Journal of Econometrics 185 (2015) 1–19

Contents lists available at ScienceDirect

Journal of Econometrics

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

Nonparametric estimation and inference on conditional quantile processes[☆]

Zhongjun Qu[a,*], Jungmo Yoon[b]

[a] Department of Economics, Boston University, United States

[b] Robert Day School of Economics and Finance, Claremont McKenna College, United States

A R T I C L E I N F O

Article history:
Received 5 September 2012
Received in revised form
11 July 2014
Accepted 20 October 2014
Available online 4 November 2014

JEL classification:
C14
C21

Keywords:
Nonparametric quantile regression
Uniform Bahadur representation
Uniform inference
Treatment effect

A B S T R A C T

This paper presents estimation methods and asymptotic theory for the analysis of a nonparametrically specified conditional quantile process. Two estimators based on local linear regressions are proposed. The first estimator applies simple inequality constraints while the second uses rearrangement to maintain quantile monotonicity. The bandwidth parameter is allowed to vary across quantiles to adapt to data sparsity. For inference, the paper first establishes a uniform Bahadur representation and then shows that the two estimators converge weakly to the same limiting Gaussian process. As an empirical illustration, the paper considers a dataset from Project STAR and delivers two new findings.

1. Introduction

Models for conditional quantiles play an important role in econometrics and statistics. In practice, it is often desirable to consider simultaneously multiple quantiles to obtain a complete analysis of the stochastic relationships between variables. This underlies the consideration of the conditional quantile process. A seminal contribution to this analysis 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 Fernandez-Val (2005) considered the same issue and suggested resampling as an alternative approach. Angrist et al. (2006) established inferential theory in misspecified models. 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.

The main focus of the above literature has been on parametric quantile models. However, there are frequent occasions where parametric specifications fail, making more flexible nonparametric methods desirable. This paper aims to achieve two goals. The first is to propose two simple nonparametric estimators for the conditional quantile process. The second is to derive an inferential theory that can be used for constructing uniform confidence bands and testing various hypotheses concerning conditional quantile processes.

The two proposed estimators are both based on local linear regressions (Fan et al., 1994; Yu and Jones, 1998), but differing in how they ensure the quantile monotonicity. Specifically, the first estimator applies local linear regressions to a grid of quantiles while imposing a set of linear inequality constraints, and then linearly interpolates between adjacent quantiles to obtain an estimate for the quantile process. The second estimator first applies local linear regressions to a grid of quantiles without any constraints and then applies rearrangement (Chernozhukov et al., 2010) if quantile crossing occurs. They share the following two features. First, the bandwidth parameter is allowed to vary across quantiles to adapt to data sparsity. This is important because data are typically more sparse near the tails of the conditional distribution. Second, the computation is feasible even for large sample sizes. More detailed

☆ First version: March 24, 2011. We thank the co-editor, the associate editor, two anonymous referees, Adam McCloskey, Denis Tkachenko, seminar participants at University of Iowa, BU, BC-BU mini conference, CUHK, 2011 Econometric Society Summer Meeting and 2011 Asian Meeting of the Econometric Society for useful comments and suggestions.

* Corresponding author.
E-mail addresses: qu@bu.edu (Z. Qu), jyoon@cmc.edu (J. Yoon).

http://dx.doi.org/10.1016/j.jeconom.2014.10.008
0304-4076/

comparisons between the two estimators are provided in Section 3 of the paper.

For inference, three sets of results are established. (1) We derive a uniform Bahadur representation for the unconstrained estimator (i.e., obtained without imposing the monotonicity constraint). This generalizes Theorem 2.1 in Koenker and Portnoy (1987) to the local linear regression setting. While being of independent interest, this representation forms a key step in proving the subsequent results. (2) We show that the first proposed estimator has the same first-order asymptotic distribution as the unconstrained estimator if a certain rate condition on the quantile grid is satisfied. Further, its asymptotic distribution is a continuous Gaussian process, whose critical values can be estimated via simulations by exploiting the fact that it is conditionally pivotal, drawing on the insights of Parzen et al. (1994) and Chernozhukov et al. (2009). (3) We show that the second proposed estimator follows the same asymptotic distribution as the first. This result broadens the application of rearrangement to the local linear regression context.

The inferential theory and methods can be used to analyze a wide range of issues. They include: (1) constructing a uniform confidence band for the conditional quantile process, (2) constructing a uniform confidence band for the difference or other linear functions of multiple such processes, and (3) testing distributional hypotheses such as the homogeneity or equality of quantile treatment effects, as well as first-order and second-order conditional stochastic dominance. They can also be potentially useful for constructing specification tests of parametrically specified conditional quantile processes. Studies considering the latter issue include Escanciano and Velasco (2010) and Mammen et al. (2013).

We evaluate the proposed methods using simulations and briefly summarize the results below. First, the two proposed estimators and the conventional quantile-by-quantile local linear estimator all perform similarly in terms of the integrated mean squared error criterion. This result confirms the finding that they all share the same limiting distribution. Second, the confidence band can have undercoverage because the bias term in the estimator can be difficult to estimate. This is not particular to our problem, but rather is a well known fact in the nonparametric literature. To address this issue, we suggest a simple modification that allows for a more flexible bias correction. The resulting confidence band is asymptotically conservative. Simulation evidence suggests that it has adequate coverage, even with small sample sizes, and that it is only mildly wider than the confidence band that uses the conventional bias correction.

As an empirical illustration, the paper considers a dataset from an experiment known as Project STAR (Student–Teacher Achievement Ratio). Two results emerge. First, the students in the upper quantiles of the test score distribution benefit more from the class size reduction. Second, the effect of the class size reduction is strongest for the classes taught by moderately experienced teachers (i.e., 6–8 years of experience). We also conduct hypotheses tests for treatment significance, homogeneity, equality as well as first order stochastic dominance. The results reconfirm the above two findings.

There are two key differences between this paper and Belloni et al. (2011). The first difference is in the estimation framework. Belloni et al. (2011) consider a series-based framework, where the conditional quantile function is modeled globally with a large number of parameters. The current paper is based on local linear regressions, where the quantile function is modeled locally by a few parameters and the modeling complexity is governed by the bandwidth. Consequently, different techniques are applied to establish the asymptotic properties of the estimators. The second difference is how the quantile monotonicity is achieved. Belloni et al. (2011) apply monotonization procedures to a preliminary series-based estimator, while in our first estimator the monotonicity enters directly into the estimation through inequality constraints.

When viewed from a methodological perspective, the current paper is related to the following two strands of literature. First, the first estimator is related to the studies on estimating nonparametric regression relationships subject to monotonicity constraints, where the main focus has been the monotonicity with respect to the covariate. For example, Mammen (1991) considered an estimator consisting of a kernel smoothing step and an isotonization step. Delecroix et al. (1996) studied a procedure that involves unconstrained smoothing followed by a constrained projection. He and Shi (1998) and Koenker and Ng (2005) considered smoothing splines subject to linear inequality constraints. In the current paper, the monotonicity constraint is with regards to the quantiles, giving rise to a different type of estimator than those discussed above, and requiring different techniques for studying its statistical properties. Note that He (1997), Dette and Volgushev (2008), Bondell et al. (2010) and Chernozhukov et al. (2010) considered monotonicity with respect to the quantiles. The connection with their works is discussed later in the paper. Second, there is an active literature that studies uniform confidence bands for nonparametric conditional quantile functions; see Hardle and Song (2010) and Koenker (2010). The former paper considered kernel-based estimators and obtained confidence bands using strong approximations. The latter considered additive quantile models analyzed with total-variation penalties and obtained confidence bands using Hotelling's tube formula. Their results are uniform in covariates but pointwise in quantiles. Therefore, their results and ours complement each other and, when jointly applied, allow one to probe a broad spectrum of topics.

The paper is organized as follows. Section 2 introduces the issue of interest. Section 3 presents the estimators while Section 4 establishes their asymptotic properties. Section 5 discusses the bandwidth selection. Section 6 shows how to construct uniform confidence bands and conduct hypothesis tests on the conditional quantile process. Section 7 reports simulation results. Section 8 contains an empirical application and Section 9 concludes. All proofs are in the two appendices, with Appendix A containing the proofs of the main results and Appendix B some auxiliary lemmas.

The following notation is used. The superscript 0 indicates the true value. $\|z\|$ is the Euclidean norm of a vector z. $1(\cdot)$ is the indicator function. supp(f) stands for the support of f. The symbols "$\Rightarrow$" and "$\to^p$" denote weak convergence under the Skorohod topology and convergence in probability, and $O_p(\cdot)$ and $o_p(\cdot)$ is the usual notation for the orders of stochastic magnitude.

2. The issues of interest

Let (X, Y) be an $\mathbb{R}^{d+1}$-valued random vector, where Y is a scalar response variable and X is an $\mathbb{R}^d$-valued explanatory variable (X does not include a constant). Let $f_{Y|X}(\cdot)$ and $f_X(\cdot)$ be the conditional density of Y and the marginal density of X. Denote the conditional cumulative distribution of Y given $X = x$ by $F_{Y|X}(\cdot|x)$ and its conditional quantile at $\tau \in (0, 1)$ by $Q(\tau|x)$, i.e.,

$$Q(\tau|x) = F_{Y|X}^{-1}(\tau|x) = \inf\left\{s : F_{Y|X}(s|x) \geq \tau\right\}.$$

In this paper, $Q(\tau|x)$ is modeled as a general nonlinear function of x and τ. We fix x and treat $Q(\tau|x)$ as a process in τ, where $\tau \in \mathcal{T} = [\lambda_1, \lambda_2]$ with $0 < \lambda_1 \leq \lambda_2 < 1$. Here, $\mathcal{T}$ falls strictly within the unit interval in order to allow the conditional distribution to have an unbounded support.

This paper has two goals. The first is to develop nonparametric estimators for the conditional quantile process. The second is to provide some asymptotic results that can be used for constructing uniform confidence bands and testing various hypotheses concerning $Q(\tau|x)$. Throughout the paper, we assume $\{(x_i, y_i)\}_{i=1}^n$ is a sample of n observations that are *i.i.d.* as (X, Y). The following examples illustrate the above issues of interest. More discussions will follow in Section 6.