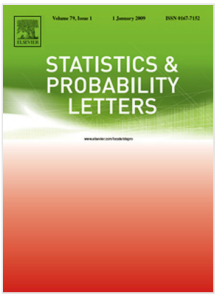
Accepted Manuscript



Imputation in nonparametric quantile regression with complex data

Yanan Hu, Yaqi Yang, Chunyu Wang, Maozai Tian

PII: S0167-7152(16)30216-4
DOI: http://dx.doi.org/10.1016/j.spl.2017.03.003
Reference: STAPRO 7881

To appear in: *Statistics and Probability Letters*

Received date: 14 October 2016
Revised date: 4 March 2017
Accepted date: 4 March 2017

Please cite this article as: Hu, Y., Yang, Y., Wang, C., Tian, M., Imputation in nonparametric quantile regression with complex data. *Statistics and Probability Letters* (2017), http://dx.doi.org/10.1016/j.spl.2017.03.003

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Imputation in Nonparametric Quantile Regression with Complex Data

Yanan Hu[a], Yaqi Yang[a], Chunyu Wang[a], Maozai Tian[a,b,c]

[a] *Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, 100872, China*
[b] *School of Statistics and Information, Xinjiang University of Finance and Economics, 830012, Xinjiang, China*
[c] *School of Statistics, Lanzhou University of Finance and Economics, Lanzhou, 730101, Gansu*

Abstract

This paper considers nonparametric quantile regression models for complex data of mixed categorical and continuous variables together with missing values at random (MAR). In consideration of the increasingly popular application of multiple imputation for handling missing data and the advantages of nonparametric quantile regression, we propose an effective and accurate multiple imputation method. The estimation procedure not only does well in modeling with mixed categorical and continuous data, but also makes full use of the entire dataset to achieve increased efficiency. The proposed estimator is asymptotically normal. In simulation study, we compare the performance of the multiple imputation method with complete case (CC), Regression imputation and nearest-neighbor imputation methods, and outline advantages and drawbacks of the different methods. Simulation studies show that the proposed multiple imputation method performs better.

Keywords: Complex data, Missing covariates, Multiple imputation, Quantile regression

1. Introduction

Variables used as covariates often have missing values (Yang & Kim, 2016). A simple approach, called a complete-case analysis, uses only observations without missing values. The complete case method can lead to biased estimators and is inefficient relative to procedures that use observations with partial information.

Imputation is a widely used method to exploit the full information contained in the data set. In imputation, the missing values are replaced with plausible values derived under a set of assumptions. Well-known imputation methods include mean imputation and nearest-neighbor imputation.

Two types of imputation procedures are single imputation and repeated imputation. In single imputation, each missing value is replaced with a single imputed value. Examples of single imputation methods include Cheng (1994) and Wang et al. (2004). In repeated imputation, each missing value is replaced with multiple imputed values. Two broad categories of repeated imputation procedures are multiple imputation and fractional imputation Kim & Shao (2013).

Repeated imputation overcomes limitations associated with single imputation (Landerman et al., 1997). The use of more than one imputed value can lead to more precise estimators than those based on a single imputed value. The use of multiple imputed values also allows estimation of the variance associated with the imputation procedure. We focus on a frequentist version of the type of multiple imputation discussed in Little & Rubin (1987). Our multiple imputation paradigm is closely related to that of Wei et al. (2012)). Multiple imputation creates $L > 1$ complete data sets. Each complete data set is analyzed by standard complete-data methods, and the results are then combined.

Quantile regression, as first introduced by (Koenker & Bassett, 1978) , is gradually developing into an important modelling tool, due to its flexibility in exploring the relationship between the response and the covariates. Compared with mean regression, quantile regression is more explicable and robust, when the

URL: The corresponding author: Prof. Maozai Tian, mztian@ruc.edu.cn (Maozai Tian)

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات