Estimation and Prediction Using Belief Functions: Application to Stochastic Frontier Analysis

Orakanya Kanjanatarakul, Nachatchapong Kaewsompong, Songsak Sriboonchitta and Thierry Denœux

Abstract We outline an approach to statistical inference based on belief functions. For estimation, a consonant belief functions is constructed from the likelihood function. For prediction, the method is based on an equation linking the unobserved random quantity to be predicted, to the parameter and some underlying auxiliary variable with known distribution. The approach allows us to compute a predictive belief function that reflects both estimation and random uncertainties. The method is invariant to one-to-one transformations of the parameter and compatible with Bayesian inference, in the sense that it yields the same results when provided with the same information. It does not, however, require the user to provide prior probability distributions. The method is applied to stochastic frontier analysis with cross-sectional data. We demonstrate how predictive belief functions on inefficiencies can be constructed for this problem and used to assess the plausibility of various assertions.

1 Introduction

Many problems in econometrics can be formalized using a parametric model

$$(Y, Z)|x \sim f_{\theta,x}(y, z), \tag{1}$$

where Y and Z are, respectively, observed and unobserved random vectors, x is an observed vector of covariates and $f_{\theta,x}$ is the conditional probability mass or density function of (Y, Z) given $X = x$, assumed to be known up to a parameter vector $\theta \in \Theta$.

O. Kanjanatarakul
Department of Economics, Faculty of Management Sciences,
Chiang Mai Rajabhat University, Chiang Mai, Thailand

N. Kaewsompong · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand

T. Denœux (✉)
Heudiasyc (UMR 7253), Université de Technologie de Compiègne and CNRS,
Compiègne, France
e-mail: thierry.denoeux@utc.fr

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_12

For instance, in the standard linear regression model, $Y = (Y_1, \ldots, Y_n)$ is a vector of n independent observations of the response variable, with $Y_i \sim \mathcal{N}(x_i'\beta, \sigma^2)$, $Z = Y_{n+1}$ is an independent random value of the response variable distributed as $\mathcal{N}(x_{n+1}'\beta, \sigma^2)$, $x = (x_1, \ldots, x_{n+1})$ and $\theta = (\beta, \sigma^2)$. Having observed a realization y of Y (and the covariates x), we often wish to determine the unknown quantities in the model, i.e., the parameter θ (assumed to be fixed) and the (yet) unobserved realization z of Z. The former problem is referred to as *estimation* and the latter as *prediction* (or *forecasting*).

These two problems have been addressed in different ways within several theoretical frameworks. The three main theories are frequentist, Bayesian and likelihood-based inference. In the following, we briefly review these three approaches to introduce the motivation for the new method advocated in this paper.

Frequentist methods provide *pre-experimental* measures of the accuracy of statistical evidence. A procedure (for computing, e.g., a confidence or prediction interval) is decided before observing the data and its long-run behavior is determined by averaging over the whole sample space, assuming it is repeatedly applied to an infinite number of samples drawn from the same population. It has long been recognized that such an approach, although widely used, does not provide a reliable measure of the strength of evidence provided by specific data. The following simple example, taken from [6], illustrates this fact. Suppose X_1 and X_2 are iid with probability mass function

$$\mathbb{P}_\theta(X_i = \theta - 1) = \mathbb{P}_\theta(X_i = \theta + 1) = \frac{1}{2}, \quad i = 1, 2, \tag{2}$$

where $\theta \in \mathbb{R}$ is an unknown parameter. Consider the following confidence set for θ,

$$C(X_1, X_2) = \begin{cases} \frac{1}{2}(X_1 + X_2) & \text{if } X_1 \neq X_2 \\ X_1 - 1 & \text{otherwise.} \end{cases} \tag{3}$$

It is a minimum length confidence interval at level 75 %. Now, let (x_1, x_2) be a given realization of the random sample (X_1, X_2). If $x_1 \neq x_2$, we know for sure that $\theta = (x_1 + x_2)/2$ and it would be absurd to take 75 % as a measure of the strength of the statistical evidence. If $x_1 = x_2$, we know for sure that θ is either $x_1 - 1$ or $x_1 + 1$, but we have no reason to favor any of these two hypotheses in particular. Again, it would make no sense to claim that the evidence support the hypothesis $\theta = x_1 - 1$ with 75 % confidence. Although frequentist procedures do provide usable results in many cases, the above example shows that they are based on a questionable logic if they are used to assess the reliable of given statistical evidence, as they usually are. Moreover, on a more practical side, confidence and prediction intervals are often based on asymptotic assumptions and their true coverage probability, assuming it is of interest, may be quite different from the nominal one for small sample sizes.

The other main approach to statistical inference is the Bayesian approach, which, in contrast to the previous approach, implements some form of *post-experimental* reasoning. Here, all quantities, including parameters, are treated as random variables, and the inference aims at determining the probability distribution of unknown quantities, given observed ones. With the notations introduced above, the estimation and prediction problems are to determine the posterior distributions of, respectively, θ and Z, given x and y. Of course, this is only possible if one provides a prior probability distribution $\pi(\theta)$ on θ, which is the main issue with this approach. There has been a long-standing debate among statisticians about the possibility to determine such a prior when the experimenter does not know anything about the parameter before observing the data. For lack of space, we cannot reproduce all the arguments of this debate here. Our personal view is that no probability distribution is truly non-informative, which weakens the conclusions of Bayesian inference in situations where no well-justified prior can be provided.

The last classical approach to inference is grounded in the likelihood principle (LP), which states that all the information provided by the observations about the parameter is contained in the likelihood function. A complete exposition of the likelihood-based approach to statistical inference can be found in the monographs [6, 8] (see also the seminal paper of Barnard et al. [3]). Birnbaum [7] showed that the LP can be derived from the two generally accepted principles of sufficiency and conditionality. Frequentist inference does not comply with the LP, as confidence intervals and significance tests depend not only on the likelihood function, but also on the sample space. Bayesian statisticians accept the LP, but claim that the likelihood function does not make sense in itself and needs to be multiplied by a prior probability distribution to form the posterior distribution of the parameter given the data. The reader is referred to Refs. [6, 8] for thorough discussions on this topic. Most of the literature on likelihood-based inference deals with estimation. Several authors have attempted to address the prediction problem using the notion of "predictive likelihood" [4, 8, 18]. For instance, the predictive profile likelihood is defined by $L_x(z) = \sup_\theta f_{\theta,x}(y, z)$. However, this notion is quite different conceptually from the standard notion of likelihood and, to some extent, arbitrary. While it does have interesting theoretical properties [18], its use poses some practical difficulties [6, p.39].

The method described in this paper builds upon the likelihood-based approach by seeing the likelihood function as describing the plausibility of each possible value of the parameter, in the sense of the Dempster-Shafer theory of belief functions [9, 10, 20]. This approach of statistical inference was first proposed by Shafer [20] and was later investigated by several authors (see, e.g., [1, 23]). It was recently justified by Denœux in [11] and extended to prediction in [16, 17]. In this paper, we provide a general introduction to estimation and prediction using belief functions and demonstrate the application of this inference framework to the stochastic frontier model. In this model, the determination of the production frontier and disturbance parameters is an estimation problem, whereas the determination of the inefficiency terms is a prediction problem. We will show, in particular, how this method makes it possible to quantify both estimation uncertainty and random uncertainty, and to evaluate the plausibility of various hypothesis about both the production frontier and the efficiencies.

The rest of this paper is organized as follows. The general framework for inference and prediction will first be recalled in Sect. 2. This framework will be particularized to the stochastic frontier model in Sects. 3 and 4 will conclude the paper.

2 Inference and Prediction Using Belief Functions

Basic knowledge of the theory of belief functions will be assumed throughout this paper. A complete exposition in the finite case can be found in Shafer's book [20]. The reader is referred to [5] for a quick introduction on those aspects of this theory needed for statistical inference. In this section, the definition of a belief function from the likelihood function and the general prediction method introduced in [16] will be recalled in Sects. 2.1 and 2.2, respectively.

2.1 Inference

Let $f_{\theta,x}(y)$ be the marginal probability mass or density function of the observed data Y given x. In the following, the covariates (if any) will be assumed to be fixed, so that the notation $f_{\theta,x}(y)$ can be simplified to $f_\theta(y)$. Statistical inference has been addressed in the belief function framework by many authors, starting from Dempster's seminal work [9]. In [20], Shafer proposed, on intuitive grounds, a more direct approach in which a belief function Bel_y^Θ on Θ is built from the likelihood function. This approach was further elaborated by Wasserman [23] and discussed by Aickin [1], among others. It was recently justified by Denœux in [11], from three basic principles: the likelihood principle, compatibility with Bayesian inference and the least commitment principle [21]. The least committed belief function verifying the first two principles, according to the commonality ordering [12] is the consonant belief function Bel_y^Θ defined by the contour function

$$pl_y(\theta) = \frac{L_y(\theta)}{\sup_{\theta' \in \Theta} L_y(\theta')}, \tag{4}$$

where $L_y(\theta) = f_\theta(y)$ is the likelihood function. The quantity $pl_y(\theta)$ is interpreted as the plaussibility that the true value of the parameter is θ. The corresponding plausibility and belief functions can be computed from pl_y as:

$$Pl_y^\Theta(A) = \sup_{\theta \in A} pl_y(\theta), \tag{5a}$$

$$Bel_y^\Theta(A) = 1 - \sup_{\theta \notin A} pl_y(\theta), \tag{5b}$$

for all $A \subseteq \Theta$. The focal sets of Bel_y^Θ are the levels sets of $pl_y(\theta)$ defined as follows:

$$\Gamma_y(\omega) = \{\theta \in \Theta | pl_y(\theta) \geq \omega\}, \tag{6}$$

for $\omega \in [0, 1]$. These sets may be called plausibility regions and can be interpreted as sets of parameter values whose plausibility is greater than some threshold ω. When ω is a random variable with a continuous distribution $\mathscr{U}([0, 1])$, $\Gamma_y(\omega)$ becomes a random set equivalent to the belief function Bel_y^{Θ}, in the sense that

$$Bel_y^{\Theta}(A) = \mathbb{P}_{\omega}(\Gamma_y(\omega) \subseteq A) \tag{7a}$$

$$Pl_y^{\Theta}(A) = \mathbb{P}_{\omega}(\Gamma_y(\omega) \cap A \neq \emptyset), \tag{7b}$$

for all $A \subseteq \Theta$ such that the above expressions are well-defined.

Example 1 Let us consider the case where $Y = (Y_1, \ldots, Y_n)$ is an i.i.d. sample from a normal distribution $\mathscr{N}(\theta, 1)$. The contour function on θ given a realization y of Y is

$$pl_y(\theta) = \frac{(2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(y_i - \theta)^2\right)}{(2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(y_i - \overline{y})^2\right)} \tag{8a}$$

$$= \exp\left(-\frac{n}{2}(\theta - \overline{y})^2\right), \tag{8b}$$

where $\overline{y}$ is the sample mean. The plausibility and belief that θ does not exceed some value t are given by the upper and lower cumulative distribution functions (cdfs) defined, respectively, as

$$Pl_y(\theta \leq t) = \sup_{\theta \leq t} pl_x(\theta) \tag{9a}$$

$$= \begin{cases} \exp\left(-\frac{n}{2}(t - \overline{y})^2\right) & \text{if } t \leq \overline{y} \\ 1 & \text{otherwise} \end{cases} \tag{9b}$$

and

$$Bel_y(\theta \leq t) = 1 - \sup_{\theta > t} pl_x(\theta) \tag{10a}$$

$$= \begin{cases} 0 & \text{if } t \leq \overline{y} \\ 1 - \exp\left(-\frac{n}{2}(t - \overline{y})^2\right) & \text{otherwise.} \end{cases} \tag{10b}$$

The focals sets (6) are closed intervals

$$\Gamma_y(\omega) = \left[\overline{y} - \sqrt{\frac{-2\ln\omega}{n}}, \overline{y} + \sqrt{\frac{-2\ln\omega}{n}}\right]. \tag{11}$$

When ω has a uniform distribution on [0, 1], $\Gamma_y(\omega)$ is a closed random interval. The cdfs of its lower and upper bounds are equal, respectively, to the lower and upper cdfs (10a, 10b) and (9a, 9b). □

2.2 Prediction

The prediction problem can be defined as follows: having observed the realization y of Y with distribution $f_\theta(y)$, we wish to make statements about some yet unobserved data $Z \in \mathbb{Z}$ whose conditional distribution $f_{y,\theta}(z)$ given $Y = y$ also depends on θ. The uncertainty on Z has two sources: (1) the randomness of the generation mechanism of Z given θ and y and (2) the estimation uncertainty on θ. In the approach outlined here, the latter uncertainty is represented by the belief function Bel_y^Θ on θ obtained by the approach described in the previous section. The random generation mechanism for Z can be represented by a sampling model such as the one used by Dempster [9] for inference. In this model, the new data Z is expressed as a function of the parameter θ and an unobserved auxiliary random variable ξ with known probability distribution independent of θ:

$$Z = \varphi(\theta, \xi), \tag{12}$$

where φ is defined in such a way that the distribution of Z for fixed θ is $f_{y,\theta}(z)$.

When Z is a real random variable, a canonical model of the form (12) can be obtained as $Z = F_{y,\theta}^{-1}(\xi)$, where $F_{y,\theta}$ is the conditional cumulative distribution function (cdf) of Z given $Y = y$, $F_{y,\theta}^{-1}$ is its generalized inverse and ξ has a continuous uniform distribution in [0, 1]. This canonical model can be extended to the case where Z is a random vector. For instance, assume that Z is a two-dimensional random vector (Z_1, Z_2). We can write

$$Z_1 = F_{y,\theta}^{-1}(\xi_1) \tag{13a}$$

$$Z_2 = F_{y,\theta,Z_1}^{-1}(\xi_2), \tag{13b}$$

where $F_{y,\theta}$ is the conditional cdf of Z_1 given $Y = y$, F_{y,θ,Z_1} is the conditional cdf of Z_2 given $Y = y$ and Z_1 and $\xi = (\xi_1, \xi_2)$ has a uniform distribution in $[0, 1]^2$.

Equation (12) gives us the distribution of Z when θ is known. If we only know that $\theta \in \Gamma_y(\omega)$ and the value of ξ, we can assert that Z is in the set $\varphi(\Gamma_y(\omega), \xi)$. As ω and ξ are not observed but have a joint uniform distribution on $[0, 1]^2$, the set $\varphi(\Gamma_y(\omega), \xi)$ is a random set. It induces belief and plausibility functions defined as

$$Bel_y(Z \in A) = \mathbb{P}_{\omega,\xi}\left(\varphi(\Gamma_y(\omega), \xi) \subseteq A\right), \tag{14a}$$

$$Pl_y(Z \in A) = \mathbb{P}_{\omega,\xi}\left(\varphi(\Gamma_y(\omega), \xi) \cap A \neq \emptyset\right), \tag{14b}$$

for any $A \subseteq \mathbb{Z}$.

Example 2 Continuing Example 1, let us assume that $Z \sim \mathcal{N}(\theta, 1)$ is a yet unoberved normal random variable independent of Y. It can be written as

$$Z = \theta + \Phi^{-1}(\xi), \tag{15}$$

where Φ is the cdf of the standard normal distribution. The random set $\varphi(\Gamma_y(\omega), \xi)$ is then the random closed interval

$$\varphi(\Gamma_y(\omega), \xi) = \left[\overline{y} - \sqrt{\frac{-2\ln\omega}{n}} + \Phi^{-1}(\xi), \overline{y} + \sqrt{\frac{-2\ln\omega}{n}} + \Phi^{-1}(\xi) \right]. \tag{16}$$

Expressions (14a, 14b) for the belief and plausibility of any assertion about Z can be approximated by Monte Carlo simulation [16]. □

As remarked by Bjornstad [8], a prediction method should have at least two fundamental properties: it should be invariant to any one-to-one reparametrization of the model and it should be asymptotically consistent, in a precise sense to be defined. An additional property that seems desirable is compatibility with Bayesian inference, in the sense that it should yield the same result as the Bayesian approach when a prior distribution on the parameter is provided. Our method possesses these three properties. Parameter invariance follows from the fact that it is based on the likelihood function; compatibility with Bayes is discussed at length in [16] and consistency will be studied in greater detail in a forthcoming paper.

3 Application to Stochastic Frontier Analysis

In this section, we apply the above estimation and prediction framework to the stochastic frontier model (SFM). To keep the emphasis on fundamental principles of inference, only the simplest case of cross-sectional data will be considered. The model as well as the inference method will be in introduced in Sect. 3.1 and an illustration with simulated data will be presented in Sect. 3.2.

3.1 Model and Inference

The SFM [2] defines a production relationship between a p-dimensional input vector $\mathbf{x}_i$ and output Y_i of each production unit i of the form

$$\ln Y_i = \boldsymbol{\beta}' \ln \mathbf{x}_i + V_i - U_i, \tag{17}$$

where $\boldsymbol{\beta}$ is a vector of coefficients, V_i is an error term generally assumed to have a normal distribution $\mathcal{N}(0, \sigma_v^2)$ and U_i is a positive inefficiency term. Usual models for

U_i are the half-normal distribution $|\mathcal{N}(0, \sigma_u^2)|$ (i.e., the distribution of the absolute value of a normal variable) and the exponential distribution. The SFM is thus a linear regression model with asymmetric disturbances $\varepsilon_i = V_i - U_i$. The inefficiency terms U_i are not observed but are of particular interest in this setting.

Assuming U_i to have a half-normal distribution, let $\lambda = \sigma_u/\sigma_v$ and $\sigma^2 = \sigma_u^2 + \sigma_v^2$ be new parameters to be used in place of σ_u^2 and σ_v^2. Although the variance of U_i is not σ_u^2 but $(1 - 2/\pi)\sigma_u^2$, λ has an intuitive interpretation as the relative variability of the two sources of error that distinguish firms from one another [2]. Using the notations defined in Sect. 1, we have $Y = (Y_1, \ldots, Y_n)$, $Z = (U_1, \ldots, U_n)$ and $\theta = (\boldsymbol{\beta}, \sigma, \lambda)$. The determination of the inefficiency terms is thus a prediction problem.

3.1.1 Parameter Estimation

Assuming the two error components U_i and V_i to be independent, the log-likelihood function is [14, p.540]

$$\ln L_y(\theta) = -n \ln \sigma + \frac{n}{2} \log \frac{2}{\pi} - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{\varepsilon_i}{\sigma}\right)^2 + \sum_{i=1}^{n} \ln \Phi \left(-\frac{\varepsilon_i \lambda}{\sigma}\right). \quad (18)$$

The maximum likelihood estimate (MLE) $\widehat{\theta}$ can be found using an iterative nonlinear optimization procedure. Parameter β may be initialized by the least squares estimate, which is unbiased and consistent (except for the constant term) [14]. However, it may be wise to restart the procedure from several randomly chosen initial states, as the log-likelihood function may have several maxima for this problem. Once $\widehat{\theta}$ has been found, the contour function (4) can be computed. The marginal contour function for any subset of parameters is the relative profile likelihood function. For instance, the marginal contour function of λ is

$$pl_y(\lambda) = \sup_{\boldsymbol{\beta}, \sigma} pl_y(\theta). \quad (19)$$

3.1.2 Prediction

The main purpose of stochastic frontier analysis is the determination of the inefficiency terms u_i, which are not observed. The usual approach is to approximate u_i by $\mathbb{E}(U_i|\varepsilon_i)$, which is itself estimated by plugging in the MLEs and by replacing ε_i by the residuals $\widehat{\varepsilon}_i$. The main result is due to Jondrow et al. [15], who showed that the conditional distribution of U given ε_i, in the half-normal case, is that of a normal $\mathcal{N}(\mu_*, \sigma_*^2)$ variable truncated at zero, with

$$\mu_* = -\frac{\sigma_u^2 \varepsilon_i}{\sigma^2} = -\frac{\varepsilon_i \lambda^2}{1+\lambda^2} \tag{20a}$$

$$\sigma_* = \frac{\sigma_u \sigma_v}{\sigma} = \frac{\lambda\sigma}{1+\lambda^2}. \tag{20b}$$

The conditional expectation of U_i given ε_i is

$$\mathbb{E}(U_i|\varepsilon_i) = \frac{\lambda\sigma}{1+\lambda^2}\left[\frac{\phi(\lambda\varepsilon_i/\sigma)}{1-\Phi(\lambda\varepsilon_i/\sigma)} - \frac{\lambda\varepsilon_i}{\sigma}\right], \tag{21}$$

where ϕ and Φ are, respectively, the pdf and cdf of the standard normal distribution. As noted by Jondrow et al. [15], when replacing the unknown parameter values by their MLEs, we do not take into account uncertainty due to sampling variability. While this uncertainty becomes negligible when the sample size tends to infinity, it certainly is not when the sample is of small or moderate size.

To implement the approach outlined in Sect. 2.2 for this problem, we may write the cdf of U_i as

$$F(u) = \frac{\Phi[(u-\mu_*)/\sigma_*] - \Phi(-\mu_*/\sigma_*)}{1-\Phi(-\mu_*/\sigma_*)}\mathbb{1}_{[0,+\infty)}(u). \tag{22}$$

Let $\xi_i = F(U_i)$, which has a uniform distribution $\mathscr{U}([0,1])$. Solving the equation $\xi_i = F(U_i)$ for U_i, we get

$$U_i = \mu_* + \sigma_* \Phi^{-1}\left[\xi_i\left(1-\Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right) + \Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right]. \tag{23}$$

Replacing μ_* and σ_* by their expressions as functions of the parameters, we have

$$U_i = \varphi(\theta, \xi_i) = \frac{\lambda}{1+\lambda^2}\left\{-\varepsilon_i\lambda + \sigma\Phi^{-1}\left[\xi_i + \Phi\left(\frac{\varepsilon_i\lambda}{\sigma}\right)(1-\xi_i)\right]\right\} \tag{24}$$

with $\varepsilon_i = \ln y_i - \boldsymbol{\beta}' \ln \mathbf{x}_i$, which gives us an equation of the same form as (12), relating the unobserved random variable U_i to the parameters and the auxiliary variable ξ_i.

To approximate the belief function on $Z = (U_1, \ldots, U_n)$, we may use the Monte Carlo method described in [16]. More specifically, we randomly generate N $n+1$-tuples $(\omega^{(j)}, \xi_1^{(j)}, \ldots, \xi_1^{(n)})$ for $j = 1, \ldots, N$ uniformly in $[0,1]^{n+1}$. For $i = 1$ to n and $j = 1$ to N, we compute the minimum and the maximum of $\varphi(\theta, \xi_i^{(j)})$ w.r.t. θ under the constraint

$$pl_y(\theta) \geq \omega^{(j)}. \tag{25}$$

Let $[\underline{u}_i^{(j)}, \overline{u}_i^{(j)}]$ be the resulting interval. The belief and plausibility of any statement $Z \in A$ for $A \subset \mathbb{R}^n$ as defined by (14a, 14b) can be approximated by

$$Bel_y(Z \in A) \approx \frac{1}{N} \# \left\{ j \in \{1, \ldots, N\} | [\underline{u}_1^{(j)}, \overline{u}_1^{(j)}] \times \cdots \times [\underline{u}_n^{(j)}, \overline{u}_n^{(j)}] \subseteq A \right\}, \tag{26a}$$

$$Pl_y(Z \in A) \approx \frac{1}{N} \# \left\{ j \in \{1, \ldots, N\} | [\underline{u}_1^{(j)}, \overline{u}_1^{(j)}] \times \cdots \times [\underline{u}_n^{(j)}, \overline{u}_n^{(j)}] \cap A \neq \emptyset \right\}, \tag{26b}$$

where # denotes cardinality. We can also approximate the belief function on any linear combination $\sum_{i=1}^n \alpha_i u_i$ by applying the same transformation to the intervals $[\underline{u}_i^{(j)}, \overline{u}_i^{(j)}]$, using interval arithmetics. For example, the belief and plausibility of statements of the form $u_i - u_k \leq c$ can be approximated as follows:

$$Bel_y(u_i - u_k \leq c) \approx \frac{1}{N} \# \left\{ j \in \{1, \ldots, N\} | \overline{u}_i^{(j)} - \underline{u}_k^{(j)} \leq c \right\}, \tag{27a}$$

$$Pl_y(u_i - u_k \leq c) \approx \frac{1}{N} \# \left\{ j \in \{1, \ldots, N\} | \underline{u}_i^{(j)} - \overline{u}_k^{(j)} \leq c \right\}. \tag{27b}$$

3.2 Simulation Experiments

To illustrate the behavior of our method, we simulated data from model (17) with $p = 1$, $\boldsymbol{\beta} = (1, 0.5)'$, $\sigma_v = 0.175$ and $\sigma_u = 0.3$. We thus have, for this model, $\lambda = 1.7143$ and $\sigma = 0.3473$. Figure 1 displays the marginal contour functions of β_0, β_1, σ and λ for a simulated sample of size $n = 100$. These plots show graphically the plausibility of any assertion of the form $\theta_j = \theta_{j0}$. For instance, we can see from Fig. 1d that the plausibility of the assertion $\lambda = 0$ is around 0.6: consequently, the hypothesis that inefficiencies are all equal to zero is quite plausible, given the data.

Figures 2 and 3 show the true and estimated inefficiencies for 20 individuals in the above simulated sample of size $n = 100$ and in a simulated sample of size $n = 1{,}000$, respectively. For the belief function estimation, we give the quantile intervals for $\alpha = 5\,\%$ and $\alpha = 25\,\%$. The lower bound of the quantile interval [16] is the α quantile of the lower bounds $\underline{u}_i^{(j)}$ of the prediction intervals, while the upper bound is the $1 - \alpha$ quantile of the upper bounds $\overline{u}_i^{(j)}$. The larger intervals in the case $n = 100$ reflect the higher estimation uncertainty.

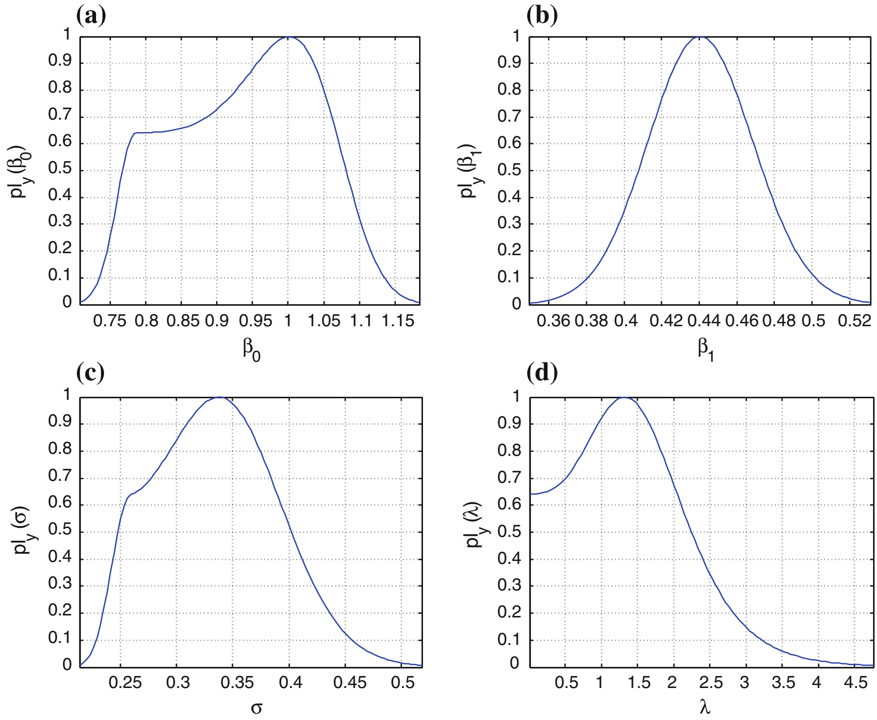

Fig. 1 Marginal contour functions for a simulated sample of size $n = 100$

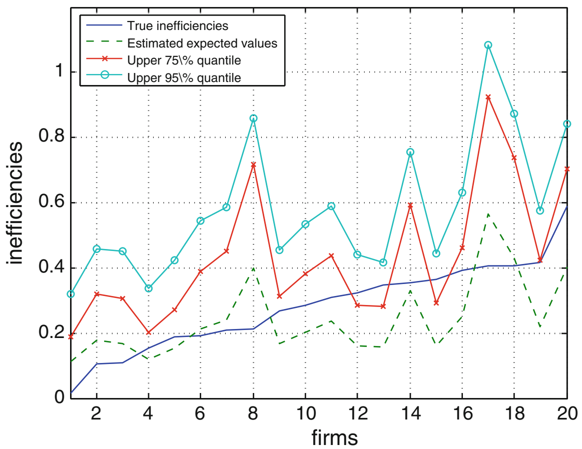

Fig. 2 True and predicted inefficiencies for 20 individuals in a simulated sample of size $n = 100$

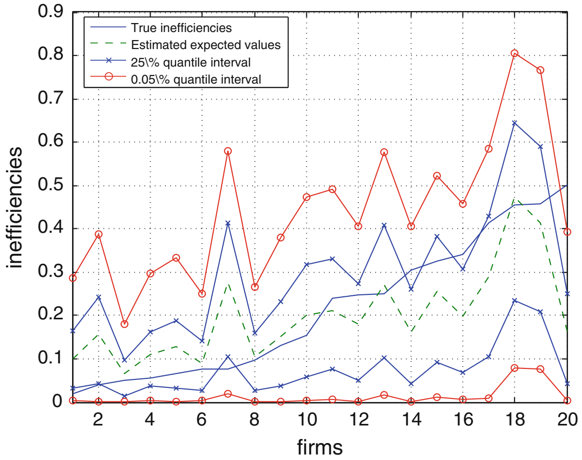

Fig. 3 True and predicted inefficiencies for 20 individuals in a simulated sample of size $n = 1{,}000$

4 Conclusions

We have shown how the estimation and prediction problems may be solved in the belief function framework, and illustrated these solutions in the case of the stochastic frontier model with cross-sectional observation. In the case of this model, the estimation problem concerns the determination of the model parameters describing the production frontier and the distributions of the noise and inefficiency terms, while the prediction problem consists in the determination of the unobserved inefficiency terms, which are of primary interest in this analysis. In our approach, uncertainties about the parameters and the inefficiencies are both modeled by belief functions induced by random sets. In particular, the random set formulation allows us to approximate the belief or plausibility of any assertion about the inefficiencies, using Monte Carlo simulation.

We can remark that parameters and realizations of random variables (here, inefficiencies) are treated differently in our approach, whereas there are not in Bayesian inference. In particular, the likelihood-based belief functions in the parameter space are consonant, whereas predictive belief functions are not. This difference is not due to conceptual differences between parameters and observations, which are just considered here as unknown quantities. It is due to different natures of the evidence from which the belief functions are constructed. In the former case, the evidence consists of observations that provide information on parameters governing a random process. In the latter case, evidence about the data generating process provides information about unobserved observations generated from that process.

The evidential approach to estimation and prediction outlined in this paper is invariant to one-to-one transformations of the parameters and compatible with Bayesian inference, in the sense that it yields the same result when provided with the same initial information. It is, however, more general, as it does not require the user to supply prior probability distributions. It is also easily implemented and does not require asymptotic assumptions, which makes it readily applicable to a wide range of econometric models.

The preliminary results reported in this paper need to be completed in several ways. First, a detailed comparison, based on underlying principles, with alternative approaches such as, e.g., the empirical Bayes method [19] or imprecise probabilities [22] remains to be performed. Secondly, it would be interesting to study experimentally how users interpret the results of the belief function analysis to make decisions in real-world situations. Finally, theoretical properties of our method, such as asymptotic consistency, are currently being studied.

References

1. Aickin, M.: Connecting Dempster-Shafer belief functions with likelihood-based inference. Synthese **123**, 347–364 (2000)
2. Aigner, D.J., Lovell, C.A.K., Schmidt, P.: Formulation and estimation of stochastic frontier production function models. J. Econom. **6**, 21–37 (1977)
3. Barnard, G.A., Jenkins, G.M., Winsten, C.B.: Likelihood inference and time series. J. R. Stat. Soc. **125**(3), 321–372 (1962)
4. Bayarri, M.J., DeGroot, M.H.: Difficulties and ambiguities in the definition of a likelihood function. J. Ital. Stat. Soc. **1**(1), 1–15 (1992)
5. Ben Abdallah, N., Mouhous Voyneau, N., Denœux, T.: Combining statistical and expert evidence using belief functions: application to centennial sea level estimation taking into account climate change. Int. J. Approx. Reason. **55**, 341–354 (2014)
6. Berger, J.O., Wolpert, R.L.: The likelihood principle: A Review, Generalizations, and Statistical Implications, 2nd edn. Lecture Notes-Monograph Series, vol 6. Institute of Mathematical Statistics, Hayward (1988)
7. Birnbaum, A.: On the foundations of statistical inference. J. Am. Stat. Assoc. **57**(298), 269–306 (1962)
8. Bjornstad, J.F.: Predictive likelihood: a review. Stat. Sci. **5**(2), 242–254, 05 (1990)
9. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. **38**, 325–339 (1967)
10. Dempster, A.P.: A generalization of Bayesian inference (with discussion). J. R. Stat. Soc. B **30**, 205–247 (1968)
11. Denœux, T.: Likelihood-based belief function: justification and some extensions to low-quality data. Int. J. Approx. Reason. **55**(7), 1535–1547 (2014)
12. Dubois, D., Prade, H.: A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets. Int. J. Gen. Syst. **12**(3), 193–226 (1986)
13. Edwards, A.W.F.: Likelihood (expanded edn.). The John Hopkins University Press, Baltimore (1992)
14. Greene, W.H.: Econometric Analysis, 6th edn. Prentice Hall, Upper Saddle River (2008)
15. Jondrow, J., Lovell, C.A.K., Materov, I.S., Schmidt, P.: On the estimation of technical efficiency in the stochastic production function model. J. Econom. **19**, 233–238 (1982)
16. Kanjanatarakul, O., Sriboonchitta, S., Denœux, T.: Forecasting using belief functions: an application to marketing econometrics. Int. J. Approx. Reason. **55**(5), 1113–1128 (2014)
17. Kanjanatarakul, O., Lertpongpiroon P., Singkharat S. and Sriboonchitta, S.: Econometric forecasting using linear regression and belief functions, In: Cuzzolin, F. (ed.), Belief Functions: Theory and Applications. Oxford, Springer, LNAI 8764, pp. 304–312 (2014)

18. Mathiasen, P.E.: Prediction functions. Scand. J. Stat. **6**(1), 1–21 (1979)
19. Robbins, H.: An empirical Bayes approach to statistics. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. Volume 1: Contributions to the Theory of Statistics, pp. 157–163 (1956)
20. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
21. Smets, Ph: Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. Int. J. Approx. Reason. **9**, 1–35 (1993)
22. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)
23. Wasserman, L.A.: Belief functions and statistical evidence. Can. J. Stat. **18**(3), 183–196 (1990)