

Prediction of conditional quantiles on the half line and the unit interval using transformation models

Predizione di quantili condizionati sulla semiretta e l'intervallo unitario tramite trasformazioni

Marco Geraci and Chris Jones

Abstract In statistical applications, transforming data may serve a number of purposes. In parametric regression, transformations are often applied to the response variable in hopes of meeting one or more assumptions of a simplified model that are otherwise unsupported by the raw (untransformed) data. A number of transformation families have been developed to address the violation of the standard assumptions of linear mean regression, and recently some of these have been extended to quantile regression to model nonlinear predictors. In this talk, we will discuss recent advances in the application of transformations to predict conditional quantiles of continuous variables with support $\mathbb{R}^+$ or the unit interval.

Abstract Nelle applicazioni statistiche, la trasformazione dei dati può assolvere diverse finalità. Nell'ambito della regressione parametrica, le trasformazioni sono spesso applicate alle variabili di risposta nella speranza di soddisfare una o più assunzioni di un modello semplificato che non sono altrimenti supportate dai dati grezzi. Numerose famiglie di trasformazioni sono state sviluppate per affrontare la violazione delle assunzioni standard di un modello di regressione lineare per la media, e alcune di queste sono state recentemente estese alla regressione su quantili per modellare predittori nonlineari. In questa presentazione, discuteremo recenti sviluppi nell'applicazione di trasformazioni per la predizione di quantili condizionati di variabili continue con supporto $\mathbb{R}^+$ o l'intervallo unitario.

Key words: bounded variable, ceiling and floor effects, nonlinear quantile functions.

Marco Geraci

Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, University College London, United Kingdom, e-mail: m.geraci@ucl.ac.uk.

Chris Jones

Department of Mathematics and Statistics, The Open University, United Kingdom, e-mail: chris.jones@open.ac.uk.

1 Background

Inference on conditional quantiles of an outcome variable which is defined on the half line or the unit interval requires appropriate modelling approaches to constrain prediction within the limits of the support. The Box-Cox [1, 5] and, more recently, the Aranda-Ordaz [2] transformation families, originally developed for modelling the mean of, respectively, continuous and binary responses, have been proposed as a parametric approach to the estimation of nonlinear quantile functions of, respectively, positive and bounded continuous outcomes. These models offer a flexible alternative to nonparametric estimation and perform satisfactorily under different scenarios. However, Box-Cox and Aranda-Ordaz transformations suffer from two drawbacks. Back-transformation may be hindered as they do not, in general, have range the entire real line. As a consequence, observations for which the objective function is not well-defined need to be excluded and information is lost. Moreover, these transformations may alter the behaviour of the quantile function at its extremities. To overcome such limitations, alternative transformations [4] are proposed. A simulation study is carried out to assess performance of different models in terms of prediction error. Different estimators of the quantile functions are considered, including two-stage [3], residual cusum process [5] and non-smooth estimators. Some applications, including assessment of educational attainment, are illustrated.

References

1. Buchinsky, M. (1995). Quantile regression, Box-Cox transformation model, and the US wage structure, 1963-1987. *J. Econometrics* 65(1), 109–154.
2. Dehbi, H.-M., M. Cortina-Borja, and M. Geraci (2014). Aranda-Ordaz quantile regression for student performance assessment. Unpublished manuscript.
3. Fitzenberger, B., R. Wilke, and X. Zhang (2010). Implementing Box-Cox quantile regression. *Econometric Rev.* 29(2), 158–181.
4. Jones, M. C. (2007). Connecting distributions with power tails on the real line, the half line and the interval. *Int. Stat. Rev.* 75(1), 58–69.
5. Mu, Y. M. and X. M. He (2007). Power transformation toward a linear regression quantile. *J. Am. Statist. Assoc.* 102(477), 269–279.