Journal of Multivariate Analysis 134 (2015) 129–145

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Tie the straps: Uniform bootstrap confidence bands for semiparametric additive models☆,☆☆

Wolfgang Karl Härdle [a,b], Ya'acov Ritov [c], Weining Wang [a,*]

[a] C.A.S.E. - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany

[b] Singapore Management University, 50 Stamford Road, Singapore 178899, Singapore

[c] Department of Statistics Hebrew University of Jerusalem Mount Scopus, Jerusalem 91905, Israel

ARTICLE INFO

Article history:
Received 9 November 2013
Available online 20 November 2014

AMS subject classifications:
62G08
62G09
62G20
62H12
62P20

Keywords:
Nonparametric regression
Bootstrap
Quantile regression
Confidence bands
Additive model
Robust statistics

ABSTRACT

This study considers the theoretical bootstrap "coupling" techniques for nonparametric robust smoothers and quantile regression, and we verify the bootstrap improvement. To handle the curse of dimensionality, a variant of "coupling" bootstrap techniques is developed for additive models both in a robust mean regression and in a quantile regression framework. Our bootstrap method can be used in many situations such as constructing confidence intervals and bands. We demonstrate the bootstrap improvement over the theoretical asymptotic band in simulations and in applications to firm expenditures and the interaction of economic sectors and the stock market.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Confidence bands are important tools for model specifications. However, it is difficult to construct precise confidence bands for nonparametric curves since a supreme norm is usually involved in the statistics. Traditional methods based on the asymptotic theory have natural drawbacks in their finite sample performance, and this motivates bootstrap methods to attain more precise bands. In this article, we deal with bootstrap bands construction for a general class of nonparametric M- and L-estimates; moreover, we adopt additive models to handle the multivariate covariates case. We believe that the developed technique is essential for componentwise shape inspection of additive models in empirical economics. Applications can be found in the work of [6] or [4].

Consider $Y, X \in \mathbb{R}^{d+1}$ with variable Y and $X \in \mathbb{R}^d$.

$$l(x) = \arg\min_{\theta} \mathsf{E}_{Y|X=x}\rho(Y-\theta), \tag{1}$$

☆ We thank the Editors and two referees for helpful comments. The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 "Ökonomisches Risiko", Humboldt-Universität zu Berlin is gratefully acknowledged.

☆☆ We appreciate the friendly working environment and atmosphere of the MFO Mathematical Research Institute of Oberwolfach during our Research in pairs stay.

* Corresponding author.
E-mail addresses: haerdle@wiwi.hu-berlin.de (W.K. Härdle), yaacov.ritov@gmail.com (Y. Ritov), wangwein@cms.hu-berlin.de (W. Wang).

http://dx.doi.org/10.1016/j.jmva.2014.11.003
0047-259X/© 2014 Elsevier Inc. All rights reserved.

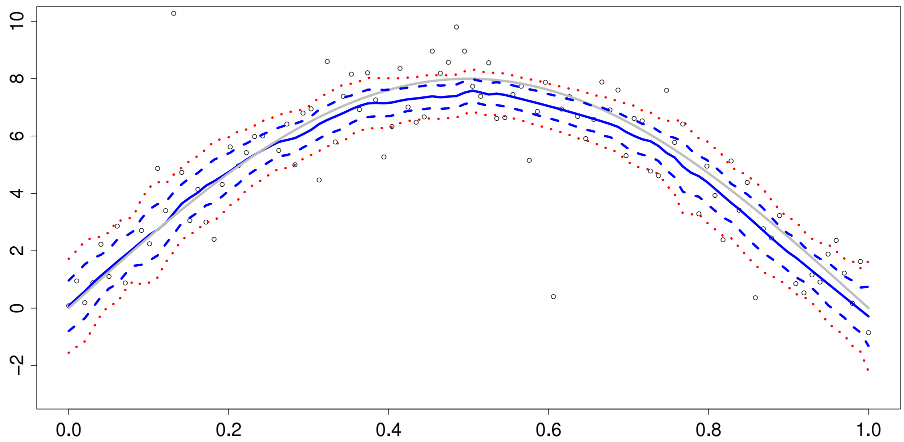

Fig. 1. Plot of true curve (gray), robust estimate with Tukey biweight loss and 5% confidence bands (blue dashed), local polynomial estimate (black), bootstrap band (red dotted), $n = 150$.

with $\rho(\cdot)$ as a loss function described in detail in Section 2.1. For the confidence bands construction, one stream of literature using empirical process theory follows the asymptotic results of [2], which is further extended by [19], and has recently been studied by [13] for L-smoothers. However, it has also been shown by [31] that such an asymptotic confidence band has a much lower coverage probability in a finite sample than what it is supposed to have. The poor performance of such a band in a finite sample has been well attributed to its slow convergence, see [7]. To address improvement of a finite sample performance, there is a class of literature on the bootstrap confidence band, see [3,12], among others. Fig. 1 shows a bootstrap confidence band and an asymptotic band for an M-smoother with $\rho(\cdot)$ being Tukey's bisquare loss. One can see that the asymptotic band is narrower than the bootstrap one. Moreover, the asymptotic band does not cover the true curve while the bootstrap band does.

Many different resampling techniques have been developed, among which the permutation tests are very popular and can be adapted to many application scenarios, see [28,1,22,23]. The bootstrap is a class of data driven resampling techniques that provide non-asymptotic approximations of finite sample distributions of different statistics. In a location model (more generally a regression model), resampling is done from the estimated residuals, and a typical theoretical analysis leads to the conclusion that "bootstrap works" in the sense that a suitably centered bootstrap estimator converges to the same asymptotic normal distribution as the original estimator under consideration. A large literature body has focused on showing bootstrap improvements and refinements of approximations via bootstrap resampling, see [8,21,15,11], which discuss the conditions for bootstrap consistency, and also prove the bootstrap accuracy as an approximation to the exact finite sample distribution for special types of statistics in a nonparametric framework. However, very few articles have focused on nonlinear statistics (e.g. maximum) in nonparametric regression. [31] proposes a bootstrap procedure and shows its improvement properties.

This stimulates the current research on finding common properties that loss functions have to share in order to attain such an improvement. Accordingly, we prove a generalized version in the univariate case for a class of loss function with bounded influence. The bootstrap becomes difficult when the dimension d of the regressors gets large. One way to avoid this problem is to impose a structure, such as an additive model, on the multivariate nonparametric function. The additive structure assumes that the covariates' effects are separable, and this effect is presented in many economic applications, [10]. Specifically, we consider the regression function

$$m(x_1, \ldots, x_d) = \sum_{j=0}^{d} m_j(x_j), \tag{2}$$

with $m_0(x_0)$ a constant. It is worth noting that the additive structure implicitly assumes that the covariates effects are separable, and of course this assumption needs to be tested in advance. [32] develops a test on testing the interaction effects; correspondingly, our method can also be extended to implement similar tests.

It is well known that (2) avoids the "curse the dimensionality" in the sense that one-dimensional convergence rates are achieved for the estimation of $m(x_1, \ldots, x_d)$, but keeps enough flexibility of the marginal influence of the different variables. See [14,17,16] among many others. [14] focuses on generalized additive models with unknown link functions, [17] proposes a two-stage estimation for quantile regression in additive models, [16] shows the equivalence between spline, kernel and other methods in terms of optimal minimax rate in the additive model estimation. The resulting estimate $\hat{m}_j(x_j)$ in (2) though needs to be screened for closeness to $m_j(x_j)$. This requires construction of confidence intervals and bands as a function of x_j. For such screening tests, our tightened bootstrap techniques will be verified. Namely, the bootstrap-based confidence bands are shown to be very close to the true finite sample distribution-based ones.

Download English Version:

https://daneshyari.com/en/article/1145595

Download Persian Version:

https://daneshyari.com/article/1145595

Daneshyari.com