period (the planning horizon, usually long enough) have to be chosen.

The calibration procedure is responsible for finding the suitable parameters of the models used in the scenarios generation. The difference between DFA and classical scenarios testing approach results from the usage of Monte Carlo simulation (including bootstrap) during scenarios generation and calibration.

The output of the scenario generator is a large number of Monte Carlo scenarios representing the future possible states of the company, while only a small set of scenarios are used in classical scenario testing approach. The company model has to react to the generated scenario according to the company operating structure along aspects like insurance, investment, the impact of reinsurance, under specific regulation, accounting and taxation. Company models used in DFA vary from simple to highly complex size. Usually, such models imitate the cash flows of the company (mainly the technical and financial accounting structures) in a risk management manner.

Running DFA analysis produces results for the output variables under interest along some future moments of time (predicted values). The size of the obtained set of results is, in general, huge enough to ask for data mining techniques used by the analysis procedure. The common approach in data mining for such task is based on statistical methods for data analysis. During analysis, the predicted values will be obtained with some accuracy to be assessed also by computer intensive methods like cross-validation or bootstrap [16]. It is possible that some scenarios will provide unacceptable results. This situation generates corrective measures and re-running the DFA. The control and optimisation module of the DFA system is responsible for such a task. After the optimisation step, the presentation module will provide reports obtained from the DFA. The management board will choose from these reports the best strategy for the company.

Even "DFA models provide generally deeper insight into risk and potential rewards of business strategies than scenario testing can do", some weaknesses of the DFA there exist [14]. The first difficulty consists of capturing the complexity of the real-life business environment. The second one deals with the strong dependence of the results on the assumptions used in the model set-up. For small models the results can correspond with intuition. However, more effort is necessary to understand and control the uncertainties and approximations for a useful DFA approach. The next section discusses the application of bootstrapping for loss models and for controlling uncertainties in the case of DFA applied in insurance.

5. **Applications and concluding remarks**

Analysing real annually aggregated number of severity losses for some financial entity, the accumulation is similar to a continuous cumulative distribution function that supports the usage of some kind of Poisson process. Even such a counting process, in general, is not a simple Poisson but a non-homogeneous Poisson process as proved in [19], we use the Poisson distribution to simulate losses as described during LDA presentation. DFA is used mainly in the insurance domain (non-life and life insurances). However, other fields, like banking and travel-industry benefit from DFA advantages.

For any insurance enterprise there are three basic elements of risk to be considered in a DFA model: the Liability Risk (LR), called obligation risk also, is the risk that the cost of settling the insurance liabilities will be greater than expected, the Asset Risk (AR) is the risk that the realizable value of assets will be less than anticipated, and the Business Risk (BR) which reflects the general business risks faced by all enterprises. A complicated factor for an international insurer (due to the globalisation effect) is related to currency-exchange rates fluctuation. Both LR and AR have a currency dimension. However the currency risk can be treated as an embedded element of LR&AR and not as a separate element of risk. Moreover, any DFA model has to provide a portfolio analysis and any update is better to be done on-line. We have identified the following occasions to use the bootstrap approach for improving the degree of certainty when apply DFA: 1) solvency-testing, 2) the asset allocation and reinsurance programs' evaluation, and 3) capital allocation determination.

There are two important approaches in solvency testing: cash flows analysis based on interest rates/inflation rates, and the analysis of market values of assets and liabilities. The last approach considers the changes in the solvency margin by projecting the cash inflows (asset values, earned premiums etc.) and outflows (dividend payments, taxes, claims etc.). A non-parametric cash-flow analysis is based on the following assumptions about the time series $J_t = \ln(I_{t+1}/I_t)$, where I_t¬ is the interest rate prevailing at time: 1) J_t is a (strictly) stationary time series, and 2) J_t has the m-dependence structure for some fixed integer $m>0$. The bootstrap, in a modified variant, can be used to estimate, for example, the distribution of the Surplus-Value process, depending on the time-series values.

When dealing with asset allocation, it is needed a multi-line business, multi-period, multi-currency and multi-asset stochastic plan generator. The input can use not only asset details or insurance business details, but also the catastrophic loss details. According to the Modern Portfolio Theory, the standard deviation which measures the extent and frequency of the variance in a portfolio's return (or how the unpredictable this return will be), is a better approach than the classical one based on the volatility related to market. The accuracy of standard deviation can also be study by bootstrap as shown above.

This short overview shows the applicability of bootstrap techniques during DFA. We do not claim that these are the only cases. We expect to identify more situations during the requirements identification for risk management software which will be developed for training during master courses in data analysis for finance, banking and insurance science.

References

1. Albeanu G., Ghica M., *Bootstrap Simulation Models for Probabilistic Approaches in Environmental Economics*, The 6th Workshop on Mathematical Modelling of Environmental and Life Sciences Problems, September 5-9, 2007, Constanţa, Romania.
2. Albeanu G., Popenţiu F., *On the Bootstrap Method: Software Reliability Assessment and Simultaneous Confidence Bands*, Annals of Oradea University, Energetic Series, 7(1), 109-113, 2001.
3. Albeanu G., *Resampling Simultaneous Confidence Bands for Nonlinear Explicit Regression Models*, Mathematical Reports, 50(5-6), 289-295, 1998.
4. Bergbauer A., Charez V., Fischer T., Perera R., Roehrl A., Schmiedl S., *Back to the Future: Dynamic Financial Analysis (DFA) for Decision Making*, 2004, http://www.approximity.com/papers/dfa_wp4en.pdf.
5. Blum P., Dacorogna M., *DFA-Dynamic Financial Analysis*, in "Encyclopaedia of Actuarial Science", John Wiley & Sons, 2004.
6. Derrig R.A., Ostaszewski K.M., Rempala G.A., *Applications of Resampling Methods in Actuarial Practice*, Proceedings of Casualty Actuarial Society, 87, 222-264, 2000.
7. Efron B., Tibshirani R.J., *An Introduction to the Bootstrap*, New York: Chapman and Hall, 1993.
8. Efron B., *Computer-Intensive Methods in Statistical Regression*, SIAM Review, 30, 3, 421-449, 1988.
9. Fishman G., *Monte Carlo: Concepts, Algorithms, and Applications*, Berlin: Springer, 1996.
10. Ghica M., *A Risk Exchange Model With a Mixture Exponential Utility Function*, Annals of Bucharest University, Mathematics-Informatics Series, 2006.
11. Ghica M., *Optimal Portfolios in Finance and Actuarial Science*, Annals of *Spiru Haret* University, Mathematics and Informatics Series, 2006.
12. Good P.I., *Resampling Methods*, 2nd ed., Boston: Birkhauser, 2001.
13. Hogg R.V., Klugman S.A., *Loss Distributions*, New York: John Wiley & Sons, 1984.
14. Kaufmann R., Gadmer A., Kett R., *Introduction to Dynamic Functional Analysis*, ASTIN Bulletin 31(1), 213-249, 2001.
15. Klugman S.A., Panjer H.H., Willmot G.E., *Loss Models: From Data to Decisions*, New York: John Wiley & Sons, Inc., 1998.
16. Kohavi R., *A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection*, International Joint Conference on Artificial Intelligence (IJCAI), 1995.
17. Ostaszewski K.M., Rempala G.A., *Parametric and Nonparametric Bootstrap in Actuarial Practice*, The Actuarial Foundation, 2000.
18. Pallini A., *Efficient Bootstrap Estimation of Distribution Functions*, METRON, 58(1-2), 81-95, 2000.
19. Rachev S.T., Chernbai A., Menn C., *Empirical Examination of Operational Loss Distributions*, Karlsruhe University Technical Report, 2005.
20. Tukey J.W., *Exploratory Data Analysis*, Reading, MA: Addison-Wesley, 1977.
21. The Phrase Finder: 2007, *Meanings and Origins of Phrases, Sayings and Idioms*, http://www.phrases.org.uk/meanings/290800.html