European Journal of Operational Research 249 (2016) 771–783

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Interfaces with Other Disciplines

Controlling for spatial heterogeneity in nonparametric efficiency models: An empirical proposal

Francesco Vidoli[a,*], Jacopo Canello[b]

[a] Department of Political Science, University of Roma Tre, Italy
[b] School of Economics, The University of Queensland, Australia

ARTICLE INFO

Article history:
Received 9 December 2014
Accepted 24 October 2015
Available online 30 October 2015

Keywords:
Productive efficiency
Conditional nonparametric efficiency
Spatial heterogeneity
Industrial districts

ABSTRACT

This paper introduces an original methodology, derived by the robust order-m model, to estimate technical efficiency with spatial autocorrelated data using a nonparametric approach. The methodology is aimed to identify *potential competitors* on a subset of productive units that are identified through spatial dependence, thus focusing on peers located in close proximity of the productive unit. The proposed method is illustrated in a simulation setting that verifies the territorial differences between the nonparametric unconditioned and the conditioned estimates. A firm-level application to the Italian industrial districts is proposed in order to highlight the ability of the new method to separate the global intangible spatial effect from the efficiency term on real data.

1. Introduction

Since long time, the dynamic process leading firms to concentrate in specific regions and specialize in particular sectors has attracted the attention of many researchers, ranging from the seminal work of Max Weber to the pioneering contributions of Marshall. Overall, there is a wide consensus on the beneficial effects of locating in close proximity to other firms: these benefits, ranging from knowledge spillovers to increased availability of inputs, are commonly defined as agglomeration economies. The locational effect is especially relevant in specific territorial contexts, such as the Marshallian industrial districts, where the role of a number of tangible and intangible factors (summarized by Marshall through the concept of 'the industrial atmosphere') has proved to be particularly effective in influencing the performances of the local actors operating in these regions: specifically, the existence of economies external to firms and internal to districts tend to generate a competitive advantage associated with a higher efficiency in the overall production system that provides sustained benefits for the economic actors located in these regions.

Given the key role played by location in determining the performance of a firm, a number of recent contributions in the field of productive efficiency have attempted to propose empirical models designed to account for the spatial heterogeneity of the production units.

Within the parametric approaches, a number of contributions have proposed different specifications to control for the spatial effect. In this respect, the most commonly used strategy involves the inclusion of contextual variables (usually defined as Z) that are believed to affect efficiency: this methodology is also aimed to limit the influence of multiple and heterogeneous production models that may affect the final estimates. In this respect, the three main alternatives proposed by the recent literature are the following: (i) inclusion of variables related to the territorial context to correct the *"average levels"* of a baseline model (Hughes, Lawson, Davidson, Jackson, & Sheng, 2011); (ii) use of contextual factors to model the inefficiency part (Lavado & Barrios, 2010); (iii) implementation of the maximum likelihood method proposed by Battese and Coelli (1995) to simultaneously estimate the production function with the inefficiency function (Azzoni, Igliori, & Schettini, 2007). Recently, Areal, Balcombe, *** Tiffin (2012) have proposed an alternative specification to evaluate spatial effects, using a Bayesian approach and including a spatial lag directly into u_i: this approach allows the inefficiency term to be splitted into a spatial component and a specific term for each firm. The specification proposed by Areal et al. (2012) has been solved by Fusco and Vidoli (2013) using the maximum likelihood function.

On the other hand, the nonparametric literature lacks of specific research regarding *spatial* dependence. Even in two-stage models, such as that of Vidoli (2011) who proposed to use jointly nonparametric methods for estimating the frontier and additive methods

* Corresponding author. Tel.: +393932281327.
E-mail addresses: fvidoli@gmail.com, francesco.vidoli@uniroma3.it (F. Vidoli), jacopo.canello@uqconnect.edu.au (J. Canello).

http://dx.doi.org/10.1016/j.ejor.2015.10.050
0377-2217/

for the specification of the functional form, the spatial dependence should be controlled in the first stage in order to prevent bias in the second stage of the estimation process. Within the non-parametric frameworks, the traditional approach to isolate the effect of contextual variables from the efficiency estimates is based on a two-step procedure (see *e.g.* Simar & Wilson, 2007, 2011) that (i) estimates the efficiency scores nonparametrically and (ii) regress them on a set of environmental variables. A more general alternative involves the probabilistic formulation of the nonparametric conditional measures (see *e.g.* Badin, Daraio, & Simar, 2012; Daraio & Simar, 2007; Jeong, Park, & Simar, 2010): even in this case, the heterogeneity between units is essentially intended as dependent from a set of contextual variables Z and not from the specific territorial location. Recently, Florens, Simar, and Van Keilegom (2014), Mastromarco and Simar (2014) proposed an alternative flexible nonparametric two-step approach on conditional efficiencies with the aim of eliminating the dependence of production inputs/outputs on common factors and of avoiding the use of smoothing methods.

Both the parametric and the parametric spatial approaches towards technical efficiency share a common limitation, as they only account for the spatial effect through a specific set of variables Z, while overlooking the global spatial trend which is not always identifiable and/or measurable in a direct way: this issue is particularly serious in specific territorial contexts, such as the Marshallian Industrial Districts, where the role of intangible aspects, often statistically and economically difficult to capture, is particularly relevant. The choice of an incomplete or erroneous set of contextual variables Z can be particularly problematic in non-parametric settings, as ex-post modelling validation processes are not available in these contexts.

In an attempt to overcome this issue, the present paper proposes a novel framework that aims to incorporate the spatial dependence into a nonparametric efficiency model by accounting for the spatial proximity of peers rather than evaluating the exact relationships between X and Y and a set of contextual factors Z. The idea of limiting the efficiency analysis to a subset of firms rather than to the entire population of production units is drawn upon the intuition of Cazals, Florens, and Simar (2002) and Daraio and Simar (2005), who proposed an order-m model to overcome one of the main limitations of the nonparametric approaches, *i.e.* the strong sensitivity to outliers. This insight can be effectively adapted in a spatial setting to isolate the role of the surrounding economic environment, thus defining the concept of *potential competitors* in a more consistent spatial framework. The proposed method is believed to be particularly valuable in contexts where technical efficiency is influenced not only by the firm's ability to produce an optimal amount of output given a particular set of inputs, but also by the presence of potential competitors whose influence tends to decline with distance. The methodology is based on a three-step approach that involves: (i) the preliminary estimation of the optimal distance in terms of spatial autocorrelation, (ii) the identification of the set of units operating within the area identified using the estimated radius for each unit i and (iii) the subsequent solution of the optimization problem. In particular, the first step is solved by estimating the semivariogram, a spatial statistical tool that is widely used in spatial statistics and can be conveniently adapted in the field of efficiency analysis: contrary to previous work, the estimation is implemented through a flexible GAM approach, which allows to identify spatial dependencies within extremely narrow spatial bounds. This method is particularly convenient to detect spatial autocorrelation in regional contexts, such as the Marshallian Industrial Districts, characterized by strong spatial dependencies that tend to decrease exponentially with increasing distance.

The remainder of the paper is structured as follows: in Section 2, an overview of the nonparametric methodologies for the efficiency analysis is presented, focusing on the conditional order-m methods that are defined to account for spatial dependence. In Section 3, the novel methodology to control for spatial proximity in a nonparametric framework is introduced and discussed, presenting the main benefits associated with its implementation in specific regional contexts characterized by spatial intangible contextual variables and sharply decreasing spatial autocorrelation. In Section 4, a simulation exercise is proposed to assess the performance of the proposed estimator against the traditional nonparametric models in a local setting characterized by both global and local spatial autocorrelation. In Section 5, the proposed model is applied to a specific context, *i.e.* the Tuscan industrial districts, characterized by peculiar spatial features that are consistent with those discussed in the preceding sections, showing the benefits associated with the used of the novel framework. Finally, the concluding remarks and some potential endeavours for future research are discussed.

2. Estimating technical efficiency in a nonparametric framework: from the traditional DEA/FDH formulation to the order-*m* models with contextual variables

The traditional nonparametric framework for efficiency analysis can be effectively described by considering a production technology characterized by a set of inputs $x \in \mathbb{R}_+^p$ that are used by a Decision Making Unit (*DMU*) to produce a set of outputs $y \in \mathbb{R}_+^q$. In this context, the production set Ψ can be defined as the set of technically feasible combinations of (x, y), as follows:

$$\Psi = \{(x, y) \in \mathbb{R}^{p+q} \mid x \text{ can produce } y\} \tag{1}$$

and the DGP (Data Generating Process) can be presented as:

$$H(x, y) = \text{Prob}(X \leq x, Y \geq y). \tag{2}$$

Against this background, Ψ can also be defined as the support of $H(x, y)$.

Following the approach originally conceived by Debreu (1951) and Farrell (1957), the efficiency scores for a given production scenario $(x, y) \in \Psi$ can be defined in terms of both the maximum amount of output potentially producible (*output oriented*) or the minimum amount of input potentially usable (*input oriented*). The choice between these two alternatives is generally driven by the nature of the problem and/or by the underlying constraints.

In an input-oriented framework, the optimization problem can be expressed as follows:

$$\theta(x, y) = \inf\{\theta \mid (\theta x, y) \in \Psi\}. \tag{3}$$

Since Ψ is unknown and has to be estimated from a random sample of production units $\chi = \{(X_i, Y_i) | i = 1, \ldots, n\}$, in a deterministic frontier framework we assume that $\text{Prob}((X_i, Y_i) \in \Psi)$. In general, the aim is to estimate the support Ψ of the random variable (X, Y), where Ψ is assumed to be compact.

The most widely used nonparametric estimators in the literature are the Free Disposal Hull (*FDH*, Deprins, Simar, & Tulkens, 1984) and the Data Envelopment Analysis (*DEA*), both originally based on the Farrell–Debreu envelopment intuition. The two estimators differ in that the former does not require the assumption of convexity, given the benchmarks are limited to the productive combinations belonging to Ψ, while the latter admit that linear combinations of the efficient units should also be part of the frontier. This difference[1] is presented schematically in Fig. 1.

The *FDH* estimator is derived from the free disposal hull of the sample points X:

$$\widehat{\Psi}_{\text{FDH}} = \{(x, y) \in \mathbb{R}^{p+q} | y < Y_i, x \geq X_i, \quad i = 1, \ldots, n\}. \tag{4}$$

The *FDH* efficiency scores are obtained by plugging $\widehat{\Psi}_{\text{FDH}}$ into Eq. (3) in place of the unknown Ψ.

[1] For the DEA model, the graph shows both the Constant Returns to Scale (CRS) and the Variable Returns to Scale (VRS) assumption.