Nonparametric Estimation and Inference on Conditional Quantile Processes

Zhongjun Qu       Jungmo Yoon
Boston University Claremont Mckenna College

January 30, 2011

Abstract

We consider the estimation and inference about a nonparametrically specified conditional quantile process. For estimation, a two-step procedure is proposed. The first step utilizes local linear regressions and maintains quantile monotonicity through simple inequality constraints. The second step involves linearly interpolating between adjacent quantiles. When computing the estimator, the bandwidth parameter is allowed to vary across quantiles to adapt to the data sparsity. The procedure is computationally simple to implement and is practical even for relatively large data sets. For inference, we first obtain a uniform Bahadur representation for the conditional quantile process. Then, we prove the estimator converges weakly to a continuous Gaussian process, whose critical values can be estimated via simulations. These results are useful for constructing uniform confidence bands and testing hypothesis about the quantile process. For application, we revisit a data set considered in Chetty (2008) on durations of unemployment and report some new findings.

Keywords: nonparametric quantile regression, monotonicity constraint, uniform Bahadur representation, uniform inference.
1 Introduction

Models for conditional quantiles play an important role in econometrics and statistics. They complement the conventional conditional mean models, offering a broader view of the diversity surrounding the mean effects. Parametric approach to conditional quantile modelling is pioneered by Koenker and Bassett (1978), while the nonparametric approach is developed in a series of papers including Stone (1977), Chaudhuri (1991), Fan, Hu and Truong (1994), Koenker, Portnoy and Ng (1994) and Koenker (2010). Comprehensive reviews of the related literature can be found in Koenker (2005).

Individual quantile functions, particularly the conditional median, are often of interest, but more often one wishes to study a full family of quantiles to obtain a more complete analysis of the stochastic relationships between variables. This motivation underlies the study of conditional quantile processes. A seminal contribution is Koenker and Portnoy (1987), which established a uniform Bahadur representation and serves as the foundation for further developments in this area. More recently, Koenker and Xiao (2002) considered the issue of testing composite hypotheses about quantile regression processes using Khmaladzation (Khmaladze, 1981). Chernozhukov and Fernández-Val (2005) considered the same issue and suggested resampling as an alternative approach. Their results can be used to study a wide range of issues, including but not restricted to (i) testing alternative model specifications, (ii) testing stochastic dominance, and (iii) detecting treatment effect significance and heterogeneity.

A main focus of the above literature has been on quantile processes in parametric models. Such models play a critical role, however, there are frequent occasions when parametric specifications fail, and more flexible nonparametric methods become desirable. Relaxing parametric assumptions is also often practical in view of the increasing availability of rich data sets in both statistics and econometrics. Such considerations motivated us to consider the estimation and inference about conditional quantile processes in a nonparametric setting. Our first goal is to propose a simple-to-implement estimator and the second is to provide a tractable inferential theory useful for constructing uniform confidence bands as well as conducting hypothesis tests.

We propose a two-step procedure to estimate the conditional quantile process. The first step applies local linear regressions (see Fan, Hu and Truong, 1994 and Yu and Jones,1998) to a grid of quantiles and maintains quantile monotonicity through a set of simple inequality constraints. The constraints depend neither on the data or the model (such as the number of the covariates) but only on the number of quantiles entering the estimation. The second step involves linearly interpolating between adjacent quantiles, returning an estimate for the quantile process. This procedure has two salient features. First, the bandwidth parameter is allowed to vary across quantiles to adapt
to the data sparsity. This is important because data are typically sparse near the tails of the conditional distribution. Second, the optimization can be written as a linear programming problem with linear inequality constraints and can be solved using the algorithm in Koenker and Ng (2005) without essential modification. Our experimentation suggests that the procedure is practical even for relatively large sample sizes.

For inference, our contribution is threefold. First, we obtain a uniform Bahadur representation, generalizing Theorem 2.1 in Koenker and Portnoy (1987) to a nonparametric framework. Second, we prove that, under a certain rate condition on the quantile grid, the proposed estimator has the same first order limiting distribution as if the inequality constraints were absent. The distribution is a continuous Gaussian process, whose critical values can be estimated using simulations. We discuss in detail how it can be used to compute optimal bandwidth, construct uniform confidence bands and conduct hypothesis tests. Finally, we establish a connection with the results in Chernozhukov, Fernández-Val and Galichon (2010). Specifically, we show that an alternative estimator, obtained from the same two-step procedure but without inequality constraints, satisfies a functional central limit theorem, therefore can be used as the basis for rearrangement. This formally justifies the applicability of rearrangement in a nonparametric context.

The paper is organized as follows. Section 2 introduces the issue of interest. Section 3 presents the estimator. Section 4 discusses the computational aspect. Section 5 presents the asymptotic results and illustrates how to construct uniform confidence bands and conduct hypothesis tests. Section 6 discusses bandwidth selection. Section 7 establishes the connection to re-arrangement. Section 8 includes some Monte Carlo experiments. Section 9 considers an empirical applications and Section 10 concludes. All proofs are included in the Appendix.