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