European Journal of Operational Research 232 (2014) 671–678

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

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

Interfaces with Other Disciplines

Nonparametric quantile frontier estimation under shape restriction

Yongqiao Wang [a,*], Shouyang Wang [b], Chuangyin Dang [c], Wenxiu Ge [d]

[a] School of Finance, Zhejiang Gongshang University, Hangzhou, Zhejiang 310018, China
[b] Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China
[c] Department of Systems Engineering and Engineering Management, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong
[d] Department of Applied Mathematics, NanHai Campus, South China Normal University, Foshan, Guangdong 528225, China

ARTICLE INFO

Article history:
Received 30 August 2012
Accepted 26 June 2013
Available online 6 July 2013

Keywords:
Productivity and competitiveness
Production frontier
Quantile regression
Shape restriction
Concavity
Non-crossing

ABSTRACT

This paper proposes a shape-restricted nonparametric quantile regression to estimate the τ-frontier, which acts as a benchmark for whether a decision making unit achieves top τ efficiency. This method adopts a two-step strategy: first, identifying fitted values that minimize an asymmetric absolute loss under the nondecreasing and concave shape restriction; second, constructing a nondecreasing and concave estimator that links these fitted values. This method makes no assumption on the error distribution and the functional form. Experimental results on some artificial data sets clearly demonstrate its superiority over the classical linear quantile regression. We also discuss how to enforce constraints to avoid quantile crossings between multiple estimated frontiers with different values of τ. Finally this paper shows that this method can be applied to estimate the production function when one has some prior knowledge about the error term.

1. Introduction

Quantile frontier with $\tau \times 100\%$ acts as a benchmark for whether a decision making unit (DMU) achieves top τ efficiency. In statistics, this τ-frontier is merely the τ-quantile of random output conditional on various input levels. If one DMU produces output not less than the τ-frontier at its input level, the DMU is regarded as τ-efficient; otherwise, τ-inefficient. In contrast with the full frontier that envelopes the production possibility set, the τ-frontier partitions the production possibility set into two portions: upper τ-efficient and lower τ-inefficient.

Linear quantile regression (LQR: Koenker & Bassett JR, 1978) has been applied to τ-frontier estimation by Behr (2010) and Bernini, Freo, and Gardini (2004). The first paper studied the productivity of German savings, cooperative and commercial banks, and the second paper analyzed the efficiency of the Italian hotel industry. In both papers, τ-frontier is assumed to be linear with the input (or log-input). But this linear functional form may fail to describe complex dependence relationships between conditional quantiles and input levels. Real applications call for flexible nonparametric methods, instead of LQR that is characterized primarily by technical convenience.

This paper provides a concave nonparametric quantile regression (CNQR) for τ-frontier estimation. Similar to previously proposed nonparametric extensions in frontier analysis, e.g. convex nonparametric least squares (CNLS: Kuosmanen, 2008), corrected concave nonparametric least squares (C^2NLS: Kuosmanen & Johnson, 2010) and stochastic non-smooth envelopment of data (StoNED: Kuosmanen & Kortelainen, 2012), this estimator has a piecewise linear form and satisfies monotonicity and concavity. The main characteristics of CNQR are described as follows. First, it is robust to functional misspecification that could undermine conventional parametric regression methods. Second, with the CNLS formulation (14a–d) developed by Kuosmanen (2008), it adheres to the regularity conditions (monotonicity and concavity) implied by microeconomic theory. Third, it enables the transformation of continuously constrained shape-restricted quantile regression into a tractable linear program.

A nonparametric quantile regression model for τ-frontier estimation is proposed in Wang and Wang (2013), which applies shape-restricted support vector regression with pinball loss. Although this estimator is smooth, it may fail to be globally nondecreasing and concave, because it only imposes the first and second-order derivative constraints on the sample points, as that in Wang and Ni (2012). However, the CNQR estimator in this paper is globally nondecreasing and concave, which is a decided advantage over Wang and Wang (2013).

Quantile frontier in this paper differs significantly from partial frontier in Aragon, Daouia, and Thomas-Agnan (2005), Daouia and Simar (2007) and Martins-Filho and Yao (2008), which is also based on conditional quantiles of random output. In these papers, the α-partial frontier at a given input level $\mathbf{x}$ is the α-quantile of

* Corresponding author. Tel.: +86 571 28877720; fax: +86 571 28877705.
E-mail addresses: wangyq@zjgsu.edu.cn (Y. Wang), swang@iss.ac.cn (S. Wang), mecdang@cityu.edu.hk (C. Dang), wenxiuge@gmail.com (W. Ge).

0377-2217/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2013.06.049

random output conditional on input level not larger than $\mathbf{x}$, whereas in this paper the τ-quantile frontier at $\mathbf{x}$ is the τ-quantile conditional on input level right equal to $\mathbf{x}$. The α-partial frontier in the aforementioned studies is to evaluate whether one specified DMU achieves the largest α output among companies with inputs less than or equal to its input. Because comparing a DMU's output to that of other DMUs with substantially less input is a meaningless endeavor, we do not test their models in the experiments.

The rest of this paper is organized as follows. Section 2 presents the shape-restricted nonparametric quantile regression for τ-frontier estimation. Section 3 investigates the performance of CNQR through Monte Carlo simulations. Section 4 discusses how to enforce non-crossing constraints in estimating multiple frontiers with different values of τ. Section 5 introduces how to estimate the production function with CNQR by adjusting τ when the modeller has some prior knowledge about the error distribution. Section 6 concludes the paper.

In this paper, $\mathbb{R}^d_+$ is the nonnegative orthant of $\mathbb{R}^d$. All vectors are column vectors written in boldface, whereas their elements are written in plain letters. For example, $\boldsymbol{\beta}_i$ is a vector and z_i is the ith element of vector $\mathbf{z}$. Vectors $\mathbf{0}$ and $\mathbf{1}$ are the vectors of appropriate dimensions with all their components equal to 0 and 1, respectively. $\forall i = 1, \ldots, n$ and $\forall j = 1, \ldots, n$ are abbreviated as $\forall i$ and $\forall j$, respectively. $\forall h$ denotes $\forall h = 1, \ldots, H$.

2. Methodology

Consider the standard multiple-input, single-output, cross-sectional model in productivity analysis

$$Y = \varphi(\mathbf{X}) + \epsilon, \tag{1}$$

where $Y \in \mathbb{R}_+$ denotes the random output, $\mathbf{X} \in \mathbb{R}^d_+$ is the random d-dimensional input, ϵ represents the error term, and $\varphi : \mathbb{R}^d_+ \to \mathbb{R}_+$ is the nondecreasing and convex production function. Quantile frontiers should be recovered on the basis of independent samples $\mathcal{X} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ from this model.

Let $Y|\mathbf{x}$ be the random output Y given input level $\mathbf{x}$ and its conditional distribution be $F_{Y|\mathbf{x}}$, i.e.,

$$F_{Y|\mathbf{x}}(y) \triangleq \text{Prob}\{Y \leqslant y| \text{ with input } \mathbf{x}\}. \tag{2}$$

Then the τ-frontier at input level $\mathbf{x}$ is defined as

$$\varphi_\tau(\mathbf{x}) \triangleq \inf\{y \in \mathbb{R}_+ | F_{Y|\mathbf{x}}(y) \geqslant 1 - \tau\}. \tag{3}$$

If $F_{Y|\mathbf{x}}$ is continuous and strictly increasing, we have $\varphi_\tau(\mathbf{x}) = F^{-1}_{Y|\mathbf{x}}(1-\tau)$. The economic meaning of τ-frontier is intuitive: one DMU with production $(\mathbf{x},y)$ is τ-efficient if $y \geqslant \varphi_\tau(\mathbf{x})$; otherwise τ-inefficient. This τ-frontier divides the production possibility set into two parts: upper τ-efficient and lower τ-inefficient. For any given input level $\mathbf{x}$, the percentages of upper τ-efficiency and lower τ-inefficiency are τ and $1-\tau$, respectively.

τ-frontier can be indirectly acquired by a two-stage procedure: first, estimating the conditional distribution $F_{Y|\mathbf{x}}$; second, computing $\varphi_\tau(\mathbf{x})$ by inverting $F_{Y|\mathbf{x}}$. Usually the performance of this procedure heavily depends on the assumptions on the distribution of ϵ and the functional form of φ. To circumvent this model specification, we can directly estimate τ-frontier by building the dependence between $\mathbf{x}$ and $\varphi_\tau(\mathbf{x})$ with quantile regression (Koenker & Bassett JR, 1978). Quantile regression has found successful applications in predicting daily supermarket sales (Taylor, 2007) and retail credit risk assessment (Somers & Whittaker, 2007). This strategy can be described by the following optimization problem

$$\min_{\varphi_\tau} \sum_{i=1}^{n} \rho_\tau(y_i - \varphi_\tau(\mathbf{x}_i)) \tag{4}$$

where $\rho_\tau : \mathbb{R} \to \mathbb{R}_+$ is an asymmetric absolute loss defined as

$$\rho_\tau(t) = \begin{cases} (1-\tau)t & t > 0 \\ -\tau t & t \leqslant 0. \end{cases} \tag{5}$$

This loss is also known as pinball loss or elbow loss.

The development from parametric to nonparametric quantile regression for τ-frontier estimation is not straightforward. Some shape-related constrains must be imposed to guarantee that the estimated quantile frontier satisfies monotonicity and concavity. Or else the flexibility of nonparametric methods may lead to counter-intuitive predictions. The conditional quantile function $\varphi_\tau : \mathbb{R}^d_+ \to \mathbb{R}_+$ should be estimated by the following nonparametric quantile regression

$$\min_{\varphi_\tau \in \mathcal{F}_2} \quad \sum_{i=1}^{n} \rho_\tau(y_i - \varphi_\tau(\mathbf{x}_i)) \tag{6}$$

where $\mathcal{F}_2$ is the set of nondecreasing and concave functions.

The shape restriction $\varphi_\tau \in \mathcal{F}_2$ is nontrivial, because it involves an infinite number of inequality constraints. Problem (6) belongs to shape-restricted nonparametric regression that has a long history in statistical literature with seminal works dating back half a century (e.g. Brunk (1955) and Hildreth (1954)). The common shapes analyzed in nonparametric econometrics are monotone, convex (concave), supermodular and homogeneous. Typical applications of shape-restricted regression in economics include the estimation of utility functions and production functions. Statistical convergences of piecewise linear convex least squares regression can be found in Aguilera, Forzani, and Morin (2011), Seijo and Sen (2011) and Lim and Glynn (2012). One can refer to Magnani and Boyd (2009), Kim, Vandenberghe, and Yang (2010), Toriello and Vielma (2012) and Lee, Johnson, Moreno-Centeno, and Kuosmanen (2013) for computational issues of piecewise linear convex least squares regression.

Given that all penalties are imposed only on sample points, the nonparametric regression (6) can be solved with the following problem

$$\min_{\mathbf{z} \in \mathcal{F}_2^{\mathcal{X}}} \quad \sum_{i=1}^{n} \rho_\tau(y_i - z_i) \tag{7}$$

where $\mathcal{F}_2^{\mathcal{X}}$ is the set of all vectors $\mathbf{z} = (z_1, \ldots, z_n)' \in \mathbb{R}^n_+$, which admits a nondecreasing and concave function $\varphi_\tau : \mathbb{R}^d_+ \to \mathbb{R}_+$ such that $\varphi_\tau(\mathbf{x}_i) = z_i$ for all $i = 1, \ldots, n$. This shape-restricted frontier estimation can therefore be divided into two steps: first, identifying the fitted values $(\hat{z}_1, \ldots, \hat{z}_n)$ that solve problem (7); second, constructing a monotone and concave estimator which links these fitted points. These two steps are comprehensively analyzed in Sections 2.1 and 2.2, respectively.

2.1. Fitting

The critical issue in solving problem (7) is how to represent $\mathbf{z} \in \mathcal{F}_2^{\mathcal{X}}$ with common inequality constraints. According to Kuosmanen (2008), this nondecreasing and concave shape restriction can be represented by the following constraints

$$z_i = \alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i \quad \forall i \tag{8a}$$

$$\alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i \leqslant \alpha_j + \boldsymbol{\beta}_j'\mathbf{x}_i \quad \forall i, \forall j \tag{8b}$$

$$\boldsymbol{\beta}_i \geqslant \mathbf{0} \quad \forall i. \tag{8c}$$

The inequalities in (8b) enforce concavity, while the nonnegativeness of $\boldsymbol{\beta}_i$ in (8c) implements nondecreasingness.

Therefore, problem (7) is equivalent to the following linear program