Conditional distribution resampling for time series

Svetlana Borovkova *(Delft University of Technology)*
Cees Diks *(CenDEF, University of Amsterdam)*

Abstract

Nonparametric methods of time series modeling and forecasting have attracted much attention recently, mainly due to availability of fast computers and advanced computation algorithms. Among those methods are local linear models, nearest neighbors methods and kernel autoregression. On the other hand, there also has been much development in the area of resampling methods for time series: moving block bootstrap of Kunsch (1989), stationary bootstrap of Politis and Romano (1994) and so on. These two areas of research pursue different goals: modeling and forecasting in the first case and generating copies of the observed series in the second. The bootstrap-like methods are concerned with preserving the distribution of the data, but they usually cannot be interpreted as nonparametric models. Nonparametric time series models, on the other hand, are primarily concerned with capturing the dependence structure of the series within an autoregression function, while not being particularly bothered with preserving marginal distributions. However, for many economic time series, these marginal distributions can be very distinct (e.g. multimodal as in case of energy prices) and carry a lot of information.

In this paper we introduce a resampling method for time series that combines the virtue of both approaches: it results into a stationary process, preserves marginal and joint distributions (up to some order) and it is interpretable as a nonparametric model. Moreover, this method is entirely data-driven and computationally fast. Applications of this method include generating copies of the original time series with the same empirical marginals and Markovian structure (up to some order), forecasting by means of providing prediction intervals and generating bootstrap confidence intervals.

The method, which we call conditional distribution resampling, is the combination of the kernel estimation of the conditional density (of the next observation in the time series given the past) with the nearest neighbor method. It is essentially a two-step procedure, whereby on the first step we select a "neighbor" of the current observation according to some kernel weights and on the second step we draw a "successor" of the current observation from a kernel density centered at the "successor of the neighbor", which is essentially the conditional density. These two steps are repeated until a series of desired length is obtained. In fact, this procedure generates a realization of a Markov process defined on the original observations. We prove that the intermediate sequence of "neighbors" exactly preserves the empirical marginal distribution of the data.

This method extends naturally to the autoregression of k-th order, by considering the series of reconstruction vectors. In this case, we suggest to use weights that depend on a distance between reconstruction vectors, defined in the way that reflects decreasing dependence of the future observation on the

past ones.

The bandwidth selection for kernel methods is usually done by means of cross-validation (i.e. "leave-one-out" technique) combined with minimization of e.g. mean-squared prediction error. Since here our "observations" are the values of the conditional density function, we employ a variant of maximum likelihood method, whereby we combine cross-validation technique with maximizing the likelihood function. We study the asymptotic properties of the bandwidth chosen in this way and indicate the optimal bandwidth rate in terms of the sample size.

Applying the method to generated as well as real-life economic time series of daily oil prices shows that the dependence structure as well as the marginal distributions of the data are captured remarkably well. We obtain oil price forecasts in terms of prediction intervals and bootstrap confidence bounds for certain functionals of the observed series.