Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Robust estimation in stochastic frontier models

Junmo Song[a,1], Dong-hyun Oh[b,1], Jiwon Kang[c,*]

[a] *Department of Computer Science and Statistics, Jeju National University, Jeju, Republic of Korea*
[b] *Department of Industrial Engineering, Inha University, Incheon, Republic of Korea*
[c] *Research Institute for Basic Sciences, Jeju National University, Jeju, Republic of Korea*

ARTICLE INFO

Article history:
Received 22 September 2015
Received in revised form 28 July 2016
Accepted 9 August 2016
Available online xxxx

Keywords:
Stochastic frontier model
Outliers
Robustness
Minimum density power divergence estimator

ABSTRACT

This study proposes a robust estimator for stochastic frontier models by integrating the idea of Basu et al. (1998) into such models. It is shown that the suggested estimator is strongly consistent and asymptotic normal under regularity conditions. The robust properties of the proposed approach are also investigated. A simulation study demonstrates that the estimator has strong robust properties with little loss in asymptotic efficiency relative to the maximum likelihood estimator. Finally, a real data analysis is performed to illustrate the use of the estimator.

1. Introduction

Technical efficiency (TE) measures have been used for several decades for benchmarking purposes. The concept of TE was first introduced by Farrell (1957). Two strands of TE measurement then developed in the late 1970s and early 1980s: data envelopment analysis (DEA), based on linear programming, and stochastic frontier analysis (SFA), which commonly uses parametric stochastic frontier (SF) models.

DEA is mainly used to measure TE scores in the research fields of managerial and economics studies. Since the DEA estimator often requires only input and output quantities, its empirical results are easy to understand and apply to develop policy implications. However, a weakness of the DEA estimator is that it is sensitive to extreme values, making it difficult to apply the estimator to data sets with outliers. Several attempts have been made to solve this problem. For example, Wilson (1993, 1995) suggested a method for detecting outliers and Cazals et al. (2002) proposed a robust estimator for the nonparametric frontier model. Simar (2003) employed the method of Cazals et al. (2002) to detect outliers by using classical DEA estimators, which is named as the order-*m* approach. Florens and Simar (2005) also proposed robust parametric estimators of nonparametric frontiers. Daouia et al. (2012) extended the idea of Cazals et al. (2002) to correct the inherent bias in the classical order-*m* approach. Quantile-based robust efficiency measurement techniques have also been developed by several studies such as Aragon et al. (2005), Wheelock and Wilson (2008), Daouia et al. (2010), and Bruffaerts et al. (2013).

The SFA framework is a counterpart to DEA in that it is a parametric approach. This means that the functional form, such as production or cost functions, needs to be assumed before estimating the TE score. One of the pioneering methodologies in the SFA framework was developed by Jondrow et al. (1982), who proposed a formula for separating a random error

* Corresponding author. Fax: +82 64 725 2579.
E-mail address: jwkang.stats@gmail.com (J. Kang).

[1] Both authors contributed equally to this work.

http://dx.doi.org/10.1016/j.csda.2016.08.005

Please cite this article in press as: Song, J., et al., Robust estimation in stochastic frontier models. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.08.005

component and a TE component. Owing to the ease of application, various models have been developed and SF models have been widely employed in efficiency measurement studies. For example, the approach suggested by Battese and Coelli (1995) provides the TE and determinants of the TE. Numerous statistical methods have been proposed for estimating SF models. For example, Park and Simar (1994) and Park et al. (1998) considered semiparametric estimation in SF panel models and Kumbhakar et al. (2007) introduced an approach for nonparametric SF models. Kopp and Mullahy (1990) and Van den Broeck et al. (1994) applied the generalized method of moments procedure and Bayesian method, respectively, to parametric SF models. Kneip et al. (2015) proposed an alternative approach for nonparametric SF models using penalized likelihood.

This study addresses the estimation of parametric SF models, particularly in the presence of very high- or low-performing observations, which can be outliers due to measurement errors or atypical observations drawn from the tails of an underlying distribution. These observations should be treated carefully because they can influence the estimation procedure. As is widely recognized in the literature, the maximum likelihood (ML) estimation method is influenced strongly by deviating observations like outliers or extreme values. Our simulation shows that applying the ML estimator (MLE) to the SF model suffers from the same problem, requiring the development of a robust estimation method to make it less sensitive to such deviating observations. However, to the best of our knowledge, little effort has been made in this regard.

The purpose of this study is to propose a robust estimator for SF models under the assumption that the SF model is not misspecified. We do not address the problem of misspecification of the chosen parametric model. To construct a robust estimator, we consider the estimation method based on divergence, which evaluates the discrepancy between any two probability distributions. The divergence-based estimation method has been used successfully in constructing robust estimators. For a review, refer to Pardo (2006) and Cichocki and Amari (2010) as well as the references therein. In this study, we employ density power divergence, as proposed by Basu et al. (1998) (henceforth BHHJ). BHHJ proposed a minimum density power divergence (MDPD) estimator (MDPDE) and demonstrated that it possesses, relative to the MLE, strong robust properties with little loss in asymptotic efficiency. Compared with other robust methods such as the minimum Hellinger distance estimation, the BHHJ method does not require any smoothing methods. Hence, it avoids the difficulty of selecting a bandwidth when estimating the nonparametric density estimation. For this reason, the BHHJ method can be applied conventionally to any parametric model to which ML estimation can be applied. For example, see Juárez and Schucany (2004), Fujisawa and Eguchi (2006), and Kim and Lee (2013).

The remainder of the paper is organized as follows. Section 2 reviews the BHHJ estimation method and proposes a robust estimator for SF models based on density power divergence. This section also examines the asymptotic and robust properties of the proposed estimator. In Section 3, we discuss our simulation study that compares the performance of the conventional MLE and MDPDE in the SFA framework. In Section 4, we analyze real data using both estimators. Lastly, Section 5 concludes the paper.

2. Robust estimation in SF models

This section reviews the MDPDE and integrates it into the SFA framework in order to estimate the TE.

2.1. MDPD estimator

In this subsection, we review the BHHJ estimation procedure that minimizes a density-based divergence measure.

Let f and g be probability densities. To measure the difference between f and g, BHHJ defined density power divergence, $d_\alpha(f,g)$, as follows:

$$d_\alpha(g,f) := \begin{cases} \displaystyle\int \left\{ f^{1+\alpha}(z) - \left(1+\frac{1}{\alpha}\right) g(z) f^{\alpha}(z) + \frac{1}{\alpha} g^{1+\alpha}(z) \right\} \mathrm{d}z, & \alpha > 0, \\ \displaystyle\int g(z)\left\{\log g(z) - \log f(z)\right\} \mathrm{d}z, & \alpha = 0. \end{cases} \tag{1}$$

Note that the divergence includes Kullback–Leibler divergence and the L_2-distance as special cases. Since $d_\alpha(f,g)$ converges to $d_0(f,g)$ as $\alpha \to 0$, the above divergence with $0 < \alpha < 1$ provides a smooth bridge between Kullback–Leibler divergence and the L_2-distance.

Consider a family of parametric distributions $\{F_\theta : \theta \in \Theta \subset \mathbb{R}^m\}$ possessing densities $\{f_\theta\}$ with respect to the Lebesgue measure, and let $\mathcal{G}$ be the class of all distributions having densities with respect to the Lebesgue measure. For a distribution $G \in \mathcal{G}$ with density g, the MDPD functional at G (i.e., $T_\alpha(G)$) with respect to $\{F_\theta : \theta \in \Theta\}$ is defined by

$$T_\alpha(G) = \underset{\theta\in\Theta}{\operatorname{argmin}}\ d_\alpha(g, f_\theta), \tag{2}$$

where it is assumed that $T_\alpha(G)$ exists and is unique, as will normally be the case. Note that when G belongs to $\{F_\theta\}$ (i.e., $G = F_{\theta_0}$ for some $\theta_0 \in \Theta$), $T_\alpha(G)$ becomes θ_0. Roughly speaking, $F_{T_\alpha(G)}$ can be considered to be a projection of G onto the space of $\{F_\theta : \theta \in \Theta\}$ in terms of the divergence, and $T_\alpha(G)$ becomes the target parameter of the MDPDE below.

Please cite this article in press as: Song, J., et al., Robust estimation in stochastic frontier models. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.08.005