European Journal of Operational Research 245 (2015) 194–202

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Decision Support

Productivity and efficiency estimation: A semiparametric stochastic cost frontier approach ☆

Kai Sun [a,*], Subal C. Kumbhakar [b,c], Ragnar Tveterås [c]

[a] Salford Business School, University of Salford, Greater Manchester, M5 4NT, UK
[b] Department of Economics, State University of New York at Binghamton, NY 13902, USA
[c] University of Stavanger Business School, Stavanger, Norway

ARTICLE INFO

Article history:
Received 13 May 2014
Accepted 2 March 2015
Available online 6 March 2015

Keywords:
Economics
Cost function
Semiparametric smooth coefficient model
Stochastic frontier model
Decomposition

ABSTRACT

This paper proposes a flexible stochastic cost frontier panel data model where the technology parameters are unknown smooth functions of firm- and time-effects, which non-neutrally shift the cost frontier. The model decomposes inefficiency into firm and time-specific components and productivity change into inefficiency change, technical change and scale change. We then apply the proposed methodology to the Norwegian salmon production data and analyze technical efficiency as well as productivity changes.

1. Introduction

The estimation of technical inefficiency was pioneered by Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977) in a stochastic frontier (SF) production function framework. Since then the SF model has been extended in many directions accommodating both cross-sectional and panel data. The models use both primal and dual approaches such as the production, input or output distance functions (Coelli & Perelman, 1999). Although most of the models use parametric functional forms, there are SF models that deal with nonparametric formulations of the frontier function (Fan, Li, & Weersink, 1996; Kumbhakar, Park, Simar, & Tsionas, 2007, among others). Sun and Kumbhakar (2013) used a semiparametric model to introduce flexibility to the parametric production frontier model. More recently, attention has been paid to the non-traditional inputs, i.e., firm characteristics, policy variables as well as factors that describe the environment in which production takes place, in addition to the traditional inputs such as capital and labor. These non-traditional inputs or environmental factors are the exogenous factors that may neutrally or non-neutrally shift the frontier of the technology.

The role of the environmental factors have been recognized in the literature using both parametric and semiparametric models. For example, Bhaumik, Kumbhakar, and Sun (2014) considered a model in which the environmental variables shift the frontier neutrally by expressing the intercept as a function of the non-traditional inputs. On the other hand, Zhang, Sun, Delgado, and Kumbhakar (2012) considered a model in which the environmental variables shift the production function non-neutrally by expressing both the intercept and slope coefficients as functions of R&D. More specifically, Zhang et al. (2012) estimated these coefficients as unknown smooth functions of R&D. The unknown functions allow the environmental variables to affect the technology in a flexible manner without imposing any specific form in their effect on the production frontier.

In this article we focus on the semiparametric cost frontier model to estimate and decompose productivity and efficiency. The advantage of the cost function approach is that it explicitly recognizes endogeneity of the input variables. To separate firm effects from time-varying technical inefficiency, we consider a panel data cost frontier model where the coefficients of the cost function are allowed to vary over time in a flexible manner. To do so, the slope coefficients of the cost function are specified as nonparametric functions of the time trend. Inefficiency is assumed to be neutral and is captured by the intercept which is allowed to vary over time and across firms in a flexible manner. Thus, firm- and time-effects on inefficiency are completely flexible. Estimation of the model is done in three steps. The semiparametric cost function is estimated in the first step. Inefficiency is estimated in the second step in which we separate the firm-effects from both persistent and time-varying technical inefficiency (Kumbhakar, Lien, & Hardaker, 2014). In doing so, we make distributional assumptions on the inefficiency components as well as

☆ We would like to thank two anonymous referees for their insightful comments. We, alone, are responsible for any remaining errors.
* Corresponding author. Tel.: +44 161 295 4369.
E-mail address: ksun1@binghamton.edu (K. Sun).

http://dx.doi.org/10.1016/j.ejor.2015.03.003
0377-2217/

on the random firm-effects. Finally, we decompose and estimate the inefficiency and productivity change components based on the estimated cost frontier from the first two steps. In particular, inefficiency is decomposed into time-invariant and time-varying components, and productivity change is decomposed into various components including the scale change, technical change, and technical inefficiency change. Although firm-effects and persistent inefficiency do not affect productivity directly (i.e., they vanish upon differentiation with respect to time), omission of these effects is likely to bias the other parameters and hence productivity and its components. As an empirical example, we use a farm-level Norwegian salmon production data set. This is an unbalanced panel data set that covers 442 farms over 14 years. Fish farming is a highly relevant industry to study as a future increase in the global production of healthy proteins has to rely on innovations and productivity growth in aquaculture. Many aquaculture sectors are young or infant industries where one can expect substantial farm heterogeneity in terms of technology, efficiency and productivity. It is important to understand quantitatively the sources of productivity differentials across firms and their development over time, as this can be used in design of policies aimed at increasing productivity, e.g., public extension services and technology transfer programs. The model framework proposed in this paper enables us to perform these kinds of analyses, as it allows a rich and flexible decomposition of sources of firms' productivity and its growth over time.

The rest of this paper is organized as follows. Section 2 explains in detail how to estimate the semiparametric stochastic cost frontier model and to decompose the inefficiency and productivity within this modeling framework. Section 3 describes the data set used for the evaluation of the model. Section 4 reports and interprets the estimation results, and Section 5 concludes.

2. Semiparametric stochastic cost frontier model

The cost function that we consider in this article is:

$$\ln C_{it} = \theta(i, t) + \phi(t)' \ln B_{it} + v_{it} \tag{1}$$

where the total cost, C_{it}, is the cost of variable inputs for firm i in year t, B_{it} is a vector of covariates (output and input prices), the intercept, $\theta(i, t)$, is an unknown function of firm- and time-effects which captures both persistent and time-varying inefficiency (to be specified later), the slope coefficient vector, $\phi(t)$, is an unknown function of time trend, and v_{it} is the noise term. We call (1) a semiparametric stochastic cost function because the structure of the cost function is parametric (Cobb-Douglas or translog depending on the construction of the B_{it} variables), but the coefficients are nonparametric functions of time trend. In principle, they can be functions of a vector of environmental variables (call them z). However, the cost frontier function is something that is common to all firms in a given year. That is, if two firms have the same values of all the covariates (input prices and output), their optimal (minimum) cost will be the same. If the cost function coefficients are functions of the z variables, the cost frontier will depend on z; and furthermore, if these z variables vary across firms, the cost frontier will be firm-specific. This is counterintuitive, especially if one thinks of the cost frontier as the benchmark technology, the value of which changes with z but not the technology itself. We avoid this problem by letting the technology parameters change with only time but not across firms. By doing this, we allow the cost frontier to change over time nonparametrically so that it is the same for all firms in any given year. More specifically, we define the cost frontier as:

$$\ln C_{it} = \alpha(t) + \phi(t)' \ln B_{it} + m_{it} + v_{it}, \tag{2}$$

where $\alpha(t) = \min_i \theta(i, t)$ and $m_{it} = \theta(i, t) - \alpha(t)$. Note that the above frontier function coefficients are invariant across firms; and therefore, we have a separate frontier for each t, unless all the functional coefficients do not vary over time (which is a testable hypothesis).

We start with a Cobb-Douglas type stochastic cost frontier model with panel data:

$$\ln C_{it} = \theta(i, t) + \sum_{q=1}^{Q} \beta_q(t) \ln Y_{qit} + \sum_{k=1}^{K} \delta_k(t) \ln W_{kit} + \sum_{p=1}^{P} \gamma_p(t) \ln Q_{pit} + v_{it}, \tag{3}$$

where C_{it} is the total cost, Y_{qit} is the qth output, W_{kit} is the kth input price, and Q_{pit} is the pth quasi-fixed input (e.g., capital), for the ith firm at time t. The i and t inside the functional parameters capture firm- and time-effects in a fully flexible manner. The random shocks v_{it} are assumed to be distributed with mean zero and variance σ_v^2. Furthermore, it is assumed to be independent of firm- and time-effects as well as $\ln Y_q$, $\forall q$, $\ln W_k$, $\forall k$, and $\ln Q_p$, $\forall p$. The functional coefficients $\theta(i, t)$, $\beta_q(t)$, $\forall q$, $\delta_k(t)$, $\forall k$, and $\gamma_p(t)$, $\forall p$, are the intercept and slope coefficients. It is worth emphasizing the economic meaning of the functional coefficients. The time-varying $\beta_q(t)$ are cost elasticities with respect to outputs. Similarly, cost elasticities with respect to input prices are $\delta_k(t)$ which also vary over time. Both $\beta_q(t)$ and $\delta_k(t)$ should be non-negative for all q, k and t. Finally, the cost elasticities with respect to quasi-fixed factors, $\gamma_p(t)$, can be either positive or negative and the sign can also change over time. We will discuss this issue in the empirical section of the paper. These unknown functions can be estimated using the kernel-based nonparametric method, which will be explained in more details in the following section.

2.1. Estimation of functional coefficients

Since the cost function is homogeneous of degree one in input prices and this property has to hold at every data point, we impose this property before estimation of the cost function. Such a restriction can be imposed by using any one of the input prices as a numeraire and express all other input prices relative to the numeraire. Using W_1, the price of the first input as the numeraire, we re-write the cost function as:

$$\ln \widetilde{C}_{it} = \theta(i, t) + \sum_{q=1}^{Q} \beta_q(t) \ln Y_{qit} + \sum_{k=2}^{K} \delta_k(t) \ln \widetilde{W}_{kit} + \sum_{p=1}^{P} \gamma_p(t) \ln Q_{pit} + v_{it}, \tag{4}$$

where $\widetilde{C}_{it} = C_{it}/W_{1it}$, and $\widetilde{W}_{kit} = W_{kit}/W_{1it}$, $\forall k = 2, \ldots, K$.

The cost function in (4) is similar to the semiparametric smooth coefficient (SPSC) model (Fan & Zhang, 1999; Li, Huang, Li, & Fu, 2002). However, the traditional SPSC model assumes that all the coefficients, the intercept and slopes, are functions of the same covariates. In our case, the intercept is a function of firm- and time-effects while the slopes are functions of time only.

For estimation we perform a Robinson (1988) type transformation and re-write (4) as:

$$\ln \widetilde{C}^*_{it} = \sum_{q=1}^{Q} \beta_q(t) \ln Y^*_{qit} + \sum_{k=2}^{K} \delta_k(t) \ln \widetilde{W}^*_{kit} + \sum_{p=1}^{P} \gamma_p(t) \ln Q^*_{pit} + v_{it}, \tag{5}$$

where $\ln \widetilde{C}^*_{it} = \ln \widetilde{C}_{it} - E(\ln \widetilde{C}_{it}|i, t)$, $\ln Y^*_{qit} = \ln Y_{qit} - E(\ln Y_{qit}|i, t)$, $\ln \widetilde{W}^*_{kit} = \ln \widetilde{W}_{kit} - E(\ln \widetilde{W}_{kit}|i, t)$, and $\ln Q^*_{pit} = \ln Q_{pit} - E(\ln Q_{pit}|i, t)$. The conditional expectations, $E(\ln \widetilde{C}_{it}|i, t)$, $E(\ln Y_{qit}|i, t)$, $E(\ln \widetilde{W}_{kit}|i, t)$, and