European Journal of Operational Research 256 (2017) 163–177

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

Relevant states and memory in Markov chain bootstrapping and simulation

Roy Cerqueti[a], Paolo Falbo[b], Cristian Pelizzari[a,*]

[a] Dipartimento di Economia e Diritto, Università degli Studi di Macerata, Via Crescimbeni, 20 - 62100 Macerata MC, Italy
[b] Dipartimento di Economia e Management, Università degli Studi di Brescia, Contrada S. Chiara, 50 - 25122 Brescia BS, Italy

ARTICLE INFO

Article history:
Received 31 March 2015
Accepted 4 June 2016
Available online 15 June 2016

Keywords:
Bootstrapping
Information theory
Markov chains
Optimization
Simulation

ABSTRACT

Markov chain theory is proving to be a powerful approach to bootstrap and simulate highly nonlinear time series. In this work, we provide a method to estimate the memory of a Markov chain (i.e. its order) and to identify its relevant states. In particular, the choice of memory lags and the aggregation of irrelevant states are obtained by looking for regularities in the transition probabilities. Our approach is based on an optimization model. More specifically, we consider two competing objectives that a researcher will in general pursue when dealing with bootstrapping and simulation: preserving the "structural" similarity between the original and the resampled series, and assuring a controlled diversification of the latter. A discussion based on information theory is developed to define the desirable properties for such optimal criteria. Two numerical tests are developed to verify the effectiveness of the proposed method.

1. Introduction

Bootstrapping and simulation procedures have been applied intensively to solve a wide variety of problems. Following such a widespread interest, several methodological contributions have appeared to improve the initial bootstrap method advanced by Efron (1979), even if the basic idea remains unchanged (e.g., see the methodological discussion on the *classical bootstrap* methods in Freedman, 1984; Freedman & Peters, 1984; Efron & Tibshirani, 1986; 1993). In particular, the heart of the bootstrap consists of resampling some given observations to the purpose of obtaining a good estimation of statistical properties of the original population.

An important restriction to the classical bootstrap methods is the hypothesis that the observations in the sample are realizations of independent and identically distributed random variables. However, in the case of time series taken from the real life, this condition is hardly true. When such hypothesis is not true, a theoretical model for the data is required and the bootstrap is then applied to the model errors.

A new group of bootstrapping methods have been advanced to reduce the risk of misspecifying the model. To this group belong the so-called *block*, *sieve*, and *local* methods of bootstrapping (see Bühlmann, 2002, for a comparison of these methods). The methods are nonparametric, and assume that observations can be (time) dependent.

This category of literature has increased in a relatively recent period, and new methods of bootstrapping based on Markov chain theory have appeared. The major advantage of this approach is that it is entirely data driven, so that it can smoothly capture the dependence structure of a time series, releasing a researcher from the risk of wrongly specifying the model, and from the difficulties of estimating its parameters.

The limitation connected to Markov chains is, of course, that they are naturally unsuitable to model continuous-valued processes. This is an unfortunate situation, since several phenomena in many areas of research are often modeled through continuous-valued processes. In economic and financial literature, there are plenty of cases of continuous-valued processes showing complex behaviors, where data show non-linear dependence. It is well known that in the financial markets, next to technological and organizational factors, psychology and emotional contagion introduce complex dynamics in driving the expectations on prices (e.g., think of the terms popular in the technical analysis such as "psychological thresholds," "price supports," "price resistances," etc.). In such cases, the selection of the correct model for complex continuous-valued stochastic processes is highly subject to uncertainty.

To overcome model risk, a researcher in the need of bootstrapping or simulating a continuous-valued stochastic process could in principle resort to partitioning its support, obtaining a discretized version of it, and then apply Markov chain bootstrapping or simulation techniques to model brilliantly any arbitrary dependence

* Corresponding author. Tel.: +390302988516.
E-mail addresses: roy.cerqueti@unimc.it (R. Cerqueti), paolo.falbo@unibs.it (P. Falbo), cristian.pelizzari@unibs.it (C. Pelizzari).

http://dx.doi.org/10.1016/j.ejor.2016.06.006
0377-2217/

structure. Such a solution has, however, a major difficulty, which is how to organize an *informational efficient partition* of the process support. Indeed, in the absence of some guide, arbitrarily fixing a partition inappropriately refined involves two major drawbacks of bootstrapping and simulation: insufficient diversification of the resampled trajectories and unsatisfactory replication of the key features of the stochastic process.

Focusing on the relevant states is crucial if we want to consider the discrete versions of complex continuous-valued processes. As mentioned previously, it is frequent in economic and financial markets that some observed states, or combinations of them, are more relevant than others in determining the future evolution of the process. In other words, not all the partitions of the support of a continuous-valued process are suitable to capture the relevant information about its dependence structure. Therefore, finding the optimal ones is crucial to capture and replicate satisfactorily the key features of the original process. To this purpose, the approach proposed in this paper first divides the support into a relatively high number of small intervals (the so-called *initial partition*). These intervals are naturally taken as (preliminary) states of the approximating Markov chain. Afterwards, these states are re-grouped following an optimal clustering procedure. Since grouping two states (to form a new one) implies an increase of disorder (or an information loss) in the sense of Kolmogorov (1965), the method proposed here seeks to minimize distance indicators respectful of the properties of disorder measures. The solution identifies partitions grouping the states with the most similar transition probabilities. In this way, the resulting groups emerge as the relevant states: the states that influence the conditional distribution of the process differently one from the others.

However, any unconstrained clustering method searching to minimize an information loss measure would end up with a partition letting all the original states separate. Such a solution is highly undesirable for bootstrapping and simulation, since it would provide insufficient diversification of the resampled series.

Overall, the method advanced in this paper is therefore a constrained optimization problem, where the search of minimal information loss is controlled by a "multiplicity" measure to guarantee a satisfactory diversification of the resampled series.

The objectives of the minimization problem are distance indicators, which are based on transition probabilities. In similar problems, some authors have focused on entropy measures. We justify our choice with two observations:

- Firstly, the key proposal of this work is a methodological approach. Therefore, choosing distance indicators in some "appropriate" way is a second level objective of the paper with respect to the proposal of the methodological approach. Indeed, we do not aim here at comparing the performances of distance indicators *per se*, but rather at providing an original satisfactory proposal specifically tailored to the problem at hand. In principle, many other distance indicators can be applied instead of the two advanced in the paper (see, e.g., Ullah, 1996; Bernardo & Rueda, 2002; Cha, 2007);
- Secondly, the choice of the two distance indicators is, however, not casual, since they have been required to respect the properties of disorder measures, in the sense of Kolmogorov (1965) (see Section 4.4). Such properties turn out to be necessary to guarantee coherent minimum and maximum values to the "singleton" and the "all comprehensive" partitions, respectively, i.e. the two extreme solutions expected in this problem (see, in particular, Section 4.1).

The optimization advanced here addresses at the same time (as it will be discussed in Section 5) the problem of determining the memory (i.e. the order) of a Markov chain.

Our work contributes to the literature on Markov chain bootstrapping and simulation in various ways.

Firstly, we develop a method to estimate jointly the parameters (states and order) of a Markov chain dedicated to bootstrap and simulation via constrained optimization. When the threshold defining the multiplicity constraint is changed according to a grid of values, an efficient frontier obtains, whose properties provide a complete description of the optimal solutions.

Secondly, we propose a non-hierarchical approach, which means that a non-sequential search of the order of the Markov chain is performed. More precisely, if some states are grouped at a given time lag w, then they are not forced to stay together at farther time lags $w + r$ (with $r > 0$). This "freedom" adds flexibility in modeling the dependence structure of a Markov chain and, to our knowledge, our approach is the first in the literature on Markov chain bootstrapping and simulation to abandon hierarchical grouping. Such feature is not of secondary importance, since it allows us to model a Markov chain with non-monotonically decreasing memory.

Thirdly, compared to the bootstrap literature developed in econometrics and applied statistics, our proposal treats states as if they were of qualitative nature, and the search of efficient partitions is based only on transition probabilities. In other words, no distance between the values of the different states is used in the decision of merging them. Again, this approach allows us a higher flexibility in the identification of the relevant states and an increased capacity to capture the dynamics of a Markov chain.

Fourthly, this paper provides the theoretical grounds for Markov chain bootstrapping and simulation of continuous-valued processes. Our search for the relevant states supplies the levels where the process modifies significantly its dynamics (i.e. its expected value, its variance, etc.). Hence, it is designed to minimize the information loss deriving from aggregating the states, so it helps to maintain highly complex nonlinearities of the original process.

Fifthly, we introduce two new non-entropic measures of the disorder of a Markov chain process, and we study their main properties.

Sixthly, given the theoretical nature of the present paper, we extend and complement Cerqueti, Falbo, Guastaroba, and Pelizzari (2013), where sub-optimal solutions are derived through a tabu search procedure.

The paper is organized as follows. Section 2 reviews the relevant literature on Markov chain bootstrapping. Section 3 introduces the settings of the problem. Section 4 discusses some theoretical properties of the criteria used here to select the optimal dimension of a Markov chain transition probability matrix. Section 5 discusses some methodological issues. In Section 6, the criteria are applied to two examples. Section 7 concludes.

2. A bibliography review on Markov chain bootstrapping

It is possible to group different contributions on resampling procedures based on Markov chain theory.

A first major category is concerned with processes that are not necessarily Markov chains. A series of stationary data is divided into blocks of length l of consecutive observations; bootstrap samples are then generated, randomly joining some blocks. The seminal idea appears first in Hall (1985) for spatial data, has been applied to time series by Carlstein (1986), but has been fully developed starting with Künsch (1989) and Liu and Singh (1992). In Hall, Horowitz, and Jing (1995), Bühlmann and Künsch (1999), Politis and White (2004), and Lahiri, Furukawa, and Lee (2007), the selection of the parameter l (a crucial point of this method)