Stochastic frontier estimation through parametric modelling of quantile regression coefficients

E. Fusco[1] · R. Benedetti[2] · F. Vidoli[3]

Abstract

Stochastic frontiers are a very popular tool used to compare production units in terms of efficiency. The parameters of this class of models are usually estimated through the use of the classic maximum likelihood method even, in the last years, some authors suggested to conceive and estimate the productive frontier within the quantile regression framework. The main advantages of the quantile approach lie in the weaker assumptions about data distribution and in the greater robustness to the presence of outliers respect to the maximum likelihood approach. However, empirical evidence and theoretical contributions have highlighted that the quantile regression applied to the tails of the conditional distribution, namely the frontiers, suffers from instability in estimates and needs specific tools and approaches. To avoid this limitation, we propose to model the parameters of the stochastic frontier as a function of the quantile in order to smooth its trend and, consequently, reduce its instability. The approach has been illustrated using real data and simulated experiments confirming the good robustness and efficiency properties of the proposed method.

✉ F. Vidoli
francesco.vidoli@uniurb.it

E. Fusco
fusco_elisa@libero.it

R. Benedetti
Benedett@unich.it

[1] Department of Economic Modelling and Statistical Analysis for Policy Making, SOGEI, Via Mario Carucci, 99, 00143 Rome, Italy

[2] Department of Economic Studies, University of Chieti-Pescara, Viale Pindaro, 42, 65127 Pescara, Italy

[3] Department of Economics, Society and Politics, University of Urbino Carlo Bo, Via Saffi 42, 61029 Urbino, Italy

1 Introduction

In the last decades, several methods have been proposed in literature to estimate the production or cost frontiers; this research stream—without claiming to be exhaustive—concerned the generalization of the baseline parametric stochastic frontier model (SFA, Kumbhakar and Lovell 2004; Kumbhakar et al. 2020) to the panel data (see e.g. Battese and Coelli 1995; Greene 2005), to the heterogeneity and spatial dependence (see e.g. Bille' et al. 2018; Fusco and Vidoli 2013; Tsionas and Michaelides 2016; Kutlu et al. 2020), to more flexible functional forms (semiparametric Fan et al. 1996, semi-nonparametric Kuosmanen 2012) and to the generalization of error term distributions (Greene 2003; Papadopoulos 2021).

A multitude of methods, therefore, have provided access to a plurality of explanatory possibilities concerning the heterogeneous behaviours of the production units in terms of efficiency.

All these methods, however, focused more on the correct methodological specification and properties than on the interpretative capabilities that the method could offer or rather on the concrete subsequent applicability in private or public policies; in other terms, the focus has been more on the long-run benchmark estimate—that is coincident with the estimated frontier—rather than on the partial benchmark references that can be useful in the short and medium term.

Quantile regression (QR, Koenker and Bassett 1978; Koenker 2005), conversely, can represent a steady approach to the long-term benchmark given that it can design gradual paths of return from inefficiency. In other terms, minimizing an asymmetrically weighted sum of absolute errors, quantile regression models allow to "*going beyond models for the conditional mean*" (Koenker and Hallock 2001) making it possible to derive different *partial benchmark references* for each quantile of the dependent variable analysed.

But it is exactly this "plurality of benchmark references" that is paradoxically the greatest weakness in the concrete use of these methods in the field of estimating production efficiency; how to identify the feasible production frontier? how to identify the stochastic part of the random noise?

Jradi et al. (2019) solves this crucial issue suggesting a heuristic algorithm to estimate the specific quantile of the conditional output distribution corresponding to the true stochastic frontier and paving the way for the use of quantile models in the field of efficiency estimation. Our proposed approach lives in this novel research stream combining a more general method such as the Frumento and Bottai (2016) quantile regression coefficients modelling (QRCM) approach with the Jradi et al. (2019) quantile selection method. This approach produces improvements both from an economic point of view, since it makes it possible to design gradual and consistent paths to recovery of inefficiency, and from a statistical point of view: in particular the absence

of assumptions on the error/inefficiency term and the robustness to outliers (Furno and Vistocco 2018) represent two crucial aspects in the practical application of frontier models.

The ultimate aim, therefore, of our paper is essentially methodological and would like to offer to the scientific and applicative debate an original flexible estimation method with the aim of bypassing the limitations (i) of the SFA in relation to assumptions about error distribution[1] and robustness in the presence of outliers and (ii) of the QR models in relation to the lack of monotonicity in the trend of the estimated coefficients as the quantile increases, or rather, in economic terms, to identify partial and feasible benchmark references that can be used to gradually reduce inefficiency. In other terms, the proposed approach aims to represent a method with clear methodological properties, but also rich in terms of application since it provides not only an estimate of inefficiency, but also and above all references that allow to identify partial benchmarks to overcome such inefficiency.

The remainder of the paper is organized as follows. In Sect. 2 methods and theoretical approaches relating to stochastic frontiers and quantile regressions will be outlined clarifying the methodological contribution of this paper; the properties of the proposed method and the application characteristics will be better highlighted on both some case studies (Sect. 3) and simulated data (Sect. 4). Section 5 is devoted to concluding remarks.

2 Frontier QRCM model

Standard quantile regression (Koenker and Bassett 1978; Koenker and Hallock 2001) is a regressive technique which aims to estimate the conditional τ_{th} quantile of a response variable y given covariates $\mathbf{x} = (x_1, \ldots, x_q)$, and—assuming a linear relationship between y and $\mathbf{x}$—it can be formulated as follows:

$$Q_y(\tau|\mathbf{x}) = x^T \beta(\tau) \tag{1}$$

where $\tau \in (0, 1)$ is the quantile and the coefficients vectors $\beta(\tau)$ are non-smooth functions. Parameter $\beta(\tau)$ plays a key role in QR models, but it can be highly variable in a random form for each quantile especially in the distribution tails (broken straight line in Fig. 1) leading to non-monotone increasing of the fitted functions (as shown in the first plot in Fig. 2). It should be noted that the shortcoming linked to the non-monotonicity represents a crucial flaw in the standard model in economic terms because it does not allow for the design coherent and consistent efficiency recovery policies.

Modelling the relationships between variables outside of the mean, QR is particular useful when outcomes are non-normally distributed and have nonlinear relationships with predictor variables relaxing the common regression assumptions and making no assumptions about the distribution of the residuals; given these premises, QR is less

---

[1] Furthermore, please note that there is no a priori economic justification to justify a given distribution of inefficiency, but that this crucial assumption is left to ex-post approximation against a set of known distributions in order to achieve the convergence of the SFA procedure.
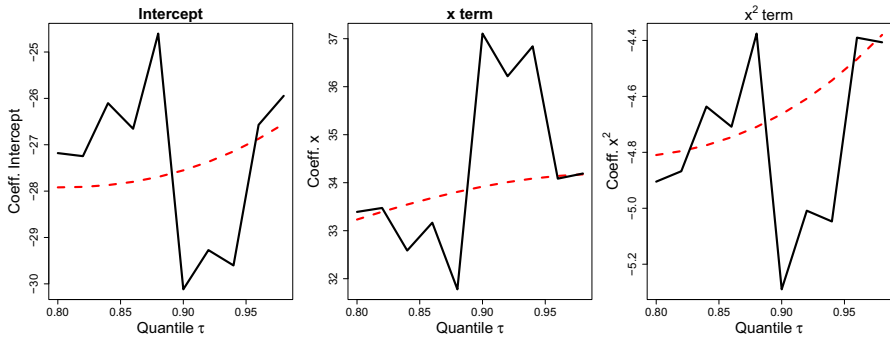
Fig. 1 Regression coefficients varying quantiles, $\tau \in [0.8, 1]$ *Note* Illustrative graph of QR coefficients estimation (continuous black line) and smoother QRCM function (dashed red line) varying quantiles τ for intercept (subgraph 1), variable x (subgraph 2) and variable x^2 (subgraph 3) on a generic sample with one input and one output. (Color figure online)

sensitive to extreme values than standard regression models proving its distributional robustness in the "*insensitivity to small deviations from the assumptions the model imposes on the data*" (Huber 1981).

Moving from the standard QR approach, Frumento and Bottai (2016, 2017) proposed to model the coefficient functions parametrically through a finite-dimensional parameter vector θ; their QRCM estimator can be, therefore, defined as:

$$Q_y(\tau|\mathbf{x}, \theta) = \mathbf{x}^T \beta(\tau|\theta) = \mathbf{x}^T \theta \mathbf{b}(\tau) \tag{2}$$

where θ parameters are estimated by minimizing the integral, with respect to τ, of the loss function of standard quantile regression, i.e.:

$$\overline{L_n}(\theta) = \int_0^1 L_n\{\beta(\tau|\theta)\}d\tau \tag{3}$$

The flexibility of the model lies, therefore, in the choice of function $\beta(\tau|\theta)$ which can be expressed as a k-th degree polynomial function, i.e. $\beta_j(\tau|\theta) = \theta_{j0} + \theta_{j1}\tau + \cdots + \theta_{jk}\tau^k$, $j = 1, \ldots, q$ or as other flexible functions as $[\log(\tau), \sqrt{-\log(1-\tau)}, -\sqrt{-\log(\tau)},\ldots]$ that better facilitate interpretation, prediction, and inference.[2]

For the sake of simplicity, in Fig. 1 the distributions of coefficients estimated with a standard QR estimator of a generic model $Q_y(\tau|\mathbf{x}, \theta) = \beta_0 + \beta_1(\tau|\theta)x + \beta_2(\tau|\theta)x^2 + \epsilon$ are reported; the continuous black line represents function $\mathbf{b}(\tau)$, namely the non-monotone trend of the coefficients estimated with the QR estimator, while the monotone dashed line in red $\beta_j(\tau|\theta)$ is the smoother QRCM function, i.e. a 3rd-degree shifted Legendre polynomial.

Continuous and monotone $\beta_j(\tau|\theta)$ functions allow to highlight two issues: (i) the first one is related to the ability to bypass the problems of instability of the estimates on

---

[2] In the application part (Sects. 3 and 4) several functional forms have been used; see Table 6 for a non-exhaustive list of application possibilities.
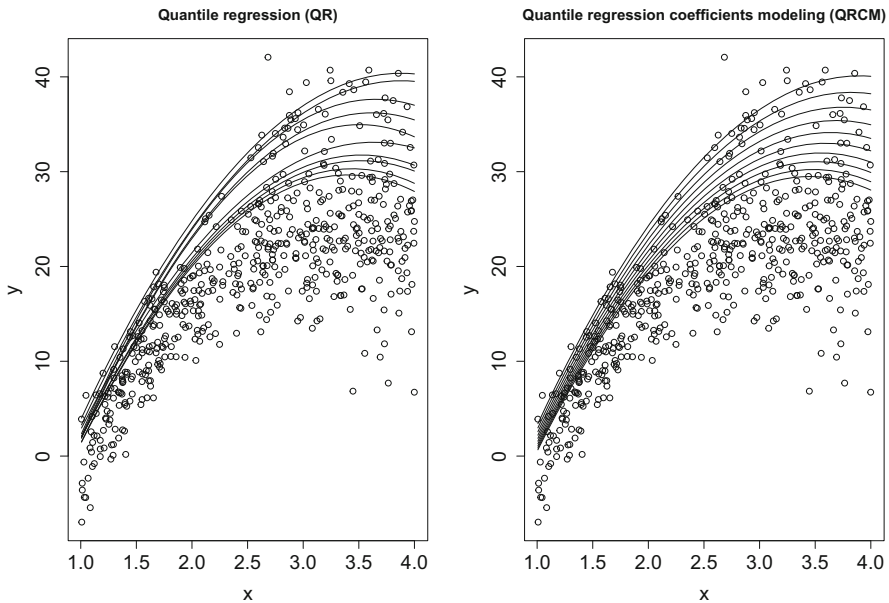
Fig. 2 QR and QRCM fitted curves varying quantiles *Note* Illustrative graph of QR (subgraph 1) and QRCM (subgraph 2) fitted curves on a generic sample with one input *x* (*x*-axis) and one output *y* (*y*-axis)

the extremal quantile as highlighted by many researchers (see e.g. Chernozhukov 2005) "*due to data sparsity*" (Li and Wang 2019) and "*heavy-tailed distribution*" (Huang and Nguyen 2017) and (ii) identify monotonous estimated curves that increase as the quantile rises and avoid quantile crossings between multiple estimated frontiers (as in Wang et al. (2014) proposal for nonparametric quantile regression) as shown in the second plot in Fig. 2; it is therefore possible, starting from these partial benchmark reference curves, define intermediate benchmarks useful in the short and medium term.

Sottile et al. (2019) suggested a penalized method that can address the selection of covariates in the QRCM modelling framework "*directly on the parameters of the conditional quantile function [and] using information on all quantiles*".

In recent years, QR has been used to estimate efficiency (Bernini et al. 2004; Liu et al. 2008; Roth and Rajagopal 2018) pointing out two improvements over models such as SFA that use maximum likelihood estimation: its robustness to the presence of outliers/abnormal points and its independence to the distributional choice providing a useful comparison for applied researchers.

Despite these methodological and empirical advantages, the main critical point has always been the discretionary choice of the "right" quantile corresponding to the production frontier: some authors, e.g. Knox et al. (2007), Liu et al. (2008) or Behr (2010) starting from the well-known finding that if no inefficiency is present in the sample the SFA frontier corresponds to an OLS estimation and hypothesizing that this is also true for the quantile regression (Horrace and Parmeter 2018), suggested (in a very subjective way, in our opinion) to choose the quantile, for production frontiers

estimation, above $\tau = 0.5$ (the median) and preferably from 0.8 to 0.975, to try to be as close as possible to a frontier but ignoring, however, distributional assumptions.

Jradi and Ruggiero (2019) and Jradi et al. (2019) finally solve this limitation proposing a heuristic method to choose the "right" quantile, demonstrating that, if the quantile is identified by considering the conditional distribution of the output given the regressors under a specific distributional setting of residuals, it is the one consistent with the location of the stochastic frontier. Tsionas (2020); Tsionas et al. (2020); Zhang et al. (2021) represent the latest methodological and empirical updates of this growing literature.

In SFA production setting, residuals are represented as a compound error $\epsilon = v - u$, assuming that v is the random term, i.e. $v \sim \mathcal{N}(0, \sigma_v^2)$ and u is the inefficiency term with a positive skewed distribution like the "Half-Normal" (Jradi et al. 2019), i.e. $u \sim \mathcal{N}^+(0, \sigma_u^2)$ or the "Exponential" (Jradi et al. 2021), i.e. $u \sim \text{Exp}(1/\sigma_u)$; given these premises, therefore, we refer to "wrong skewness" when the distribution of the term u presents a negative skewness. Finally, the compound error ϵ follows a negative skewed distribution as, respectively, "Normal Half-Normal" or "Normal Exponential".

Given these assumptions, and following the Jradi and Ruggiero (2019) proposal, the optimal quantile corresponding to the true location of the production frontier, in the case of "Normal Half-Normal" distribution, can be expressed as:

$$\tau^* = 0.5 + \frac{\arcsin(-E[\epsilon]/E[|\epsilon|])}{\pi} \tag{4}$$

where the term $\frac{-E[\epsilon]}{E[|\epsilon|]} = \frac{-\sigma_u}{\sigma}$ gives information about the quantity of inefficiency in the sample; following Fan et al. (1996), Jradi and Ruggiero (2019) also derived the $\lambda = \sigma_u/\sigma_v$ parameter that gives an immediate suggestion of the amount of inefficiency with respect to the noise; this parameter can be expressed as[3]:

$$\lambda = \tan(\pi(\tau^* - 0.5)) \tag{5}$$

Given these premises, the empirical algorithm for estimate the "right" quantile involves the iteration over different quantiles (e.g. $\tau = 0.5, 0.51, \ldots, 0.99$) and the comparison with the related likelihood choosing the one with the highest likelihood value in order to minimize $\tau^* - \tau$.

In this paper, the two methods outlined above are combined in order to gain the flexibility and independence from functional assumptions of the QRCM method with the objectivity in the choice of optimal production frontier of the Jradi and Ruggiero (2019) approach. In mathematical terms:

$$\begin{cases} Q_y(\widetilde{\tau}^* | \mathbf{x}, \theta) = \mathbf{x}^T \beta(\widetilde{\tau}^* | \theta) = \mathbf{x}^T \theta \mathbf{b}(\widetilde{\tau}^*) + \widetilde{\epsilon} \\ \widetilde{\tau}^* = 0.5 + \frac{\arcsin(-E[\widetilde{\epsilon}]/E[|\widetilde{\epsilon}|])}{\pi} \end{cases} \tag{6}$$

---

[3] This result is derived from the following—well known in SFA literature—equations $\sigma = \sigma_v \sqrt{1 + \lambda^2}$ and $\sigma^2 = \sigma_v^2 + \sigma_u^2$, by substituting $\frac{-E[\epsilon]}{E[|\epsilon|]} = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

where $\widetilde{\tau}^*$ is the "right" QRCM quantile obtained by estimating QRCM, also in this case, for different quantiles (e.g. $\tau = 0.5, 0.51, \ldots, 0.99$) and chosen by minimizing the difference $\widetilde{\tau}^* - \tau$.

Therefore, from a technical point of view, by imposing a parametrization and some degree of smoothness on coefficients β, the fitted values—and consequently the residuals $\widetilde{\epsilon}$—are estimated by using information on all quantiles simultaneously; following this approach, it is possible to inherit the other advantages of parametric modelling like parsimony, the ease of interpretation (Frumento and Bottai 2016) and the possibility to be applied to cases—latent variables, missing or partially observed data, causal inference—where "*parameters are harder to estimate in closed form*" (Waldmann 2018)—where applying standard QR proves to be difficult and computationally inefficient. Please note that a useful criterion for finding the best smoothness function can be using a goodness-of-fit test; in this paper, following Frumento and Bottai (2016), a Kolmogorov–Smirnov test has been considered[4] (more detailed information can be found in Sect. 3).

Moreover, the choice of the form, which leads to coefficients highly correlated on the frontier with SFA's ones, allows to include in the efficiency estimation not only all properties of the ML approach, but also to transfer them to lower quantiles.

On the other hand, from an economic point of view, the possibility to have increasing monotonous functions at the observed covariate values across quantiles lets to estimate different partial benchmark references in the short, medium and long term.

3 Properties of the proposed method: some case studies

In this Section, two empirical applications, based on two datasets well known in the literature, are proposed: in the first one (Sect. 3.1) the focus is on the comparison of QR/QRCM methods with respect to SFA; this is the most favourable scenario for SFA since there are no outliers in the data and the assumptions about the error distribution are met. In the second one (Sect. 3.2), instead, SFA estimation brings out wrong skewness on the inefficiency term showing very clearly the advantage of using QRCM-type estimation methods in this context.

3.1 Philippine rice farming dataset

In this Subsection, as previously stated, QRCM is compared to standard QR approach in order to bring out two evidences: (i) the QRCM capability to estimate more stable β parameters at quantile variations and (ii) the greater approximation, in terms of estimation, with respect to the SFA taken as a reference model given an optimal β's smooth function and the estimation of an optimal quantile.

Philippine rice farming dataset is widely used in literature to compare frontier methods (see for example Coelli et al. 2005, Rho and Schmidt 2015, Parmeter et al. 2019

---

[4] Other tests e.g. are the Chi-square (Snedecor and Cochran (1989)) or the Anderson–Darling (Stephens (1974)) test.

or Jradi et al. 2019). The dataset contains annual data collected from 43 smallholder rice producers in the Tarlac region of the Philippines between 1990 and 1997[5].

In this dataset, the output variable (y) is tonnes of freshly threshed rice and the main input variables are being area (area) of planted rice (hectares), total labour (labour) used (man-days of family and hired-labour) and fertilizer (npk) used (kilograms); the relative *translog* production frontier specification is defined as:

$$\begin{aligned}
\ln(y_i) = {} & \beta_0 + \beta_1 \ln(\text{area}_i) + \beta_2 \ln(\text{labour}_i) + \beta_3 \ln(npk_i) \\
& + \beta_{11} \ln(\text{area}_i)^2/2 + \beta_{12} \ln(\text{area}_i) \cdot \ln(\text{labour}_i) \\
& + \beta_{13} \ln(\text{area}_i) \cdot \ln(npk_i) + \beta_{22} \ln(\text{labour}_i)^2/2 \\
& + \beta_{23} \ln(\text{labour}_i) \cdot \ln(npk_i) + \beta_{33} \ln(npk_i)^2/2 \\
& + \theta t + v_i - u_i
\end{aligned} \tag{7}$$

Frontier specification reported in equation (7) has been estimated by the three methods; specifically, the QRCM approach[6] needed to identify the best smooth function for the quantile coefficients: this choice is clearly related to the empirical framework under consideration either by choosing the smooth function based on its theoretical properties or by using adjustment criteria.

In this application, following Frumento and Bottai (2016), a Kolmogorov-Smirnov goodness-of-fit test has been used for this purpose, indeed, they suggest to test the null hypothesis $H_0 : \tau_1, \ldots, \tau_n \sim U(0, 1)$, since by definition, at the true model $\tau_1, \ldots, \tau_n$ are independently and identically distributed draws from a standard uniform distribution. Moreover, with the aim to better approximate the functional form on the frontier, in this paper, a further criterion to select from the functional forms shown in Table 6 has been added, that is, a high correlation of the obtained βs with those of the SFA. The mix of the two approaches has led to the choice of the function $I(qnorm(\tau^3)) + I(\log(\tau))$[7].

In Fig. 3, the QR and the QRCM β coefficients smooth functions for each quantile from 0.6 and 1 are plotted, showing that QR coefficients are too volatile especially for quantile values greater than 0.8—those that are most relevant in terms of the production frontier—while QRCM smooth function is able to well approximate all translog terms.

In Table 1, a comparison of production translog frontier β coefficients, the efficiency specific parameters, namely, the λ, the total variance σ^2 and the mean Fan et al. (1996) efficiency values (standard deviations in brackets), estimated with corrected ordinary least squares (COLS, Winsten 1957), SFA, QR and QRCM, are reported. Moreover, the obtained optimal quantile τ^* for QR and QRCM are shown.

It can be noted that, QRCM method is able to approximate better, in terms of linear coefficients, the functional form of SFA with respect QR suggesting an economic interpretation closer to the SFA one. More in particular and similarly to Jradi et al. (2019), the optimal τ quantiles estimated with QR and QRCM: (i) are very similar

---

[5] In order to make our exercise comparable to Jradi et al. (2019)'s analysis, even if the dataset contains panel data, also in this paper, a cross-sectional setting is used by ignoring time effects.

[6] QRCM estimates have been carried out with the R package Mqrcm (Frumento 2021).

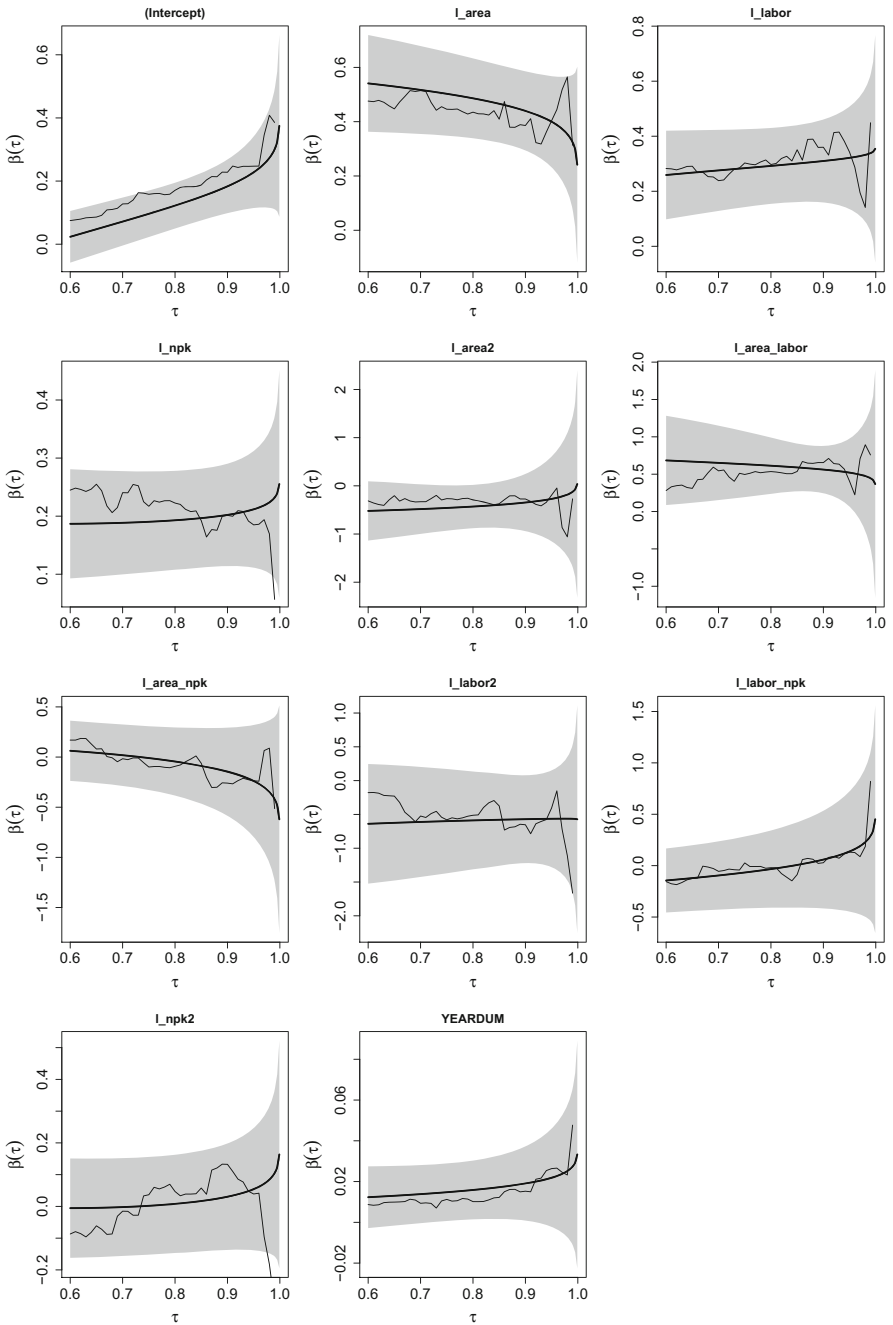[7] Test and correlation results can be found in Table 7.

Fig. 3 QR and QRCM regression coefficients by quantile—Philippines rice farming

Table 1 Parameters estimation results by method—Philippines rice farming

	COLS	SFA	QR	QRCM
Intercept	0.827	0.274∗∗∗	0.228∗∗∗	0.143∗∗∗
β_1	0.588∗∗∗	0.531∗∗∗	0.389∗∗∗	0.508∗∗∗
β_2	0.192∗∗	0.231∗∗∗	0.359∗∗∗	0.306∗∗∗
β_3	0.198∗∗∗	0.203∗∗∗	0.200∗∗∗	0.179∗∗∗
β_{11}	− 0.436∗	− 0.476∗∗	− 0.272	− 0.486
β_{12}	0.679∗∗∗	0.609∗∗∗	0.655∗∗	0.619∗∗∗
β_{13}	0.064	0.062	− 0.257	0.029
β_{22}	− 0.742∗∗	− 0.564∗∗	− 0.646	− 0.543
β_{23}	− 0.178	− 0.137	0.022	− 0.074
β_{33}	0.020	− 0.007	0.133	− 0.028
θ	0.013	0.015∗∗	0.015∗	0.014
λ	3.361	2.755	2.778	3.111
σ^2	0.098	0.222	0.107	0.114
τ^*	–	–	0.890	0.901
Efficiency	0.467	0.729	0.744	0.732
	(0.125)	(0.146)	(0.162)	(0.169)

* p<0.1; **p<0.05; ***p<0.01

to the one computed *a posteriori*, for a merely comparison purpose, for SFA and COLS frontiers (0.908 and 0.889, respectively); (ii) are close to the upper decile. The obtained average level of efficiency is about 0.467 for COLS and goes up to 0.729 for SFA, 0.744 for QR and 0.732 for QRCM. Moreover, the sum of the linear terms of the translog production frontier is close to one for all methods indicating a slightly decreasing returns of scale for the Philippines rice farms. Finally, the Spearman correlation index on efficiency values, in Table 8 in "Appendix", shows how QR and even further QRCM's rankings differ more from the COLS than those of SFA (respectively, 0.954, 0.931 and 0.989). Such a result, in this case, overcomes the criticism highlighted in Ondrich and Ruggiero (2001) claiming that *[...] we show that rankings for firm-specific inefficiency estimates produced by traditional stochastic frontier models do not change from the rankings of the composed errors. As a result, the performance of the deterministic models is qualitatively similar to that of the stochastic frontier models.[...].*

3.2 NBER manufacturing dataset

In this Subsection, the NBER manufacturing productivity dataset[8] (Bartelsman and Gray 1996) has been considered to highlight the properties of the proposed QRCM

---

[8] https://www.nber.org/research/data/nber-ces-manufacturing-industry-database.

method with respect to SFA when in the presence of "wrong skewness" problem[9] (Green and Mayes 1991).

Wrong skewness, in fact, may can be ascribed to incorrect or outlier data, an incorrect or incomplete specification of the production model[10] or to both making the choice of the form of the inefficiency term "*sometimes a matter of computational convenience*" Bonanno and Domma (2017).

This dataset, already used in literature to study the above mentioned problem, has been used primarily to propose new skewed densities for the compound error (see, among others, Li 1996; Carree 2002; Tsionas 2007; Almanidis and Sickles 2012; Almanidis et al. 2014; Bonanno and Domma 2017; Hafner et al. 2018) or the adjustment of the estimator for finite sample (see, among others, Simar and Wilson 2009; Cai et al. 2021). In our case, however, no error distribution is to be assumed a priori to obtain model convergence.

NBER dataset contains information on 473 US manufacturing industries for 54 years (from 1958 to 2011) and, by following Bonanno and Domma (2017) and Hafner et al. (2018), 54 sub-sectors from the textile industry over the years 1958-2011 are analysed. Even in this case and following Hafner et al. (2018) approach, with the aim to compare QRCM model with SFA and QR in terms of "wrong skewness", a cross-sectional estimation for each year has been carried out.

In this dataset, the output variable (y) is the total value added and, as input variables, total employment (*labour*), cost of materials (*materials*), energy cost (*energy*) and capital stock (*capital*) are used; the *Cobb-Douglas* production frontier specification is defined as:

$$\ln(y_i) = \beta_0 + \beta_1 \ln(\text{labour}_i) + \beta_2 \ln(\text{materials}_i) \\ + \beta_3 \ln(energy_i) + \beta_4 \ln(capital_i) + v_i - u_i \tag{8}$$

For each year, the production frontier has been estimated by using OLS, SFA (Normal-Half Normal, Normal-Exponential and Normal-t-Normal specification), QR and QRCM models.

More specifically, as resulted by the best approximation among functions in Table 6, the smooth function $I(qnorm(\tau)) + I(\log(\tau))$ has been chosen for QRCM for the most numbers of years.

Results are reported in Fig. 4; it can be seen that as long as the skewness is "correct" (values in the bottom plot below 0—years from 1958 to 1998) all methods work in a similar way, but in the presence of "wrong" skewness (values in the bottom plot above 0—for the last few years) the *a posteriori* τ^* parameter collapses to the median because it fails to estimate the inefficiency (no convergence of the maximum likelihood optimizer) for SFA, regardless of specifications of the residuals (dark green straight

---

[9] It occurs when the sign of the empirical OLS residuals skewness is positive instead of negative, when on the contrary as pointed out in Sect. 2 at page 13, in production efficiency $\epsilon = v - u$ and so follows a negative skewed distribution.

[10] Although this is not always true; Hafner et al. (2018) in particular claims that "*when observing the "wrong" skewness, most researchers are tempted to believe that the model is wrong, and we know that even a correct SFM allowing inefficient firms may produce the wrong sign for the skewness. This happens more often with small sample sizes or when the ratio Var(V)/Var(U) increases*".
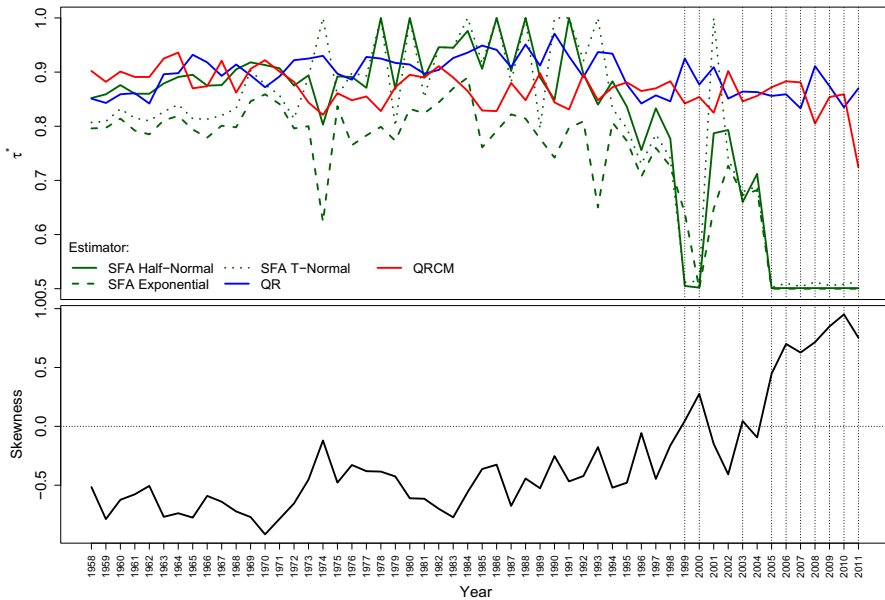
Fig. 4 Optimal quantile comparison, by method, in the presence of wrong skewness

line for the Half-Normal, dark green dashed line for Exponential and dark green dotted for the t-Normal).

Finally, in Table 9 in the "Appendix", the estimated efficiency results, by method, are reported. In particular, it is noteworthy that in the years where residuals present a "wrong" skewness the efficiency is close to 1 for SFA as it fails to detect inefficiency, unlike QR and QRCM which are able to estimate it.

4 Simulations

The aim of this section is to assess, in a more systematic way, the properties of the QRCM model both in terms of estimating the frontier and the inefficiency of individual units. SFA and Jradi et al. (2019) QR have been chosen as contrasting methods as they represent the natural comparison both on the side of stochastic efficiency and quantile approach. The production simulation setting mimics the Banker and Natarajan (2008) proposal—also followed by Johnson and Kuosmanen (2011)—generating sample data by a cubic polynomial in x:

$$\phi(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 \tag{9}$$

choosing $\alpha_0 = -37$, $\alpha_1 = 48$, $\alpha_2 = -12$ and $\alpha_3 = 1$ in order to ensuring the monotonicity and the concavity in the range $x = [1, 4]$. Finally, in the efficiency setting, two key parameters must be defined: the error term v set, as usually, from a two-side Normal distribution $\mathcal{N}(\mu_v, \sigma_v)$ with $\mu_v = 0$ and $\sigma_v = 1$ and the inefficiency

term u which will be varied, in the following simulations, in absolute terms and distributional form.

After drawing the random variables x as a uniform[1,4] for 200 units, the logarithm of the output $y = \phi(x)$ has been set as:

$$\ln(y) = \ln(-37 + 48x - 12x^2 + x^3) + v - u \tag{10}$$

Finally, two measures have been used to evaluate the performance of the proposed model against both the simulated frontier and the SFA and QR methods:

- the mean squared error (MSE), that is, the average squared difference between the simulated and the estimated values, in order to verify the accuracy of the frontier estimate; $\text{MSE} = 1/n \sum_{i=1}^{n} (y_i - \widehat{y_i})^2$;
- the average of the absolute value of the differences (Mean abs diff.) between estimated and true efficiencies with the aim of evaluating the models on the efficiency estimation side; $\text{Mean\_diff} = 1/n \sum_{i=1}^{n} |(\text{eff}_i - \widehat{\text{eff}_i})|$.

4.1 First simulation: half-normal inefficiency

Once the general framework of the simulation had been set up, some settings have been varied in order to assess the stability and flexibility of the models.

In this first simulation, the inefficiency has been generated from a Half-Normal distribution with parameters $\mu_u = 0$ and $\sigma_u \in [0.6, 1.2, 1.8, 2.4, 3]$; as a result, keeping in mind that σ_v is equal to 1, in the following simulations $\lambda = \sigma_u / \sigma_v$ is equal, respectively, to [0.6, 1.2, 1.8, 2.4, 3]. Moreover, the choice of $\beta(\tau|\theta)$ has been set among all those proposed in Table 6; this simulation setting is, therefore, the more favourable for the SFA since inefficiency follows a standard Half-Normal distribution and no outlier/out-of-scale data are present.

Figure 5 allows to verify how the three models (QRCM, SFA and QR) substantially allow a good fit to the frontier and how this result is quite stable as the inefficiency varies.

From a preliminary analysis (Fig. 6), the QR estimator seems to be substantially less accurate than the corresponding QRCM estimator (as stable as the quantile chosen by the Jradi et al. (2019) algorithm); these initial impressions will be hereafter verified.

The setting proposed has been the starting point for checking the performance of the three chosen methods in terms of frontier fitting and efficiency estimation.

Table 2 reports mean and standard deviation for MSE and mean difference in absolute value of efficiencies over 1000 iterations varying λ.

In terms of MSE, it can be seen that the difference between quantile-based methods and SFA tends to decrease as the inefficiency in the data increases, while the QRCM ever outperforms QR.

This result is confirmed—even more clearly—by the average difference in absolute value between the inefficiency estimates and the true values, confirming a substantial equivalence of the methods under analysis in the case most favourable to the SFA, i.e. that in which the form of the inefficiency is Half-Normal and no outliers are included.
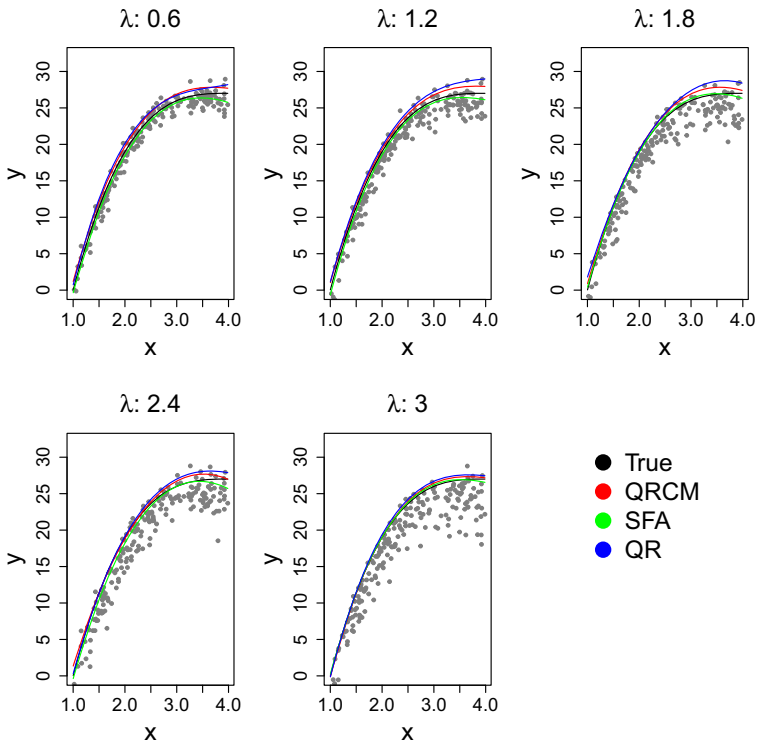
Fig. 5 Frontier estimation by method and level of inefficiency

But what if the form of the inefficiency is no longer standard, an issue that often occurs in real-world data? Section 4.2 will try to answer this question by varying the inefficiency distribution and including outliers in the simulated data.

4.2 Second simulation: varying inefficiency distribution

In this second simulation, therefore, always starting from the baseline setting proposed in Sect. 4, the finite sample performance of the proposed estimator has been examined by means of Monte Carlo simulation (1000 replications) by varying the distributional form of inefficiency and for three percentage levels of outliers of the total number of cases (1%, 3%, 5%); more specifically, outliers have been generated according to equation (9) in which the term α_0 has been set equal to -32.

More specifically, six different distributions (see Table 3 and Fig. 7) for the u term have been chosen: *(1)* Half-Normal, with the aim of verifying the impact of outliers; *(2)* Skew-Normal (Azzalini and Valle 1996) with high positive marginal skewness; *(3)* Skew-Normal with low positive marginal skewness; *(4)* Skew-Normal with low negative marginal skewness; *(5)* Skew-Normal with high negative marginal skewness; *(6)* Gamma.
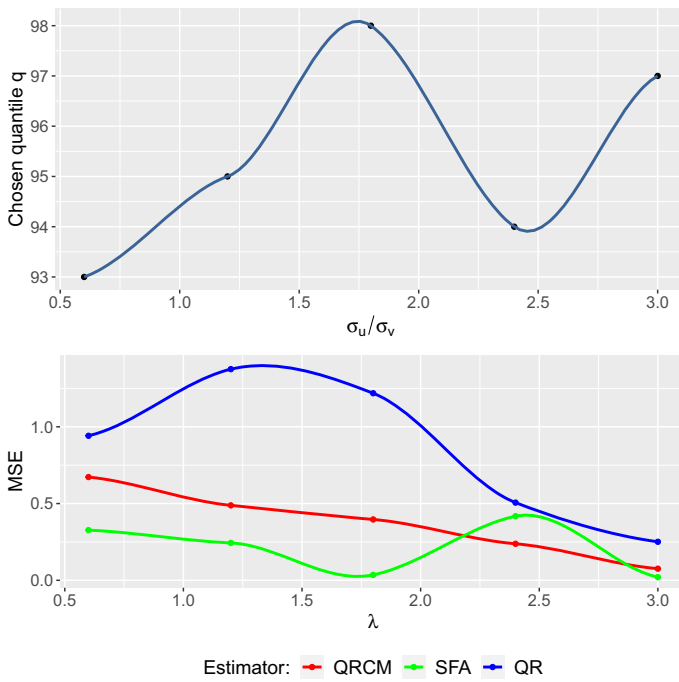
Fig. 6 MSE by method and level of inefficiency

Table 2 MSE and mean absolute difference for efficiencies by method and level of inefficiency

Method	λ	MSE		Mean abs diff.	
		Mean	Sd	Mean	Sd
QRCM	0.6	0.70	0.31	0.31	0.04
SFA		0.17	0.12	0.25	0.07
QR		1.17	0.57	0.36	0.04
QRCM	1.2	0.38	0.23	0.22	0.02
SFA		0.20	0.30	0.24	0.11
QR		0.80	0.53	0.25	0.03
QRCM	1.8	0.23	0.17	0.17	0.01
SFA		0.14	0.27	0.18	0.06
QR		0.59	0.47	0.19	0.02
QRCM	2.4	0.19	0.18	0.15	0.01
SFA		0.12	0.18	0.15	0.03
QR		0.60	0.71	0.16	0.02
QRCM	3	0.20	0.26	0.13	0.02
SFA		0.12	0.13	0.13	0.01
QR		0.55	0.68	0.13	0.02

Table 3 Summary statistics—u term

	Half Normal	Skew Normal ++	Skew Normal +	Skew Normal −	Skew Normal –	Gamma
Mean	2.374	2.903	2.936	2.938	2.938	1.925
SD	1.786	1.765	1.677	1.48	1.432	1.812
Skewness	0.945	0.768	0.455	− 0.075	− 0.277	1.509

Not all distributions could be expressed in terms of mean and variance like the Half Normal; therefore, in order to make the simulations comparable, the parameters of each distribution (for analytical specification of parameters, please see Table 10) have been set in such a way as to obtain similar mean and variance; Table 3 verifies this result (results are reported for $\sigma_u = 3$; similar results for $\sigma_u = 1$ are available from the authors) by also highlighting another key parameter, namely skewness, which—as highlighted in Sect. 2—in SFA models is necessary to be positive in order to obtain convergence.

Table 4 and Table 5 show, respectively, the average values of the MSE and the absolute mean difference for efficiencies by method, distribution of inefficiency and percentage of outliers included in the simulated data. Some implications arise:

- SFA approach—in the case of Half-Normal distribution—proves to be very sensitive to the presence of outliers; this result is most evident when inefficiency in the data is strongest.
- In the case of "wrong skewness" (negative Skew-Normal distributions) SFA does not converge—as already highlighted in Sect. 3.2—in all iterations and, therefore, performs worse than quantile models; this effect is increasing as the skewness of the u term decreases reflecting the fact that, as soon as the inefficiency data depart from the standard assumptions, the SFA model tends to inaccurately estimate the production frontier.
- QRCM performs better than QR both in terms of MSE and absolute difference for all inefficiency distributions and all percentage levels of outliers.

5 Final remarks

In this paper, the effects on efficiency estimates of the presence of outliers in the observed data and the non-occurrence of distributional assumptions have been analysed. After a brief review of some recent developments in robust estimation of stochastic boundaries, based on quantile regression approaches, the use of a variant of these methods based on modelling β parameters as a function of the quantile τ has been proposed. We have focused on this approach because the introduction of a similar model, very flexible and smooth, is fast and stable and offers a very practical approach to a solid estimate of the efficiencies not sensitive to the presence of anomalous data and to distributions even very far from the classical Half-Normal and Exponential
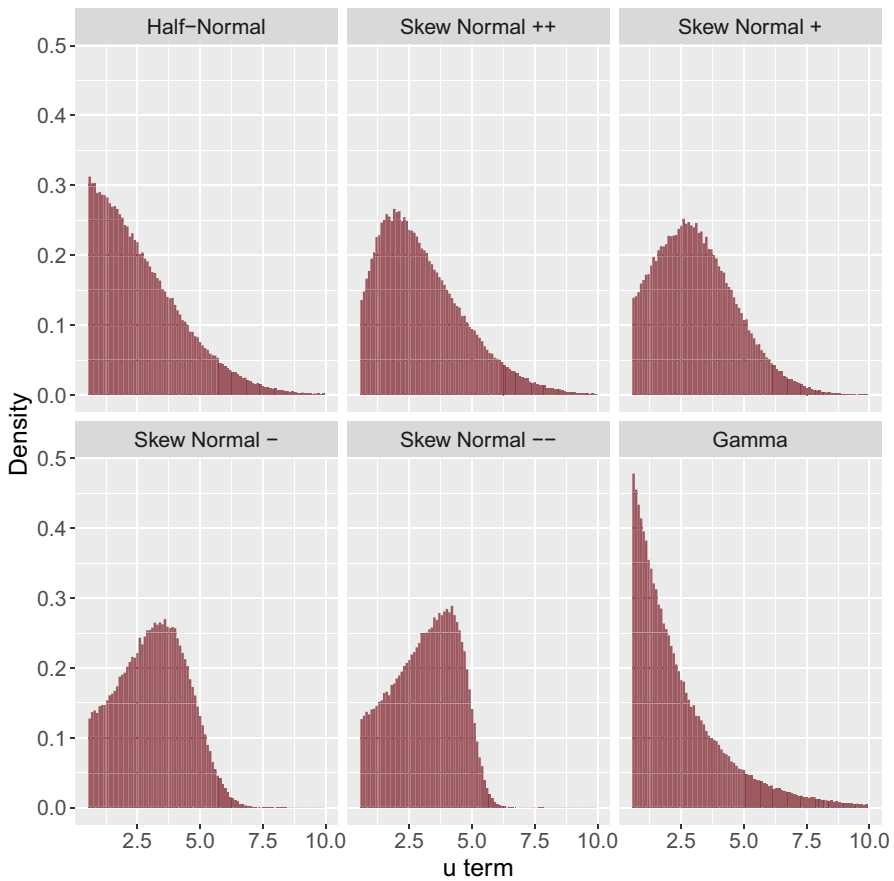
Fig. 7 Inefficiency *u* term distributions, 1000 replications

assumptions. The approach has been then illustrated using real data, already used in literature, and simulated experiments.

The results confirm that the proposed method offers good robustness properties and, in many cases, may be more efficient than the two main alternative estimation approaches, both robust like quantile regression and not robust like maximum likelihood.

As has already been verified in previous studies (Song et al. 2017; Wheat et al. 2019; Zulkarnain and Indahwati 2021), the latter are extremely compromised by anomalous data and often, if the efficiency distribution is different from that specified, the algorithms used for its optimization do not converge and fail in the search of a maximum (Meesters 2014).

On the contrary, quantile regression does not seem to suffer from similar problems, but its estimation capabilities are seriously compromised by a known and evident instability of the parameters relating to the higher quantiles which unfortunately are those necessary for stochastic frontier models. Our suggested method appears to be successful in simultaneously solving the drawbacks of its competitors. At the acceptable price

Table 4 MSE by method, distribution of inefficiency and outliers (1%, 3%, 5%)

MSE	Distribution	$\sigma_u = 3$ $(\lambda = 3)$			$\sigma_u = 1$ $(\lambda = 1)$		
		1%	3%	5%	1%	3%	5%
QRCM	Half Normal	0.50	1.57	2.35	0.87	1.82	2.69
SFA		0.66	2.56	4.54	0.84	1.35	1.76
QR		1.26	4.11	4.78	1.55	4.00	5.11
QRCM	Skew Normal ++	0.52	1.27	2.10	0.53	1.33	2.26
SFA		2.60	7.01	8.54	1.47	1.98	2.36
QR		0.86	3.62	4.47	1.08	3.75	4.90
QRCM	Skew Normal +	0.67	1.22	1.99	0.38	1.02	1.88
SFA		5.96	8.83	9.26	2.57	3.06	3.41
QR		0.60	2.81	3.95	0.70	3.30	4.54
QRCM	Skew Normal −	1.01	1.22	1.88	0.63	1.06	1.77
SFA		8.83	9.10	9.28	4.41	4.71	5.11
QR		0.66	2.14	3.42	0.58	2.98	4.42
QRCM	Skew Normal −	1.17	1.26	1.87	0.75	1.14	1.78
SFA		8.87	9.11	9.32	4.86	5.22	5.46
QR		0.71	1.86	3.19	0.63	2.90	4.47
QRCM	Gamma	0.89	2.09	3.04	0.37	1.02	1.90
SFA		0.48	0.87	1.33	2.12	4.25	4.87
QR		2.66	5.80	6.07	0.83	4.35	5.04

of a slight loss of estimation efficiency when the data are not contaminated by outliers and there is no doubt about the theoretical distribution of efficiencies, it has provided reliable estimates in any real and simulated case. It should also be emphasized the advantage of not necessarily requiring any preliminary tests to verify distributional hypotheses and regression diagnostics, nor the application of complex procedures for the automatic identification of outliers.

Finally, the recovery of the parametrization within the quantile regression approach gives greater flexibility to its practical use, allowing with extreme simplicity to impose constant parameters or frontiers that do not overlap as the quantile adopted increases. This flexibility could allow new and simpler developments also in the methodological field by introducing, for example, in these models some dependence parameters in time, space or network data. But this is left for further research, along with the possible extension of the proposed model to panel data, being currently only defined for cross-sectional data.

Code availability The software code that supports the findings of this study is available from the corresponding author, upon reasonable request.

Table 5 Mean absolute difference by method, distribution of inefficiency and outliers (1%, 3%, 5%)

Mean abs diff.	Distribution	$\sigma_u = 3$ ($\lambda = 3$)			$\sigma_u = 1$ ($\lambda = 1$)		
		1%	3%	5%	1%	3%	5%
QRCM	Half Normal	0.13	0.15	0.16	0.26	0.29	0.31
SFA		0.17	0.33	0.50	0.44	0.47	0.48
QR		0.14	0.17	0.18	0.30	0.36	0.37
QRCM	Skew Normal ++	0.11	0.11	0.11	0.19	0.20	0.22
SFA		0.27	0.64	0.76	0.59	0.61	0.61
QR		0.10	0.12	0.12	0.22	0.26	0.27
QRCM	Skew Normal +	0.12	0.11	0.12	0.17	0.16	0.16
SFA		0.56	0.80	0.82	0.72	0.73	0.73
QR		0.10	0.11	0.12	0.17	0.19	0.20
QRCM	Skew Normal −	0.14	0.11	0.12	0.19	0.16	0.15
SFA		0.84	0.84	0.83	0.82	0.82	0.81
QR		0.11	0.11	0.12	0.16	0.15	0.15
QRCM	Skew Normal −	0.15	0.12	0.12	0.20	0.17	0.15
SFA		0.85	0.84	0.84	0.84	0.83	0.82
QR		0.12	0.11	0.11	0.16	0.14	0.15
QRCM	Gamma	0.18	0.21	0.22	0.15	0.13	0.13
SFA		0.16	0.18	0.19	0.44	0.74	0.79
QR		0.22	0.25	0.25	0.14	0.15	0.16

Declarations

Conflict of interest The authors have declared that no competing interests exist; the authors have no relevant financial or non-financial interests to disclose.

Appendix A. Smooth functions

See Table 6.

Table 6 M-quantile smooth functions depending on the order of the quantile τ and R code

M-quantile smooth functions	R code
τ	`tau`
$\tau + \tau^2$	`tau+I(tau^2)`
$\tau + \tau^2 + \tau^3$	`tau+I(tau^2)+I(tau^3)`
$\tau + \tau^2 + \tau^3 + \tau^4$	`tau+I(tau^2)+I(tau^3)+I(tau^4)`
$\tau + \tau^2 + \tau^3 + \tau^4 + \tau^5$	`tau+I(tau^2)+I(tau^3)+I(tau^4)+I(tau^5)`
$\tau^{(1/2)} + (1 - \tau)^{(1/2)}$	`I(tau^(1/2))+I((1-tau)^(1/2))`
$\tau^{(1/3)} + (1 - \tau)^{(1/3)}$	`I(tau^(1/3))+I((1-tau)^(1/3))`
$\tau^{(1/4)} + (1 - \tau)^{(1/4)}$	`I(tau^(1/4))+I((1-tau)^(1/4))`
$\tau + cos(\pi * \tau) + sin(\pi * \tau)$	`tau+I(cos(pi*tau))+I(sin(pi*tau))`
$\tau + \log(\tau) + \log(1 - \tau)$	`tau+I(log(tau))+I(log(1-tau))`
Square root $(-\log(1 - \tau))$	`I(sqrt(-log(1-tau)))`
$-$ Square root $(-\log(\tau))$	`I(-sqrt(-log(tau)))`
Quantile function for the Normal distribution (τ)	`I(qnorm(tau))`
Quantile function for the Normal distribution $(\tau) + \log(\tau)$	`I(qnorm(tau))+I(log(tau))`
Quantile function for the Normal distribution $(\tau^2) + \log(\tau)$	`I(qnorm(tau^2))+I(log(tau))`
Quantile function for the Normal distribution $(\tau^3) + \log(\tau)$	`I(qnorm(tau^3))+I(log(tau))`
Quantile function for the Normal distribution $(\tau^4) + \log(\tau)$	`I(qnorm(tau^4))+I(log(tau))`
Quantile function for the Normal distribution $(\tau^5) + \log(\tau)$	`I(qnorm(tau^5))+I(log(tau))`
Quantile function for the Normal distribution $(\tau) + \log(1 - \tau)$	`I(qnorm(tau))+I(log(1-tau))`
Shifted Legendre polynomials (τ), degree 2	`slp(tau, k=2)`
Shifted Legendre polynomials (τ), degree 3	`slp(tau, k=3)`
Piecewise Linear Function (τ), knots [0.1,0.9]	`plf(tau, knots=c(0.1,0.9)`

Appendix B. Philippines rice farming

See Tables 7 and 8.

Table 7 Kolmogorov–Smirnov GOF test and correlation results—Philippines rice farming

M-quantile smooth functions	KS p value	QRCM versus SFA β correlations
τ	1.000	0.882
$\tau + \tau^2$	0.977	0.883
$\tau + \tau^2 + \tau^3$	1.000	0.803
$\tau + \tau^2 + \tau^3 + \tau^4$	1.000	0.751
$\tau + \tau^2 + \tau^3 + \tau^4 + \tau^5$	1.000	0.291
$\tau^{(1/2)} + (1 - \tau)^{(1/2)}$	0.328	0.692
$\tau^{(1/3)} + (1 - \tau)^{(1/3)}$	0.418	0.391
$\tau^{(1/4)} + (1 - \tau)^{(1/4)}$	0.873	0.266
$\tau + cos(\pi * \tau) + sin(\pi * \tau)$	1.000	−0.076
$\tau + \log(\tau) + \log(1 - \tau)$	0.977	0.924
Square root $(-\log(1 - \tau))$	0.012	0.880
− Square root $(-\log(\tau))$	0.772	0.955
Quantile function for the Normal distribution (τ)	0.975	0.963
Quantile function for the Normal distribution $(\tau) + \log(\tau)$	0.994	0.951
Quantile function for the Normal distribution $(\tau^2) + \log(\tau)$	0.994	0.985
Quantile function for the Normal distribution $(\tau^3) + \log(\tau)$	0.994	0.989
Quantile function for the Normal distribution $(\tau^4) + \log(\tau)$	0.991	0.985
Quantile function for the Normal distribution $(\tau^5) + \log(\tau)$	0.989	0.979
Quantile function for the Normal distribution $(\tau) + \log(1 - \tau)$	0.866	0.966
Shifted Legendre polynomials (τ), degree 2	0.977	0.883
Shifted Legendre polynomials (τ), degree 3	1.000	0.803
Piecewise Linear Function (τ), knots [0.1,0.9]	1.000	0.202

Table 8 Efficiency results, Philippine rice farming dataset, Spearman correlation

	COLS	SFA	QR	QRCM
COLS	1	0.989	0.954	0.931
SFA	0.989	1	0.980	0.969
QR	0.954	0.980	1	0.989
QRCM	0.931	0.969	0.989	1

Appendix C. Efficiency results—NBER manufacturing

See Table 9.

Table 9 Efficiency results, NBER manufacturing dataset, by method

Year	SFA Half-Normal	SFA Exponential	SFA T-Normal	QR	QRCM
1958	0.810	0.967	0.845	0.823	0.769
1959	0.820	0.971	0.855	0.843	0.808
1960	0.806	0.967	0.840	0.829	0.786
1961	0.812	0.970	0.846	0.821	0.791
1962	0.812	0.971	0.848	0.835	0.790
1963	0.808	0.970	0.847	0.813	0.780
1964	0.807	0.969	0.842	0.816	0.768
1965	0.815	0.974	0.861	0.806	0.837
1966	0.822	0.975	0.861	0.813	0.835
1967	0.814	0.972	0.851	0.816	0.780
1968	0.803	0.973	0.845	0.805	0.832
1969	0.789	0.969	0.803	0.816	0.804
1970	0.806	0.970	0.836	0.845	0.803
1971	0.809	0.972	0.843	0.830	0.814
1972	0.812	0.973	0.849	0.790	0.812
1973	0.808	0.974	0.810	0.796	0.841
1974	0.849	0.989	0.663	0.778	0.848
1975	0.809	0.971	0.836	0.817	0.834
1976	0.816	0.979	0.811	0.828	0.846
1977	0.820	0.977	0.807	0.794	0.839
1978	0.751	0.974	0.735	0.800	0.846
1979	0.825	0.979	0.865	0.809	0.833
1980	0.749	0.973	0.748	0.802	0.811
1981	0.794	0.972	0.824	0.803	0.802
1982	0.786	0.974	0.786	0.824	0.816
1983	0.794	0.973	0.793	0.815	0.834
1984	0.777	0.970	0.765	0.809	0.841
1985	0.796	0.979	0.790	0.797	0.850
1986	0.768	0.977	0.739	0.802	0.850
1987	0.818	0.977	0.834	0.825	0.837
1988	0.780	0.978	0.790	0.806	0.855
1989	0.827	0.981	0.873	0.824	0.830
1990	0.839	0.982	0.719	0.781	0.847
1991	0.771	0.978	0.757	0.810	0.859

Table 9 continued

Year	SFA Half-Normal	SFA Exponential	SFA T-Normal	QR	QRCM
1992	0.817	0.978	0.800	0.833	0.825
1993	0.824	0.989	0.723	0.772	0.833
1994	0.806	0.976	0.839	0.783	0.822
1995	0.813	0.976	0.845	0.796	0.788
1996	0.843	0.980	0.865	0.791	0.755
1997	0.859	0.983	0.886	0.855	0.842
1998	0.878	0.984	0.897	0.851	0.819
1999	0.998	0.988	0.994	0.752	0.819
2000	0.999	1	0.994	0.803	0.820
2001	0.846	0.987	0.683	0.789	0.837
2002	0.839	0.979	0.874	0.814	0.750
2003	0.912	0.983	0.902	0.781	0.791
2004	0.889	0.983	0.901	0.803	0.804
2005	0.999	1	0.999	0.797	0.774
2006	0.999	1	0.993	0.748	0.705
2007	0.999	1	0.998	0.798	0.734
2008	0.999	1	0.993	0.728	0.819
2009	0.999	1	0.996	0.738	0.758
2010	0.999	1	0.995	0.802	0.774
2011	0.999	1	0.992	0.768	0.862

Appendix D. Inefficiency distributions used for simulations

See Table 10.

Table 10 Inefficiency distributions and relative R code

Distribution	R code
	for $\sigma_u = 3$
Half Normal	`abs(rnorm(n=200, 0, 3))`

Skew Normal ++	`u = rsn(n=200, dp=cp2dp(cp = c(mean=10, s.d.= 6, gamma1=0.8), family="SN"));`
	`ifelse(u<=0, abs(u),u)`
	`u = ((u - min(u))/mean(u)) * 3`

Skew Normal +	`u = rsn(n=200, dp=cp2dp(cp = c(mean=10, s.d.= 6, gamma1=0.2), family="SN"));`
	`ifelse(u<=0, abs(u),u)`
	`u = ((u - min(u))/mean(u)) * 3`

Skew Normal −	`u = rsn(n=200, dp=cp2dp(cp = c(mean=10, s.d.= 6, gamma1=-0.65), family="SN"));`
	`ifelse(u<=0, abs(u),u)`
	`u = ((u - min(u))/mean(u)) * 3`

Skew Normal −	`u = rsn(n=200, dp=cp2dp(cp = c(mean=10, s.d.= 6, gamma1=-0.9), family="SN"));`
	`ifelse(u<=0, abs(u),u)`
	`u = ((u - min(u))/mean(u)) * 3`

Gamma	`rgamma(n=200, shape=1, scale = 2)`
	for $\sigma_u = 1$
Half Normal	`abs(rnorm(n=200, 0, 1))`

Skew Normal ++	`u = rsn(n=200, dp=cp2dp(cp = c(mean=10, s.d.= 2, gamma1=0.8), family="SN"));`
	`ifelse(u<=0, abs(u),u)`
	`u = ((u - min(u))/mean(u)) * 3`

Skew Normal +	`u = rsn(n=200, dp=cp2dp(cp = c(mean=10, s.d.= 2, gamma1=0.2), family="SN"));` `ifelse(u<=0, abs(u),u)`

Table 10 continued

Distribution	R code
Skew Normal –	```u = ((u - min(u))/mean(u)) * 3```
	```u = rsn(n=200, dp=cp2dp(cp = c(mean=10, s.d.= 2, gamma1=-0.65), family="SN"));```
	```ifelse(u<=0, abs(u),u)```
Skew Normal –	```u = ((u - min(u))/mean(u)) * 3```
	```u = rsn(n=200, dp=cp2dp(cp = c(mean=10, s.d.= 2, gamma1=-0.9), family="SN"));```
	```ifelse(u<=0, abs(u),u)```
	```u = ((u - min(u))/mean(u)) * 3```
Gamma	```rgamma(n=200, shape=1, scale = 1)+1```

References

Almanidis P, Sickles RC (2012) The skewness issue in stochastic frontiers models: Fact or fiction? In: Van Keilegom I, Wilson PW (eds) Exploring research Frontiers in contemporary statistics and econometrics: a Festschrift for Léopold Simar, pp 201–227. Physica-Verlag HD, Heidelberg. ISBN 978-3-7908-2349-3

Almanidis P, Qian J, Sickles RC (2014) Stochastic frontier models with bounded inefficiency. In: Sickles RC, Horrace WC (eds) Festschrift in honor of Peter Schmidt: econometric methods and applications, pp 47–81. Springer, New York. ISBN 978-1-4899-8008-3

Azzalini A, Valle AD (1996) The multivariate skew-normal distribution. Biometrika, 83(4):715–726. ISSN 00063444

Banker RD, Natarajan R (2008) Evaluating contextual variables affecting productivity using data envelopment analysis. Oper Res 56(1):48–58

Bartelsman E, Gray W (1996) The NBER manufacturing productivity database. NBER Technical Working Papers 0205, National Bureau of Economic Research, Inc

Battese GE, Coelli TJ (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empir Econ 20(2):325–332

Behr A (2010) Quantile regression for robust bank efficiency score estimation. Eur J Oper Res 200(2):568–581

Bernini C, Freo M, Gardini A (2004) Quantile estimation of frontier production function. Empir Econ 29(2):373–381

Bille' AG, Salvioni C, Benedetti R (2018) Modelling spatial regimes in farms technologies. J Product Anal 49(2–3):173–185

Bonanno D, De Giovanni G, Domma F (2017) The 'wrong skewness' problem: a re-specification of stochastic frontiers. J Prod Anal 47:49–64

Cai J, Feng Q, Horrace WC, Wu GL (2021) Wrong skewness and finite sample correction in the normal-half normal stochastic frontier model. Empir Econ

Carree MA (2002) Technological inefficiency and the skewness of the error component in stochastic frontier analysis. Econ Lett 77(1):101–107. ISSN 0165-1765

Chernozhukov V (2005) Extremal quantile regression. Ann Stat 33(2):806–839

Coelli T, Rao DSP, Battese GE (2005) An introduction to efficiency and productivity analysis, 2nd edn. Kluwer Academic Publishers, Norwell

Fan Y, Li Q, Weersink A (1996) Semiparametric estimation of stochastic production frontier models. J Bus Econ Stat 14:460–468

Frumento P (2021) Mqrcm: M-quantile regression coefficients modeling. https://CRAN.R-project.org/package=Mqrcm. R package version 1.2

Frumento P, Bottai M (2016) Parametric modeling of quantile regression coefficient functions. Biometrics 72(1):74–84

Frumento P, Bottai M (2017) Parametric modeling of quantile regression coefficient functions with censored and truncated data. Biometrics 73(4):1179–1188. ISSN 0006-341X

Furno M, Vistocco D (2018) Quantile regression: estimation and simulation, vol 216. Wiley

Fusco E, Vidoli F (2013) Spatial stochastic frontier models: controlling spatial global and local heterogeneity. Int Rev Appl Econ 27(5):679–694

Green A, Mayes D (1991) Technical inefficiency in manufacturing industries. Econ J 101(406):523–538

Greene W (2005) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. J Econom 126(2):269–303

Greene WH (2003) Simulated likelihood estimation of the normal-gamma stochastic frontier function. J Product Anal 19(2):179–190

Hafner CM, Manner H, Simar L (2018) The "wrong skewness" problem in stochastic frontier models: a new approach. Econom Rev 37(4):380–400

Horrace WC, Parmeter CF (2018) A Laplace stochastic frontier model. Econom Rev 37(3):260–280

Huang ML, Nguyen C (2017) High quantile regression for extreme events. J Stat Distrib Appl 4(1):4. ISSN 2195-5832

Huber PJ (1981) Robust statistics. Wiley, New York

Johnson A, Kuosmanen T (2011) One-stage estimation of the effects of operational conditions and practices on productive performance: asymptotically normal and efficient, root-n consistent Stonezd method. J Product Anal 36:219–230. ISSN 0895-562X

Jradi S, Ruggiero J (2019) Stochastic data envelopment analysis: a quantile regression approach to estimate the production frontier. Eur J Oper Res 278(2):385–393. ISSN 0377-2217

Jradi S, Parmeter CF, Ruggiero J (2019) Quantile estimation of the stochastic frontier model. Econ Lett 182:15–18. ISSN 0165-1765

Jradi S, Parmeter CF, Ruggiero J (2021) Quantile estimation of stochastic frontiers with the normal-exponential specification. Eur J Oper Res. ISSN 0377-2217

Knox K, Blankmeyer E, Stutzman J (2007) Technical efficiency in Texas nursing facilities: a stochastic production frontier approach. J Econ Finance 31(1):75–86

Koenker R (2005) Quantile regression. Cambridge Books, Cambridge University Press

Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46(1):33–50. ISSN 00129682, 14680262

Koenker R, Hallock KF (2001) Quantile regression. J Econ Perspect 15(4):143–156

Kumbhakar S, Parmeter CF, Zelenyuk V (2020) Stochastic frontier analysis: foundations and advances I. In: Subhash RC, Chambers R, Kumbhakar S (eds) Handbook of production economics, pp 1–39. Springer, Singapore. ISBN 978-981-10-3450-3

Kumbhakar SC, Lovell CAK (2004) Stochastic frontier analysis. Cambridge University Press, Cambridge

Kuosmanen T (2012) Stochastic semi-nonparametric frontier estimation of electricity distribution networks: application of the stoned method in the finnish regulatory model. Energy Econ 34(6):2189–2199

Kutlu L, Tran KC, Tsionas MG (2020) A spatial stochastic frontier model with endogenous frontier and environmental variables. Eur J Oper Res 286(1):389–399

Deyuan Li and Huixia Judy Wang (2019) Extreme quantile estimation for autoregressive models. J Bus Econ Stat 37(4):661–670

Li Q (1996) Estimating a stochastic production frontier when the adjusted error is symmetric. Econ Lett 52(3):221–228. ISSN 0165-1765

Liu C, Laporte A, Ferguson B (2008) The quantile regression approach to efficiency measurement: insights from Monte Carlo simulations. Health Econ 17:1073–1087

Meesters A (2014) A note on the assumed distributions in stochastic frontier models. J Product Anal 42(2):171–173. ISSN 15730441

Ondrich J, Ruggiero J (2001) Efficiency measurement in the stochastic frontier model. Eur J Oper Res 129(2):434–442. ISSN 0377-2217. A global view of industrial logistics

Papadopoulos A (2021) Stochastic frontier models using the generalized exponential distribution. J Product Anal 55(1):15–29

Parmeter CF, Wan ATK, Zhang X (2019) Model averaging estimators for the stochastic frontier model. J Prod Anal 51:91–103

Rho S, Schmidt P (2015) Are all firms inefficient? J Prod Anal 43:327–349

Roth J, Rajagopal R (2018) Benchmarking building energy efficiency using quantile regression. Energy 152:866–876

Simar L, Wilson PW (2009) Inferences from cross-sectional, stochastic frontier models. Econom Rev 29(1):62–98

Snedecor GW, Cochran WG (1989) Statistical methods, 8th edn. The Iowa State University Press

Song J, Oh DH, Kang J (2017) Robust estimation in stochastic frontier models. Comput Stat Data Anal 105:243–267

Sottile G, Frumento P, Chiodi M, Matteo B (2019) A penalized approach to covariate selection through quantile regression coefficient models. Stat Model 1471082X1982552. ISSN 1471-082X

Stephens MA (1974) EDF statistics for goodness of fit and some comparisons. J Am Stat Assoc 69(347):730–737

Tsionas EG (2007) Efficiency measurement with the Weibull stochastic frontier*. Oxf Bull Econ Stat 69(5):693–706

Tsionas EG, Michaelides PG (2016) A spatial stochastic frontier model with spillovers: evidence for Italian regions. Scott J Political Econ 63(3):243–257

Tsionas MG (2020) Quantile stochastic frontiers. Eur J Oper Res 282(3):1177–1184. ISSN 0377-2217

Tsionas MG, Assaf AG, Andrikopoulos A (2020) Quantile stochastic frontier models with endogeneity. Econ Lett 188:108964. ISSN 0165-1765

Waldmann E (2018) Quantile regression: a short story on how and why. Stat Model 18(3–4):203–218

Wang Y, Wang S, Dang C, Ge W (2014) Nonparametric quantile frontier estimation under shape restriction. Eur J Oper Res 232(3):671–678. ISSN 0377-2217

Wheat P, Stead AD, Greene WH (2019) Robust stochastic frontier analysis: a student's t-half normal model with application to highway maintenance costs in England. J Product Anal 51(1):21–38. ISSN 15730441

Winsten CB (1957) Discussion on Mr. Farrell's paper. J R Stat Soc 120:282–284

Zhang N, Huang X, Liu Y (2021) The cost of low-carbon transition for China's coal-fired power plants: a quantile frontier approach. Technol Forecast So Change 169:120809. ISSN 0040-1625

Zulkarnain R (2021) Robust stochastic frontier using Cauchy distribution for noise component to measure efficiency of rice farming in East Java. J Phys Conf Ser 1863(012031):1–10. ISSN 1742-6588