Book Section:
Stead, A, Wheat, P and Greene, WH (2023) Robustness in stochastic frontier analysis. In: Macedo, P, Moutinho, V and Madaleno, M, (eds.) Lecture Notes in Economics and Mathematical Systems. Lecture Notes in Economics and Mathematical Systems, 692 (692). Springer, Cham, Switzerland, pp. 197-228. ISBN: 978-3-031-29583-6.

https://doi.org/10.1007/978-3-031-29583-6_12

Robustness in Stochastic Frontier Analysis

Alexander D. Stead, Phill Wheat, and William H. Greene

Abstract A number of recent studies have addressed the issue of robustness in the context of stochastic frontier analysis, and alternative models and estimation methods have been proposed that appear more robust to outliers. For example, several models assuming heavy-tailed noise distributions appeared in the literature, including the logistic, Laplace, and Student's t distributions. Despite this, there has been little explicit discussion of the what is meant by 'robustness' and how models might be compared in terms of robustness to outliers. This chapter discusses two different aspects of robustness in stochastic frontier analysis: first, robustness of parameter estimates, by comparing the influence of outlying observations across different specifications – a familiar approach in the wider literature on robust estimation; second, the robustness of efficiency predictions to outliers across different specifications – a consideration unique to the efficiency analysis literature.

1 Introduction

Stochastic frontier (SF) analysis involves the estimation of an efficient frontier function – e.g. a production, cost, or revenue frontier. The robustness of our estimators and predictors is critical to accurate prediction of efficiency levels or rankings. The presence of contaminating outliers and other departures from distributional assumptions

Alexander D. Stead
Institute for Transport Studies, University of Leeds, 34-40 University Rd, Leeds, LS2 9JT, UK
e-mail: a.d.stead@leeds.ac.uk

Phill Wheat
Institute for Transport Studies, University of Leeds, 34-40 University Rd, Leeds, LS2 9LJ, UK
e-mail: p.e.wheat@its.leeds.ac.uk

William H. Greene
Department of Economics, Stern School of Business, New York University, 44 West 4th Street, New York, NY 10012, USA e-mail: wgreene@stern.nyu.edu

is problematic, since maximum likelihood estimation (MLE) is the most commonly employed estimation method in the SF literature and it is well known that MLE and other classical estimation methods are usually non-robust, in that they perform poorly in the presence of departures from distributional assumptions.

In SF modelling, robustness has additional relevance since a primary concern, especially in a regulatory context, is the prediction of efficiency levels of individual firms against the estimated frontier function. We therefore have reason to be concerned not only with the robustness of our estimation of the frontier, but the robustness of our efficiency predictions.

Despite this relevance, relatively little attention has been given to the issue of robustness in the SF literature. In recent years, several studies have proposed alternative models or estimators relevant to the discussion of robustness, though with little explicit discussion of what is meant by 'robustness' and how it might be measured and compared across specifications and estimation methods. This chapter discusses two different aspects of robustness in stochastic frontier analysis: first, robustness of parameter estimates, by comparing the influence of outlying observations across different specifications – a familiar approach in the wider literature on robust estimation; second, the robustness of efficiency predictions to outliers across different specifications – a consideration unique to the efficiency analysis literature.

The remainder of this chapter is structured as follows. Sections 2 and 3 briefly introduce the SF model and some background and key concepts from the robustness literature, respectively. Section 4 discusses robust estimation of the SF model, drawing on the concepts introduced in Section 3 and relevant SF literature. Section 5 discusses the robustness of efficiency prediction. Section 6 summarises and concludes.

2 The Stochastic Frontier Model

Introduced by Aigner et al. (1977) and Meeusen and van Den Broeck (1977), the basic cross-sectional stochastic frontier model is given by

$$y_i = \boldsymbol{x_i'}\boldsymbol{\beta} + \varepsilon_i, \qquad \varepsilon_i = v_i - su_i, \tag{1}$$

where for the i^{th} firm, y_i is the dependent variable, $\boldsymbol{x_i}$ is a vector of independent variables, $\boldsymbol{\beta}$ is a vector of frontier coefficients, and ε_i is an error term. The latter is composed of a two-sided term, v_i, representing random measurement error and other noise factors, and a non-negative term, u_i, representing departure from the frontier function as a result of inefficiency. In the case of a production frontier, firms may only be on or below the frontier, in which case $s = 1$. On the other hand, in the case of a cost frontier, firms may only be on or above the frontier, in which case $s = -1$. However, note that the composed error ε_i may take on either sign owing to the presence of the two-sided noise term v_i. Many extensions of this model, particularly to panel data settings, have been proposed; see reviews of the literature by Kumbhakar and Lovell (2000), Murillo-Zamorano (2004), Coelli et al. (2005), Greene (2008), Parmeter and Kumbhakar (2014), and Stead et al. (2019).

Though alternative parametric and semiparametric approaches to estimation of the model have been explored, some of which are of particular interest from a robustness perspective and discussed further in Sect. 4, MLE is the most common approach in theoretical and applied SF literature. This necessitates specific distributional assumptions regarding the error terms v_i and u_i; Aigner et al. (1977) and Meeusen and van Den Broeck (1977) explored estimation of the model under the assumption that

$$v_i \sim \mathcal{N}(0, \sigma_v^2), \qquad u_i = |w_i|, \qquad w_i \sim \mathcal{N}(0, \sigma_u^2),$$

known as the normal-half normal (N-HN) model, or alternatively that

$$v_i \sim \mathcal{N}(0, \sigma_v^2), \qquad u_i \sim \text{Exponential}(1/\sigma_u),$$

known as the normal-exponential (N-EXP) model. In both cases it is assumed that v_i and u_i are independent. The marginal density of the composed error, ε_i, is then derived via the convolution

$$f_\varepsilon(\varepsilon_i, \boldsymbol{\theta}) = \int_0^\infty f_{v,u}(\varepsilon_i + su_i, u_i, \boldsymbol{\theta}) \mathrm{d}u_i, \qquad \boldsymbol{\theta} = \left(\boldsymbol{\beta}', \boldsymbol{\vartheta}'\right)', \tag{2}$$

where $f_{v,u}$ is the joint density of v_i and u_i–which under the assumption of independence, is just the product of their marginal densities–and $\boldsymbol{\theta}$ is a vector of parameters. This is then used to form the log-likelihood function.

The next step in SF modelling is then to predict observation-specific efficiency, u_i, relative to the estimated frontier. These predictions are based on the conditional distribution

$$f_{u|\varepsilon}(u_i|\varepsilon_i, \boldsymbol{\theta}) = \frac{f_{v,u}(\varepsilon_i + su_i, u_i, \boldsymbol{\theta})}{f_\varepsilon(\varepsilon_i, \boldsymbol{\theta})}, \tag{3}$$

following Jondrow et al. (1982). Both the log-likelihood function and $f_{u|\varepsilon}$ clearly depend on our underlying assumptions about the distribution of v_i and u_i. Many generalisations of and departures from the N-HN and N-EXP cases have been proposed, and a detailed discussion of these is beyond the scope of this chapter, but is available in the aforementioned reviews of the SF literature. From a robustness perspective, the essential point is that distributional assumptions about v_i and u_i are critical to both estimation of the frontier function and the prediction of efficiency relative to it.

3 Definitions and Measures of Robustness

Robustness, broadly defined, means a lack of sensitivity to small departures from our model assumptions. Ideally, we would like a robust estimator, i.e. one that is not unduly sensitive such departures. The main concern in many applications is robustness to ‘outliers’, i.e. outlying observations drawn from some contaminating distribution. Robustness to outliers will also be the main focus of this chapter, though given the critical role of specific distributional assumptions in SF modelling, it is worth keeping in mind a broader definition, especially in relation to issues from the SF literature such as ‘wrong skew’. Two related but distinct concepts in the literature on robust estimation are *robustness*, measured in terms of the *influence* of contaminating observations on an estimator, and *resistance*, measured in terms of the estimator’s *breakdown point*. These are discussed below.

3.1 The Influence Function

Analysis of robustness often centres around the influence of contaminating observations on a statistical functional. Introduced by Hampel et al. (1986), the *influence function* measures the effect on a statistical functional T of an infinitesimal perturbation of the distribution of the data F at a point $(y, \boldsymbol{x})$, and is given by

$$L_T(y, \boldsymbol{x}) = \lim_{\epsilon \to 0} \frac{T\big((1-\epsilon)F + \epsilon\delta_y\big) - T(F)}{\epsilon}, \tag{4}$$

where δ_y is a point mass at y. A statistical functional, in turn, is any function that maps a distribution to a real scalar or vector. The greater the magnitude of L_T, the greater the influence of the contaminating observation. The usefulness of this concept is clear, since it can be applied to many different kinds of estimators and predictors, enabling the discussion of the robustness of SF estimation and efficiency prediction in terms of a consistent set of concepts and measures.

The influence function can be derived as a limiting case of the *Gâteaux derivative*, a generalisation of the directional derivative to differentiation in vector spaces. In the current context, we could generalise the influence function given by Eq. 4 to the Gâteaux derivative

$$\mathrm{d}T(F, G) = \lim_{\epsilon \to 0} \frac{T\big((1-\epsilon)F + \epsilon G\big) - T(F)}{\epsilon},$$

where G is potentially any contaminating distribution. This potentially offers a useful way of analysing the influence of other kinds of departures from our distributional assumptions, though such a discussion is beyond the scope of this chapter. However, we will exploit a useful property of the influence function which it owes to this relationship. As with ordinary derivatives, a chain rule exists for Gâteaux derivatives, and therefore for influence functions. If we have a statistical functional which can be expressed in terms of J functionals such that

$$T(F) = T\big(T_1(F), \ldots, T_J(F)\big),$$

then the overall influence on the functional $T(F)$ of an infinitesimal peturbation of the data F at y is given by

$$L_T(y, \boldsymbol{x}) = \sum_{j=1}^{J} \frac{\partial T(F)}{\partial T_j(F)} L_j(y, \boldsymbol{x}), \tag{5}$$

where $T_j(F)$ is the jth functional, and L_j is its corresponding influence function.

Some key measures derived from the influence function include the *rejection point*, the *gross-error sensitivity*, and the *local-shift sensitivity*, discussed below. We also discuss the related concept of *leverage*.

3.1.1 Gross-error sensitivity

A key measure of the robustness of a functional is the *gross-error sensitivity*

$$\gamma_T^*(\boldsymbol{x}) = \inf_y \left| L_T(y, \boldsymbol{x}) \right| \tag{6}$$

which is the infimum of the magnitude of the influence function over all points for which the influence function exists. If the gross-error sensitivity is finite, that is the influence is bounded, we say that a functional is *bias-robust* or simply robust; the lower the gross-error sensitivity, the more robust the model is to contaminating outliers.

3.1.2 Local-shift sensitivity

An alternative metric, the *local-shift sensitivity*

$$\lambda_T^*(\boldsymbol{x}) = \inf_{y \neq z} \frac{\left| L_T(y, \boldsymbol{x}) - L_T(z, \boldsymbol{x}) \right|}{|y - z|}, \tag{7}$$

where the infimum is taken over all y, z for which $y \neq z$ and the influence functions exist. The local-shift sensitivity measures the sensitivity of the estimator to a *small* change in observed values.

3.1.3 Rejection point

Another measure of interest is the *rejection point*

$$\rho_T^*(\boldsymbol{x}) = \inf_{r > 0} \left\{ r : L_T(y, \boldsymbol{x}) = 0, |y| > r \right\}, \tag{8}$$

which tells us how large an outlier must become before the influence function becomes zero. If the influence of gross outliers becomes zero, the effect is the same as removing them from the sample, hence the name *rejection point*.

3.1.4 Leverage

In regression-type settings which involve linear functions of vectors of covariates, such as our SF model described by Eq. 1, we are interested in the predicted values of the dependent variable for some $\boldsymbol{x_i}$, given by $\hat{y} = \boldsymbol{x_i'}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is our estimator for $\boldsymbol{\beta}$. Predicted values are, of course, of particular interest in SF modelling, given our interest in estimating the distance of the firm from the frontier. The sensitivity of $\hat{y}_i$ to contaminating observations is therefore of particular interest. Given that we can think of $\hat{y}_i$ as a function of functionals as in Eq. 5, it has an influence function. Applying the influence function chain rule as given in Eq. 4, this is simply

$$L_{\hat{y}_i}(y, \boldsymbol{x}) = \boldsymbol{x_i'}\boldsymbol{L}_{\hat{\boldsymbol{\beta}}}(y, \boldsymbol{x}), \tag{9}$$

where $\boldsymbol{L}_{\hat{\boldsymbol{\beta}}}$ is the influence function for $\hat{\boldsymbol{\beta}}$. It is of natural interest here to consider the evaluation of $L_{\hat{y}_i}$ at $(y_i, \boldsymbol{x_i})$,

$$L_{\hat{y}_i}(y_i, \boldsymbol{x_i}) = \boldsymbol{x_i'}\boldsymbol{L}_{\hat{\boldsymbol{\beta}}}(y_i, \boldsymbol{x_i}), \tag{10}$$

that is the influence on the predicted value for a observation i of a peturbation of the data at $(y_i, \boldsymbol{x_i})$. This is known as the self-influence or *leverage* of the observation, and depends fundamentally on the on the values of the independent variables. The direct relationship between the leverage, the influence of an observation on $\hat{\boldsymbol{\beta}}$, and the values of the covariates is apparent from Eq. 10.

3.2 Breakdown point

Aside from influence-based measures and definitions of robustness, there is a related but distinct concept of the *breakdown point*. The concept of the breakdown point, introduced by Hampel (1971; 1968), gives the smallest fraction of sampled observations that would be needed to make an estimator take on an arbitrary value. There is a distinction between the *replacement breakdown point* (the minimum fraction of existing observations one would need to contaminate) and the *addition breakdown point* (the fraction of additional contaminating observations one would need to add to an existing sample). Both asymptotic and finite sample definitions exist – see Donoho and Huber (1983). An estimator with a high breakdown point is said to be *resistant*.

The concept of the breakdown point is easily illustrated by contrasting the sample mean and sample median; the former has a breakdown point of $1/n$ in small samples and zero asymptotically, while the latter has a breakdown point of $1/2$. The sample mean and median therefore represent opposite extremes in terms of breakdown points and resistance to contaminating observations.

In terms of the asymptotic breakdown points of functionals, the asymptotic addition breakdown point is, from Huber (1981)

$$\epsilon_T^* = \inf\left\{\epsilon > 0 : \sup_G \left| T\big((1-\epsilon)F + \epsilon G\big) - T(F) \right| = \infty \right\}, \tag{11}$$

where the supremum is taken over all possible contaminating distributions.

3.3 Summary of measures

In this section, we have introduced several related concepts from the literature on robustness. With the exception of the breakdown point these measures relate to the influence of contaminating observations, which can be evaluated using the influence function describing the relationship between the influence of an observation and its value. The gross-error sensitivity is defined as the supremum of the magnitude of the influence function, while the local-shift sensitivity reflects its slope, and the rejection point is how large $|y|$ must be for influence to become zero.

The breakdown point offers an alternative approach. In contrast to the influence function, we are not aware of convenient formulae for the breakdown point of general classes of functionals, and therefore most of the discussion hereafter will focus on influence-based measures. The influence function also offers a more natural way to extend the discussion to the robustness of post-estimation predictions. It will, however, be useful to discuss breakdown points in the context of, e.g. conditional quantile regression estimation approaches that have been proposed in the SF literature.

4 Robustness and stochastic frontier estimation

In this section, we move on to discuss SF estimation in light of the concepts from the robustness literature introduced in the previous Sect 3. MLE is the workhorse of the SF literature, in terms of both the attention it has received in theoretical literature, and its useage in empirical applications. Other estimation methods we consider are the corrected ordinary least squares (COLS), and quantile regression (QR) approaches, and some generalisations of MLE. We discuss the robustness properties of each of these estimators in the context of SF modelling, and some of the approaches that have been taken to deal with outliers, including both alternative approaches to estimation and alteration of distributional assumptions.

This discussion is aided by the fact that all of the estimation methods under consideration belong to a broader class of *M-estimators*. Following Huber (1964, 1981), an M-estimator is any that can be defined as the solution to the problem

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \rho(y_i, \boldsymbol{x_i}, \boldsymbol{\theta}). \tag{12}$$

where n is the number of observations and ρ is a loss function. If the loss function has a derivative with respect to $\boldsymbol{\theta}$, denoted $\boldsymbol{\psi}$, we can define the M-estimator as the solution to the equation

$$\sum_{i=1}^{n} \boldsymbol{\psi}(y_i, \boldsymbol{x_i}, \hat{\boldsymbol{\theta}}) = \mathbf{0}, \qquad \boldsymbol{\psi}(y_i, \boldsymbol{x_i}, \hat{\boldsymbol{\theta}}) = \left.\frac{\partial \rho(y_i, \boldsymbol{x_i}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{13}$$

where $\mathbf{0}$ denotes a column vectors of zeros of the same length as $\boldsymbol{\theta}$. From Huber (1981), the influence function for an M-estimator is given by

$$L_{\hat{\boldsymbol{\theta}}}(y, \boldsymbol{x}) = -\left(\mathbb{E}\left(\left.\frac{\partial \boldsymbol{\psi}(y, \boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)\right)^{-1} \boldsymbol{\psi}(y, \boldsymbol{x}, \hat{\boldsymbol{\theta}}). \tag{14}$$

This class of M-estimators encompasses many classical estimators, including MLE and ordinary least squares (OLS), two methods commonly employed in the SF literature. The robustness properties of M-estimators have been studied extensively, and several robust M-estimators have been proposed in the literature on robust estimation, though application of these in the SF literature has so far been limited.

4.1 Corrected ordinary least squares

Where we have a stochastic frontier model of the form introduced in Sect. 2, the corrected ordinary least squares (COLS) estimation method proceeds by noting that, assuming independence of v_i and u_i, we may re-write Eq. 1 as

$$y_i = \alpha^* + \boldsymbol{x_i^{*\prime}}\boldsymbol{\beta^*} + \varepsilon_i^*, \qquad \varepsilon_i^* = v - su_i^*, \tag{15}$$
$$\alpha^* = \alpha - s\mathbb{E}(u_i), \quad u_i^* = u_i - \mathbb{E}(u_i), \quad \boldsymbol{\beta} = (\alpha, \boldsymbol{\beta^{*\prime}}), \quad \boldsymbol{x_i} = (1, \boldsymbol{x_i^{*\prime}})',$$

where $\mathbb{E}(\varepsilon_i^*) = 0$, and that therefore OLS may be used to obtain unbiased estimates of α^* and $\boldsymbol{\beta^*}$. That is, OLS yields unbiased estimates of all of the frontier parameters apart from the intercept, which is biased downward (upward) in a production (cost) frontier model by $\mathbb{E}(u_i)$. Parameters of the distributions of v_i and u_i are then obtained based on the sample moments of the distribution of the OLS residuals, given a set of distributional assumptions. The moment-based estimator of $\mathbb{E}(u_i)$ is then used to correct our estimated intercept, hence the name *corrected ordinary least squares*[1]. Solutions have been derived for several different models – see Aigner et al. (1977) and Olson et al. (1980) for the N-HN model, Greene (1980) for the N-EXP model, and Greene (1990) for the normal-gamma (N-G) model.

It is straightforward, however, to show that COLS is non-robust. If we partition the parameter vector such that

[1] Not to be confused with the *modified ordinary least squares* (MOLS) approach to estimating a deterministic frontier function. The two terms are often used interchangeably; for an extensive discussion, see Parmeter (2021).

$$\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^{*\prime}, \boldsymbol{\vartheta}')',$$

where $\boldsymbol{\vartheta}$ is the vector of parameters of the distribution of the composed error, the COLS estimator $\tilde{\boldsymbol{\theta}}$ may expressed as a function of the OLS estimator

$$\tilde{\boldsymbol{\theta}} = \left(\hat{\alpha}^* + sg\big(\boldsymbol{\mu}(\hat{\boldsymbol{\theta}})\big), \hat{\boldsymbol{\beta}}^{*\prime}, \boldsymbol{h}\big(\boldsymbol{\mu}(\hat{\boldsymbol{\theta}})'\big)\right)', \qquad \hat{\boldsymbol{\theta}} = (\hat{\alpha}^*, \hat{\boldsymbol{\beta}}^{*\prime}), \tag{16}$$

where $\hat{\boldsymbol{\theta}}$ is the OLS estimator, $\boldsymbol{\mu}(\hat{\boldsymbol{\theta}})$ is a vector of moments of the distribution of estimated OLS residuals, and g and $\boldsymbol{h}$ are functions of the latter yielding our moment-based estimators of $\mathbb{E}(u_i)$ and $\boldsymbol{\vartheta}$, respectively. Applying the influence function chain rule gives

$$L_{\tilde{\boldsymbol{\theta}}}(y, \boldsymbol{x}) = L_{\hat{\boldsymbol{\theta}}}(y, \boldsymbol{x})' \frac{\partial \tilde{\boldsymbol{\theta}}}{\partial \hat{\boldsymbol{\theta}}'}, \tag{17}$$

from which it is clear that the robustness properties of COLS follow directly from those of OLS. The non-robustness of OLS is well-known, but one convenient way to demonstrate this is by recognising that OLS is an M-estimator where

$$\rho(y_i, \boldsymbol{x_i}, \boldsymbol{\theta}) = (y_i - \boldsymbol{x_i'\theta})^2,$$

$$\boldsymbol{\psi}(y, \boldsymbol{x}, \hat{\boldsymbol{\theta}}) = -2(y - \boldsymbol{x}'\hat{\boldsymbol{\theta}})\boldsymbol{x}, \qquad \left.\frac{\partial \boldsymbol{\psi}(y, \boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 2\boldsymbol{x}\boldsymbol{x}',$$

and substituting these expressions in to Eq. 14 leads to a simple formula for influence

$$L_{\hat{\boldsymbol{\theta}}}(y, \boldsymbol{x}) = (y - \boldsymbol{x}'\hat{\boldsymbol{\theta}})\big(\mathbb{E}(\boldsymbol{x}\boldsymbol{x}')\big)^{-1}\boldsymbol{x}, \tag{18}$$

which is clearly unbounded. The gross-error sensitivity of OLS is infinite, and it has the lowest possible breakdown point of $1/n$. Substituting Eq. 18 in to Eq. 17, we then see that COLS inherits these properties and is likewise non-robust.

This is unfortunate for two main reasons. First, COLS is otherwise an attractive estimator in its own right because of its simplicity and ease of implementation, and is possibly second only to MLE in terms of its usage and coverage in the SF literature. In addition, evidence from Monte Carlo experiments suggests that COLS may perform well relative to MLE in small samples (Olson et al., 1980). Second, COLS is often used as a means of obtaining starting values for iterative optimisation algorithms such as those used for MLE. The sensitivity of these starting values to contaminating outliers may be problematic given that in SF modelling, the log-likelihood is not always well-behaved. Similarly, obtaining good starting values can become particularly important in the context of certain robust M-estimators, for which multiple local optima may exist.

Robust alternatives to COLS

The preceding discussion leads us to consider potential robust alternative to COLS. In principle we could estimate α^* and $\boldsymbol{\beta^*}$ in Eq. 15 using any linear unbiased estimator and, redefining $\hat{\boldsymbol{\theta}}$ as an alternative robust estimator, substitute in to Eq. 16 for the corresponding 'corrected' estimator for the SF model. To give a concrete example, we could use the least absolute deviations (LAD) estimator to obtain robust estimates of α^* and $\boldsymbol{\beta^*}$, and then estimate $\mathbb{E}(u_i)$ and remaining parameters of the marginal distribution of ε_i based on moments of the LAD residuals, and making the necessary correction to the intercept. By analogy with COLS, we might name this the CLAD estimator.

LAD is a well-known estimator with a long history, predating even OLS; for an extensive background and discussion see Dielman (2005). LAD is an M-estimator where

$$\rho(y_i, \boldsymbol{x_i}, \boldsymbol{\theta}) = |y_i - \boldsymbol{x_i'}\boldsymbol{\theta}|, \qquad \boldsymbol{\psi}(y, \boldsymbol{x}, \hat{\boldsymbol{\theta}}) = -\text{sgn}(y - \boldsymbol{x'}\hat{\boldsymbol{\theta}})\boldsymbol{x},$$

however are not able to derive the influence function via Eq. 14 owing to the singularity of the Hessian[2]. Following Koenker (2005), the influence function for the class of quantile regression estimators to which LAD belongs is given by

$$L_{\hat{\boldsymbol{\theta}}}(y, \boldsymbol{x}) = \boldsymbol{Q}^{-1}\text{sgn}(y - \boldsymbol{x'}\hat{\boldsymbol{\theta}})\boldsymbol{x}, \tag{19}$$

$$\boldsymbol{Q} = \int \boldsymbol{x}\boldsymbol{x'} f(\boldsymbol{x'}\hat{\boldsymbol{\theta}})\mathrm{d}G(\boldsymbol{x}), \qquad \mathrm{d}F = \mathrm{d}G(\boldsymbol{x})f(y|\boldsymbol{x}),$$

and by comparing this to the OLS influence function, we can see that LAD gives less weight to outlying observations. Regarding the resistance of of the estimator, a distinction needs to be made between the *finite sample breakdown point* or *conditional breakdown point* (Donoho and Huber, 1983) and the ordinary breakdown point. The conditional breakdown point considers contamination in y only, taking $\boldsymbol{x}$ as fixed, while the ordinary breakdown point considers contamination with respect to both y and $\boldsymbol{x}$. It has been shown that the conditional breakdown point of the LAD estimator can be greater than $1/n$ (He et al., 1990; Mizera and Müller, 2001; Giloni and Padberg, 2004). On the other hand, the ordinary breakdown point of LAD is $1/n$. Giloni et al. (2006) propose a weighted LAD estimator with a high breakdown point.

In addition to LAD, there are many other alternatives to OLS notable for their outlier robustness or resistance, which we could consider as the first stage in some corrected regression approach. A comprehensive discussion of robust and resistant regression estimators is beyond the scope of this chapter, but some prominent examples include M-estimation approaches, least median of squares (LMS) (Siegel, 1982), and least trimmed squares (LTS) (Rousseeuw, 1984). Early resistant regression techniques such as LMS and LTS have high breakdown points, but low efficiency. More

[2] Another consequence of the singularity of the Hessian is that the usual asymptotic results cannot be applied. Bassett and Koenker (1978) show that LAD is asymptotically normal. A simpler proof is given by Pollard (1991).

recently, techniques have been developed with higher efficiency under the assumption of normality, though computational issues can be significant – for a review, see Yu and Yao (2017). Seaver and Triantis (1995) discuss the sensitivity of COLS to outliers and compare production function estimates under OLS, LTS, LMS, and weighted least squares (WLS).

Of particular relevance in the context of SF modelling, small sample results from Lind et al (1992) suggest that LAD significantly outperforms OLS in terms of bias and mean squared error when the error distribution is asymmetric. This suggests that CLAD or and corrected robust or resistant regression estimators could offer significant improvements over COLS.

4.2 Quantile regression

An alternative estimation approach that has gained some attention in the SF literature in recent years is quantile regression (QR). Introduced by Koenker and Bassett (1978), the QR estimator is an M-estimator where the loss function corresponding to the τ^{th} conditional quantile is given by

$$\rho(y_i, \boldsymbol{x_i}, \boldsymbol{\theta}) = \rho_\tau(y_i - \boldsymbol{x_i'\theta}), \qquad \rho_\tau(\varepsilon_i) = \varepsilon_i\left(\tau - \mathbb{I}_{\varepsilon_i<0}\right), \tag{20}$$

where $\mathbb{I}$ denotes the indicator function. Note that when $\tau = 0.5$, we have the conditional median or LAD estimator discussed in Sect. 4.1. Rather than estimating a conditional mean function as in OLS, we are estimating a conditional median or other quantile. The intuitive appeal of QR is clear, since in SF modelling we expect most of the observations to lie below (above) the estimated production (cost) function. Moreover, for an appropriate choice of τ, the estimator will be unbiased, even for the intercept. To see this, note that – assuming that ε_i is independent of $\boldsymbol{x_i}$ – the τ^{th} conditional quantile of y_i is given by

$$Q(\tau|\boldsymbol{x_i}) = \boldsymbol{x_i'\beta} + F_\varepsilon^{-1}(\tau),$$

where F_ε^{-1} denotes the quantile function of composed error distribution. The QR estimator of the intercept will then be biased by $F_\varepsilon^{-1}(\tau)$, so the problem is to choose some optimal τ^* such that

$$F_\varepsilon^{-1}(\tau^*) = 0,$$

which depends on specific distributional assumptions. Early applications of quantile regression to frontier estimation chose τ arbitrarily, e.g. Bernini et al. (2004), Knox et al. (2007), Liu et al. (2008), and Behr (2010). More recently, Jradi and Ruggiero (2019) noted that a formula for τ^* can be derived by evaluating F_ε, the distribution function for the composed error, at zero, showing that for the N-HN model this gives

$$\tau^* = \frac{1}{2} + s\frac{1}{\pi}\arctan\left(\frac{\sigma_u}{\sigma_v}\right),$$

and Jradi et al. (2021) show that for the N-EXP model

$$\tau^* = \frac{1}{2} + s \exp\left(\frac{1}{2}\left(\frac{\sigma_v}{\sigma_u}\right)^2\right)\Phi\left(-\frac{\sigma_u}{\sigma_v}\right).$$

Jradi and Ruggiero (2019) suggest comparing likelihood values for different values of τ, while Jradi et al. (2019) choose the τ that minimises $\tau - \tau^*$. A simpler approach approach based on evaluating F_ε at the expected value of the OLS residuals is proposed by Jradi et al. (2021) and applied to the N-EXP model; Zhao (2021) implement this approach in the N-HN case. Bayesian approaches to QR estimation of the SF model are explored by Tsionas (2020) and Tsionas et al. (2020). For a recent in-depth review and discussion of the application of QR to estimate SF models, see Papadopoulos and Parmeter (2022).

Robustness properties of quantile regression

The perceived robustness of QR estimation relative to MLE has been cited as a key motivation for the its use in a SF context. Papadopoulos and Parmeter (2022) argue that, under usual assumptions about independence of the errors from the regressors, the appeal of QR is diminished and "the one remaining advantage of quantile regression is its robustness to outliers". The robustness properties of QR have not been discussed in detail in the SF literature, and warrant further consideration. Of particular relevance is the fact that the choice of τ, which has been a focus of the previous discussion, has a direct bearing on the resistance of the estimator to outliers.

In a linear regression model, at least k points will lie exactly on the estimated regression plane under QR, where k is the number of regressors; in other words, $\hat{\varepsilon}_i = 0$ for at least k observations. If we can correctly guess the sign of the residuals, we can run the regression excluding all the observations above and below the estimated regression plane and obtain exactly the same results. This property has been exploited in order to design faster algorithms for QR estimation; see Portnoy and Koenker (1997) and Chernozhukov et al. (2022). But it is also helpful in understanding intuitively the effect of contaminating outliers on the estimator. When $\tau = 0.5$, contaminating outliers are likely to be points above or below the estimated QR plane; yet as we move to more extreme quantiles, it becomes increasingly likely that they will be included in the k observations interpolated by the plane. This presents a potential problem in the context of SF estimation where the quantile τ^* could plausibly be rather extreme, the problem becoming more acute when k is large relative to n.

The influence function and breakdown points of QR have been examined in some detail. As noted previously, the LAD estimator is a special case of QR that arises when $\tau = 0.5$. As such, much of our previous discussion of the properties of the LAD estimator in Sect. 4.1 applies to QR in general. The influence function is as given in Eq. 19, meaning that the influence of an contaminating observation depends only on its sign and $\boldsymbol{x}_i$. In that sense, in line with the previous discussion, QR is robust provided we don't venture too far into the tails of the distribution.

The ordinary breakdown point of QR is $1/n$ and thus no better than that of OLS, while the conditional breakdown point can be larger than $1/n$. This is easily understood in terms of the preceding discussion and the influence function, since we can potentially alter y_i for $n-k$ observations without influencing the estimator, so long as the residuals don't change sign, while even for a single observation, we could choose y_i and $\boldsymbol{x_i}$ such that the QR plane is forced to intersect it. Accordingly He et al. (1990) show that the conditional breakdown point of the QR estimator has an upper bound which decreases with τ.

To summarise, QR is a potentially robust estimation method, and may also have a conditional breakdown point over $1/n$. On the other hand, its resistance to outliers is reduced for more extreme quantiles, potentially lessening its appeal in the context of SF modelling. Further examination of this issue is needed to assess the appeal of QR as a robust estimator of the SF model.

4.3 Maximum likelihood estimation

Beginning with Aigner et al. (1977) and Meeusen and van Den Broeck (1977), MLE has been the most commonly used estimation method in the SF literature. As mentioned previously, as with COLS and QR estimation, MLE belongs to the class of M-estimators. In the case of MLE

$$\begin{aligned}\rho(y_i, \boldsymbol{x_i}, \boldsymbol{\theta}) &= -\ln f_\varepsilon(y_i - \boldsymbol{x_i'\beta}, \boldsymbol{\theta}),\\ \psi(y, \boldsymbol{x}, \hat{\boldsymbol{\theta}}) &= -\left.\frac{\partial \ln f_\varepsilon(y - \boldsymbol{x'\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},\end{aligned}$$

where f_ε denotes the density of the composed error. Substituting these expressions into Eq. 14, the influence function is given by

$$L_{\hat{\boldsymbol{\theta}}}(y, \boldsymbol{x}) = -\big(\boldsymbol{I}(\hat{\boldsymbol{\theta}})\big)^{-1} \left.\frac{\partial \ln f_\varepsilon(y - \boldsymbol{x'\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{21}$$

where $\boldsymbol{I}(\hat{\boldsymbol{\theta}})$ denotes the Fisher information. It is straightforward to calculate the influence for each observation in this way, since both of the terms on the right-hand side of Eq. 15 are ordinarily available as by-products of MLE in most statistical and econometric software packages: the inverse information may be estimated by the covariance matrix which, following the Cramér-Rao Theorem, is typically the inverse of the observed information, and the second term is simply the observation-level score vector.

It is useful at this point to dispel a common misconception that MLE is non-robust in general. MLE is sometimes said to be 'non-robust' because of its typical $1/n$ ordinary breakdown point (i.e. non-resistance to leverage points). However, this is a property shared by most M-estimators[3], even those that are robust in the sense of having finite gross-error sensitivity to contamination in y. In other cases, it is perhaps taken for granted that the gross-error sensitivity will always be infinite under MLE, but as we will see, though that is often the case, it is not always so.

From Eq. 21 we can see that the influence function for the maximum likelihood estimator is simply a linear transformation, by a matrix of constants, of the observation-level score vector. In order to determine whether or not MLE is robust, it is therefore sufficient to check whether the score vector is bounded. This is true of M-estimators generally, but in the case of MLE, this depends entirely on the assumed error distribution. It is therefore of interest to consider how the behaviour of the influence function differs under various assumptions considered in the SF literature, and whether or not some distributional assumptions can be considered robust. We will begin by discussing the canonical cases such as the N-HN and N-EXP specifications, which have received the most attention in the literature, and then consider potentially robust alternatives.

Influence under canonical distributional assumptions

As discussed in Sect. 2, Aigner et al. (1977) and Meeusen and van Den Broeck (1977) originally considered the N-HN and N-EXP cases. An immediate generalisation of the N-HN model is the normal-truncated normal (N-TN) model considered by Stevenson (1980). In each of these cases, f_ε has a convenient closed-form expression, making derivation of the influence function particularly straightforward. Given this, and the large amount of attention given to the N-HN and N-EXP models in the SF literature, it is worth discussing their robustness properties specifically. The 'wrong skew' problem has particular relevance in these cases.

As shown by Waldman (1982), the OLS estimator is a stationary point in the log-likelihood function in the N-HN case, and a sufficient condition for the stability of this stationary point is that the skewness of the OLS residuals has the wrong sign[4]. Horrace and Wright (2020) generalise this result to the N-TN model and the normal-doubly truncated normal (N-DTN) model proposed by Almanidis et al. (2014), and show that the OLS estimator is also a stable stationary point in the N-EXP case. This is well-known in the SF literature as the 'wrong skew' problem. It follows immediately MLE of the SF model is non-robust in these cases: it is straightforward to show via the influence function chain rule that the skewness of the OLS residuals is itself a non-robust functional, implying that even a single contaminating outlier in the 'wrong' direction may result in the OLS estimator being returned. In the event of wrong skew, the influence function therefore becomes that of the OLS estimator.

In other instances, the influence function is more complex. For the N-HN model, where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_v, \sigma_u)'$,

[3] Other M-estimation techniques, e.g. QR, are therefore no more 'robust' than MLE in this sense.

[4] Assuming independence of v_i and u_i, and that the distribution of u_i is positively skewed, $\text{sgn}\big(\text{Skew}(v_i - su_i)\big) = \text{sgn}(-s)$, and therefore we expect the OLS residuals to be negatively skewed in the production frontier case, and positively skewed in the cost function case.

$$L_{\hat{\boldsymbol{\theta}}}(y, \boldsymbol{x}) = -\left(\boldsymbol{\mathcal{I}}(\hat{\boldsymbol{\theta}})\right)^{-1} \begin{pmatrix} -\dfrac{\hat{\sigma}_u/\hat{\sigma}_v}{\sqrt{\hat{\sigma}_v^2 + \hat{\sigma}_u^2}}\left(\left(\dfrac{\hat{\sigma}_v}{\hat{\sigma}_u}\right)^2 z + sh(z)\right)\boldsymbol{x} \\ \dfrac{\hat{\sigma}_v}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2}\left(1 - 2\left(\dfrac{\hat{\sigma}_v}{\hat{\sigma}_u}\right)^2 z^2 - s\left(2 + \left(\dfrac{\hat{\sigma}_u}{\hat{\sigma}_v}\right)^2\right) zh(z)\right) \\ \dfrac{\hat{\sigma}_u}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2}\left(1 - 2\left(\dfrac{\hat{\sigma}_v}{\hat{\sigma}_u}\right)^2 z^2 + s\left(\dfrac{\hat{\sigma}_v}{\hat{\sigma}_u}\right)^2 zh(z)\right) \end{pmatrix}, \quad (22)$$

$$h(z) = \frac{\phi(sz)}{1 - \Phi(sz)}, \qquad z = \frac{y - \boldsymbol{x'}\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}_v^2 + \hat{\sigma}_u^2}}\frac{\hat{\sigma}_u}{\hat{\sigma}_v},$$

where ϕ and Φ are the standard normal probability density and distribution functions, respectively. From this, we can conclude that MLE is non-robust in the N-HN case generally, since each element of the influence function is unbounded, approaching infinity as we increase $\hat{\varepsilon}_i$ in either direction – though the influence increases more slowly for outliers below (above) the production (cost) frontier, there is no robustness to outliers caused by extremely inefficient firms.

The unboundedness of the influence function in the N-TN case is similar and has previously been shown by Song et al. (2017). In the N-EXP case, intuition suggests the model should be better at handling outliers, since the exponential distribution has heavier tails than the half normal distribution. However the influence function for the N-EXP model

$$L_{\hat{\boldsymbol{\theta}}}(y, \boldsymbol{x}) = -\left(\boldsymbol{\mathcal{I}}(\hat{\boldsymbol{\theta}})\right)^{-1} \begin{pmatrix} s\left(\dfrac{1}{\hat{\sigma}_v} h\left(z - \dfrac{\hat{\sigma}_v}{\hat{\sigma}_u}\right) - \dfrac{1}{\hat{\sigma}_u}\right)\boldsymbol{x} \\ \dfrac{\hat{\sigma}_v}{\hat{\sigma}_u^2} - \dfrac{s}{\hat{\sigma}_v}\left(z + \dfrac{\hat{\sigma}_v}{\hat{\sigma}_u}\right) h\left(z - \dfrac{\hat{\sigma}_v}{\hat{\sigma}_u}\right) \\ \dfrac{1}{\hat{\sigma}_u}\left(\dfrac{\hat{\sigma}_v}{\hat{\sigma}_u} h\left(z - \dfrac{\hat{\sigma}_v}{\hat{\sigma}_u}\right) - 1 + z - \dfrac{\hat{\sigma}_v^2}{\hat{\sigma}_u^2}\right) \end{pmatrix}, \quad (23)$$

$$h\left(z - \frac{\hat{\sigma}_v}{\hat{\sigma}_u}\right) = \frac{\phi\left(z - \frac{\hat{\sigma}_v}{\hat{\sigma}_u}\right)}{\Phi\left(z - \frac{\hat{\sigma}_v}{\hat{\sigma}_u}\right)}, \qquad z = -s\frac{y - \boldsymbol{x'}\hat{\boldsymbol{\beta}}}{\hat{\sigma}_u},$$

is likewise unbounded, though it is apparent that influence of outlying observations ought to be less than in the N-EXP case given the contrast between the linear functions of the error term found in Eq. 23 and the quadratic functions that appear in Eq. 22.

Alternative distributional assumptions

As mentioned in Sect. 2, many alternative distributional assumptions have been suggested in the SF literature. Following the intuition that outliers should be modelled as instances of large $|v_i|$ rather than contaminating the distribution of u_i, several studies have explored the use of noise distributions with heavier tails than the usual normal distribution, which ought to better accomodate contaminating outliers.

In an early example of this approach, Janssens and van Den Broeck (1993) suggest the use of an 'approximative t' distribution, and derive f_ε for an approximative t-exponential (AT-EXP) SF model. However the authors do not operationalise the model, and to our knowledge its use remains unexplored. More recent proposals for the heavy-tailed noise distributions include the logistic (Stead et al., 2018), Laplace (Nguyen, 2010; Horrace and Parmeter, 2018), Cauchy (Gupta and Nguyen, 2010), scale contaminated normal (Stead et al., Forthcoming), and Student's t (Tancredi, 2002; Wheat et al., 2019) distributions[5]. Note that the contaminated normal and Student's t distributions contain the the normal distribution is nested as a limiting case; Stead et al. (Forthcoming) and Wheat et al. (2019) discuss testing against the N-HN model, which can be interpreted as testing for heavy-tails in the noise distribution.

A common characteristic of these sub-Gaussian SF models is the atypical tail behaviour of the efficiency predictor, which is discussed in Sect. 5. In terms of parameter estimates, results suggest that adopting such distributions for v_i can lead to substantive changes relative to the normal case, even when assumptions about u_i are unaltered. In empirical applications, estimates from the logistic-half normal (LOG-HN), contaminated normal-half normal (CN-HN) and Student's t-half normal (T-HN) models can differ significantly from the N-HN estimates (Stead et al., Forthcoming; Wheat et al., 2019), especially with respect to σ_u.

An interesting feature of the aforementioned sub-Gaussian noise distributions is that each may be characterised as a scale mixture of normal distributions. That is, in each case

$$f_v(v_i, \boldsymbol{\theta}_v) = \int_0^\infty \frac{1}{\varsigma\sigma_v}\phi\Big(\frac{v_i}{\varsigma\sigma_v}\Big) f_\varsigma(\varsigma)\mathrm{d}\varsigma,$$

where ϕ is the standard normal density and f_ς is the density of the mixing distribution. In the degenerate case, we recover the normal distribution

$$f_\varsigma(\varsigma) = \delta(\varsigma - 1) \Longrightarrow f_v(v_i, \boldsymbol{\theta}_v) = \frac{1}{\sigma_v}\phi\Big(\frac{v_i}{\sigma_v}\Big),$$

while the CN-HN case explored by Stead et al. (2018) is obtained by assuming that ς follows a generalised multinomial distribution, such that

$$f_\varsigma(\varsigma) = \sum_j [\varsigma = \varsigma_j] p_j \Longrightarrow f_v(v_i, \boldsymbol{\theta}_v) = \sum_j p_j \frac{1}{\varsigma_j\sigma_v}\phi\Big(\frac{v_i}{\varsigma_j\sigma_v}\Big).$$

This discrete approach may be used to approximate any mixing distribution arbitrarily well, though at the cost of increasing the number of parameters to be estimated in more flexible cases. Moving from a finite to an infinite mixture setting, Stefanski

[5] For packages facilitating estimation some of these models, see https://github.com/AlexStead.

(1991) shows that the logistic distribution is obtained when the mixing distribution is Kolmogorov-Smirnov; that is

$$f_\varsigma(\varsigma) = 2\varsigma \sum_{m=1}^{\infty} (-1)^{m+1} m^2 e^{-\frac{1}{2}(m\varsigma)^2} \implies f_v(v_i, \boldsymbol{\theta}_v) = \frac{\exp\left(-\frac{v_i}{\sigma_v}\right)}{\sigma_v\left(1+\exp\left(-\frac{v_i}{\sigma_v}\right)\right)^2}.$$

Similarly, Andrews and Mallows (1974) show that the Laplace distribution is obtained when the mixture distribution is exponential with a variance of 4, such that

$$f_\varsigma(\varsigma) = \frac{1}{2} e^{-\frac{\varsigma}{2}} \implies f_v(v_i, \boldsymbol{\theta}_v) = \frac{1}{2\sigma_v} \exp\left(-\frac{|v_i|}{\sigma_v}\right),$$

and that the Student's t distribution is obtained when ς^2 follows an inverse gamma distribution with scale and shape parameter both equal to $\frac{1}{2\alpha}$, such that

$$f_\varsigma(\varsigma) = \frac{(2\alpha\sqrt{\varsigma})^{-\frac{1}{2\alpha}}}{\Gamma\left(\frac{1}{2\alpha}\right)\varsigma} e^{-\frac{1}{2\alpha\sqrt{\varsigma}}} \implies f_v(v_i, \boldsymbol{\theta}_v) = \sqrt{\frac{2\alpha}{\pi}} \frac{\Gamma\left(\frac{2\alpha+1}{4\alpha}\right)}{\Gamma\left(\frac{1}{2\alpha}\right)\sigma_v} \left(1 + 2\alpha\left(\frac{v_i}{\sigma_v}\right)^2\right)^{-\frac{2\alpha+1}{4\alpha}}, \tag{24}$$

where α is an inverse degrees of freedom parameter[6]. This characterisation is useful since it provides a way of conceptualising these alternative specifications in terms of departures from normality driven by explicit contamination models. Assuming independence of v_i and u_i, the marginal density of ε_i is given by

$$f_\varepsilon(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}, \boldsymbol{\theta}) = \int_0^\infty \int_0^\infty \frac{1}{\varsigma\sigma_v} \phi\left(\frac{y_i - \boldsymbol{x}_i'\boldsymbol{\beta} + su_i}{\varsigma\sigma_v}\right) f_\varsigma(\varsigma) f_u(u, \boldsymbol{\theta}_u) \mathrm{d}\varsigma \mathrm{d}u_i.$$

Making use of the Fubini-Tonelli theorem, we can interchange the order of integration and then integrate out u_i in order to express the contaminated density as

$$f_\varepsilon(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}, \boldsymbol{\theta}) = \int_0^\infty f_\varepsilon^*\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}, (\boldsymbol{\beta}', \varsigma\sigma_v, \boldsymbol{\theta}_u')'\right) f_\varsigma(\varsigma) \mathrm{d}\varsigma, \tag{25}$$

where f_ε^* denotes the corresponding density that arises when $v_i \sim \mathcal{N}(0, \sigma_v^2)$. For example, in the T-HN case, the expression above becomes

$$f_\varepsilon(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}, \boldsymbol{\theta}) = \int_0^\infty \frac{2}{\sigma} \phi\left(\frac{y_i - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right) \Phi\left(\frac{-(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})\lambda}{\sigma}\right) \frac{(2\alpha\sqrt{\varsigma})^{-\frac{1}{2\alpha}}}{\Gamma\left(\frac{1}{2\alpha}\right)\varsigma} e^{-\frac{1}{2\alpha\sqrt{\varsigma}}} \mathrm{d}\varsigma,$$

$$\sigma = \sqrt{\varsigma^2\sigma_v^2 + \sigma_u^2}, \qquad \lambda = \frac{\sigma_u}{\varsigma\sigma_v}.$$

---

[6] Wheat et al. (2019) parameterise the T-HN model in terms of the degrees of freedom $\upsilon = \frac{1}{\alpha}$, but for the purposes of the following discussion, the inverse will be more useful.

This result is potentially useful in solving for or approximating $f_\varepsilon(\varepsilon_i, \boldsymbol{\theta})$, but also suggests an alternative way of conceptualising MLE under these alternative distributional assumptions; if the true distribution is such that $v_i \sim \mathcal{N}(0, \sigma_v^2)$, we are effectively maximising a pseudolikelihood that reflects an particular contamination model.

Simulation evidence suggests that these models perform well under misspecification – Horrace and Parmeter (2018) show that the Laplace-Exponential (L-EXP) performs better when the true data generating process (DGP) is N-EXP than vice-versa, and likewise Wheat et al. (2019) show that that Student's t-half normal (T-HN) model performs well when the DGP is N-HN, while the N-HN model performs poorly when the DGP is T-HN. In addition, the L-EXP and T-HN models appear to perform better than the N-EXP and N-HN models in the presence of 'wrong skew' (Horrace and Parmeter, 2018; Wheat et al., 2019).

While simulation results offer an encouraging indication that sub-Gaussian SF models may be less sensitive to misspecification in a general sense, none of the aforementioned studies examine the behaviour of the influence function under their models explicitly, leaving the actual robustness of the various specifications to outliers an open question. Examination of the properties of influence function in each case would be tedious, especially in cases where f_ε lacks a convenient analytical expression. Such an approach would also shed relatively little light on how we might choose distributional assumptions with robustness in mind.

Recent results from Stead et al. (2023) are useful in this respect. The authors derive some sufficient conditions for boundedness of the score vector and hence the robustness of MLE of the SF model, and discuss distributional assumptions which satisfy these. An exhaustive discussion of these conditions is beyond the scope of this chapter, however they relate to the boundedness of the logarithmic derivatives of the joint density function $f_{v,u}$. The authors show that these conditions are not satisfied under the canonical distributional assumptions – e.g. N-HN, N-EXP – nor even in most of the sub-Gaussian cases just discussed.

In fact, of the proposed alternative noise distributions discussed, only the Student's t distribution (including its Cauchy special case) is found to satisfy the Stead et al. (2023) conditions for robust MLE. Under the assumption of independence, it may be paired with any one-parameter scale family – eg. half normal, exponential – for u_i and these conditions will be satisfied. A particularly attractive feature of the Student's t model is that it appears to be robust to contamination in both y and $\boldsymbol{x}$. In other words, it is robust to both outliers and leverage points, and has an ordinary breakdown point greater than $1/n$; its robustness properties in this sense are better than those of the QR estimator. It should be noted that these results may depend on treating the 'degrees of freedom' or shape parameter as a fixed tuning parameter, rather than estimating it via MLE as in Wheat et al. (2019). The authors note that proposed flexible distributions of u_i with two or more parameters, e.g. gamma, do not satisfy their conditions, nor do copulae that have been proposed for modelling dependence between v_i and u_i.

These results demonstrate that simply adopting a sub-Gaussian distribution for v_i, such as the logistic or Laplace distributions, is not enough to ensure robustness of

MLE. On the other hand, they indicate the flexibility of the Student's t distribution in robust MLE of the SF model. The T-HN and Student's t-exponential (T-EXP) models may be estimated robustly via MLE, and the results may be used as a guide to identify further robust pairings. Usefully, this offers a path to robust estimation while remaining within the framework of MLE.

4.4 Alternative M-estimators

One particular class of M-estimators generalises MLE by specifying some alternative loss function containing $-\ln f_\varepsilon$ as a limiting case, such that

$$\rho(y_i, \boldsymbol{x_i}, \boldsymbol{\theta}) = \begin{cases} \rho_\alpha(y_i, \boldsymbol{x_i}, \boldsymbol{\theta}), & \alpha > 0 \\ -\ln f_\varepsilon(y_i - \boldsymbol{x_i'}\boldsymbol{\beta}, \boldsymbol{\theta}), & \alpha = 0 \end{cases}, \tag{26}$$

$$\boldsymbol{\psi}_\alpha(y, \boldsymbol{x}, \hat{\boldsymbol{\theta}}) = \left.\frac{\partial \rho_\alpha(y, \boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

where $\alpha \geq 0$ is some tuning parameter and $\boldsymbol{\psi}_\alpha(y, \boldsymbol{x}, \hat{\boldsymbol{\theta}})$ is bounded when $\alpha > 0$. The gross-error sensitivity and efficiency of these estimators both decrease as α increases, creating a trade-off between robustness and efficiency. A loss function $\rho_\alpha(y_i, \boldsymbol{x_i}, \boldsymbol{\theta})$ is chosen such that this trade-off is minimised.
Examples of this approach include minimum density power divergence estimation (MDPDE) (Basu et al., 1998), in which

$$\rho_\alpha(y_i, \boldsymbol{x_i}, \boldsymbol{\theta}) = \int f_\varepsilon^{1+\alpha}(y - \boldsymbol{x_i'}\boldsymbol{\beta}, \boldsymbol{\theta})\mathrm{d}y + \frac{\alpha+1}{\alpha} f_\varepsilon^\alpha(y_i - \boldsymbol{x_i'}\boldsymbol{\beta}, \boldsymbol{\theta}).$$

Comparable methods are maximum L_q-likelihood estimation (ML_qLE) (Ferrari and Yang, 2010) and maximum Ψ-likelihood estimation (MΨLE) (Eguchi and Kano, 2001; Miyamura and Kano, 2006). ML_qLE replaces the logarithm of the density with its Box-Cox transformation

$$\rho_\alpha(y_i, \boldsymbol{x_i}, \boldsymbol{\theta}) = -\frac{\alpha+1}{\alpha}\left(f_\varepsilon^{\frac{\alpha}{\alpha+1}}(y_i - \boldsymbol{x_i'}\boldsymbol{\beta}, \boldsymbol{\theta}) - 1\right),$$

while (MΨLE) makes one of several transformations of the likelihood. Miyamura and Kano (2006) propose

$$\rho_\alpha(y_i, \boldsymbol{x_i}, \boldsymbol{\theta}) = \frac{1}{\alpha+1}\int f_\varepsilon^{\alpha+1}(y - \boldsymbol{x_i'}\boldsymbol{\beta}, \boldsymbol{\theta})\mathrm{d}y - \frac{1}{\alpha} f_\varepsilon^\alpha(y_i - \boldsymbol{x_i'}\boldsymbol{\beta}, \boldsymbol{\theta}),$$

which, up to a constant factor, is identical to the minimum density power divergence estimator. MDPDE, ML_qLE, and MΨLE are equivalent to maximising weighted likelihood functions, where outlying observations are downweighted[7].

The use of MDPDE as a robust estimator of the SF model is explored by Song et al. (2017), who provide simulation evidence suggesting that the estimator outperforms MLE in the presence of contaminating outliers, as expected, but also that its small-sample performance is comparable to that of MLE. Similar results with respect to MΨLE are shown by Bernstein et al. (2021)[8], who also explore use of ML_qLE, for which the results are by contrast mixed.

We can understand these approaches as maximising some quasi-likelihood function – see White (1982). As discussed in Sect. 4.3, MLE under alternative distributional assumptions can be conceptualised in the same way. In particular, following Eqs. 24 and 25, if $v_i \sim \mathcal{N}(0, \sigma_v^2)$ then the Student's t model can be understood as a Student's t-based robust M-estimator such that,

$$\rho_\alpha(y_i, \boldsymbol{x_i}, \boldsymbol{\theta}) = -\ln \int_0^\infty f_\varepsilon\big(y_i - \boldsymbol{x_i'\beta}, (\boldsymbol{\beta}, \varsigma\sigma_v, \boldsymbol{\theta}_u)'\big) \frac{(2\alpha\varsigma)^{-\frac{1}{2\alpha}}}{\Gamma\big(\frac{1}{2\alpha}\big)\varsigma} e^{-\frac{1}{2\alpha\varsigma}} \mathrm{d}\varsigma,$$

which fits the general framework described by Eq. 26. As with MDPDE, MΨLE, and ML_qLE, we have an tuning parameter α which governs the trade-off between robustness and efficiency, and $-\ln f_\varepsilon$ is recovered when $\alpha = 0$. This provides an alternative motivation for the Student's t model as a robust M-estimator based on a particular model of contamination. This is advantageous, since the contamination model is then incorporated into efficiency prediction in a consistent way by use of the corresponding Student's t based efficiency predictor. By contrast, other robust M-estimation methods leave robust efficiency prediction as a separate problem.

Further investigation would be useful to compare the performance of these methods in estimating SF models, in terms of robustness to differing contamination models and the trade-off between robustness and efficiency. Intuition suggests that these quasi-likelihood methods may outperform QR and similar methods. The choice of α is crucial in each case. Song et al. (2017) discuss the choice of α under MDPDE, and follow Durio and Isaia (2011) in using an approach to select α based on a measure of the similarity of MDPDE and MLE results. Bernstein et al. (2021) discuss the choice of α under ML_qLE, noting that the estimator is biased except when $\alpha = 0$ and suggest setting α equal to a function of n such that $\alpha \to 0$ in large samples, though in this case we approach MLE. For the Student's t based M-estimator, Wheat et al. (2019) discuss hypothesis testing in the case that α is estimated directly via MLE; if we treat it instead as a fixed tuning parameter, information criteria could be used.

[7] In the case of ML_qLE, Ferrari and Yang (2010) use a different formulation for in terms of $q = \frac{1}{1+\alpha}$ and allow for $q > 1$, which effectively upweights outliers.

[8] note that the authors use the Miyamura and Kano (2006) transformation. As noted previously, this is equivalent to MDPDE, so the similarity of these results is to be expected.

5 Robustness and efficiency prediction

As discussed previously, the robustness literature is concerned with ensuring that departures from our model assumptions, such as contaminating observations, cannot push some functional to a boundary of its sample space. In the case of functionals that are defined as measures of technical or cost efficiency, these must belong to the interval $(0,1]$. In most cases we might consider a finding that *all* firms are on the frontier, in other words that sample mean efficiency is 1, to be unrealistic. Such a situation can only arise when we are at a boundary of the parameter space, and is an issue of the robustness of the estimation method.

We will be concerned if efficiency predictions start approaching zero. On the other hand, when it comes to predicting firm-specific efficiency scores we may not regard an efficiency prediction of 1 as problematic; identification of firms on the frontier may be of particular interest in some applications. However, when seeking to identify the most efficient firms, we may wish to exclude extreme outliers. This motivates consideration of potentially robust efficiency predictors that are not unduly sensitive to extreme outliers.

As discussed in Sect 2, firm-specific efficiency prediction is based on the conditional distribution given by Eq. 2. Since the true parameter values are not known in practice, we use some estimator $\hat{\boldsymbol{\theta}}$. By definition, the influence of a contaminating point $(y_i, \boldsymbol{x_i})$ on some predictor $\hat{u}_j$ of (the natural logarithm of) efficiency evaluated for the j^{th} observation is

$$L_{\hat{u}_j}(y_i, \boldsymbol{x_i}) = \hat{u}_j(y_j - \boldsymbol{x'_j}\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}, \hat{\boldsymbol{\theta}}_{\boldsymbol{i}}) - \hat{u}_j(y_j - \boldsymbol{x'_j}\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$$

where $\hat{\boldsymbol{\theta}}_{\boldsymbol{i}} = \left(\hat{\beta}'_{\boldsymbol{i}}, \hat{\vartheta}'_{\boldsymbol{i}}\right)'$ and $\hat{\boldsymbol{\theta}} = \left(\hat{\beta}', \hat{\vartheta}'\right)'$ denote the estimator including and excluding $(y_i, \boldsymbol{x_i})$, respectively. From this, we can derive the expression

$$\begin{aligned}\hat{u}_l(y_l - \boldsymbol{x'_l}\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}, \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}) - \hat{u}_j(y_j - \boldsymbol{x'_j}\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}, \hat{\boldsymbol{\theta}}_{\boldsymbol{i}}) &= \hat{u}_l(y_l - \boldsymbol{x'_l}\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \hat{u}_j(y_j - \boldsymbol{x'_j}\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) \\ &\quad + L_{\hat{u}_l}(y_k, \boldsymbol{x_k}) - L_{\hat{u}_j}(y_i, \boldsymbol{x_i}). \quad (27)\end{aligned}$$

When $l = j$, Eq. 27 gives us the influence on the prediction for the j^{th} observation of removing a contaminating point $(y_i, \boldsymbol{x_i})$ and replacing it with another, $(y_k, \boldsymbol{x_k})$,

$$\hat{u}_j(y_j - \boldsymbol{x'_j}\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}, \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}) - \hat{u}_j(y_j - \boldsymbol{x'_j}\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}, \hat{\boldsymbol{\theta}}_{\boldsymbol{i}}) = L_{\hat{u}_j}(y_k, \boldsymbol{x_k}) - L_{\hat{u}_j}(y_i, \boldsymbol{x_i}), \quad (28)$$

while if $j = i, l = k$, we have an expression for the effect of replacing $(y_i, \boldsymbol{x_i})$ with $(y_k, \boldsymbol{x_k})$ on the efficiency predictor evaluated at the contaminating point

$$\begin{aligned}\hat{u}_k(y_k - \boldsymbol{x'_k}\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}, \hat{\boldsymbol{\theta}}_{\boldsymbol{k}}) - \hat{u}_i(y_i - \boldsymbol{x'_i}\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}, \hat{\boldsymbol{\theta}}_{\boldsymbol{i}}) &= \hat{u}_l(y_k - \boldsymbol{x'_k}\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \hat{u}_i(y_i - \boldsymbol{x'_i}\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) \\ &\quad + L_{\hat{u}_k}(y_k, \boldsymbol{x_k}) - L_{\hat{u}_i}(y_i, \boldsymbol{x_i}). \quad (29)\end{aligned}$$

Eq. 28 may be interpreted as the effect of changing the i^{th} observation on the j^{th} efficiency prediction where $i \neq j$, and depends only on the change in influence, whereas Eq. 29 gives the effect of changing the i^{th} observation on its own efficiency prediction, which depends also on the direct effect of evaluating the predictor at a different point. Stead et al. (2023) note that, since $\hat{u}_i$ is a function of $\hat{\boldsymbol{\theta}}$, the influence of some contaminating observation $(y, \boldsymbol{x})$ on $\hat{u}_i$ can be derived via the influence function chain rule

$$L_{\hat{u}_i}(y, \boldsymbol{x}) = \frac{\partial \hat{u}_i}{\partial \hat{\boldsymbol{\theta}}'} L_{\hat{\boldsymbol{\theta}}}(y, \boldsymbol{x}). \tag{30}$$

From Eq. 30 we can see that one sufficient, though not necessary, condition for the robustness of $\hat{u}_i$ is that robustness of $\hat{\boldsymbol{\theta}}$, and thus the boundedness of $L_{\hat{\boldsymbol{\theta}}}(y, \boldsymbol{x})$. Stead et al. (2023) apply this formula to compare the sensitivity of efficiency predictions across specifications, finding that predictions from robust specifications appear less sensitive than those from non-robust specifications. The authors, confine their attention to the conditional mean predictor, though it is clear from Eq. 30 that the choice of predictor will also have an impact on sensitivity. We will now consider various predictors and their robustness properties, in terms of both their influence functions and their tail behaviour.

5.1 Conditional mean

The most commonly used efficiency predictor is the conditional mean. Following Jondrow et al. (1982), this is given by

$$\mathbb{E}(u_i|y_i - \boldsymbol{x}_i'\boldsymbol{\beta})\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \int_0^\infty u_i f_{u|\varepsilon}(u_i|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}, \boldsymbol{\theta}) \mathrm{d}u_i\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{31}$$

or alternatively following Battese and Coelli (1988), we use

$$\mathbb{E}(e^{-u_i}|y_i - \boldsymbol{x}_i'\boldsymbol{\beta})\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \int_0^\infty e^{-u_i} f_{u|\varepsilon}(u_i|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}, \boldsymbol{\theta}) \mathrm{d}u_i\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

It will be convenient to limit discussion to the Jondrow et al. (1982) predictor. From Jensen's inequality, we can see that

$$\exp\big(-\mathbb{E}(u_i|y_i - \boldsymbol{x}_i'\boldsymbol{\beta})\big) \leq \mathbb{E}(e^{-u_i}|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}),$$

and in practice the difference between the two predictors is usually negligible. Let us consider the influence of contaminating observations on the Jondrow et al. (1982) predictor given by Eq. 27. From Eq. 30, the influence function is given by

$$L_{\mathbb{E}(u_i|\varepsilon_i)}(y, \boldsymbol{x}) = \frac{\partial \mathbb{E}(u_i|y_i - \boldsymbol{x}_i'\boldsymbol{\beta})}{\partial \boldsymbol{\theta}}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} L_{\hat{\boldsymbol{\theta}}}(y, \boldsymbol{x}).$$

The robustness of the conditional mean efficiency predictor therefore depends not only on the robustness of the estimator $\hat{\boldsymbol{\theta}}$ but also on the derivative of the predictor with respect to the estimated parameter vector, which will depend on the model's distributional assumptions. Again, it would be tedious to examine this derivative under all proposed distributional assumptions, but in the N-HN case

$$\mathbb{E}(u_i|y_i - \boldsymbol{x_i'\beta})\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{\hat{\sigma}_v\hat{\sigma}_u}{\sqrt{\hat{\sigma}_v^2+\hat{\sigma}_u^2}}\left(\frac{\phi(z_i)}{1-\Phi(sz_i)} - sz_i\right),$$

$$\frac{\partial\mathbb{E}(u_i|y_i - \boldsymbol{x_i'\beta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \begin{pmatrix} -s\dfrac{\hat{\sigma}_u^2}{\hat{\sigma}_v^2+\hat{\sigma}_u^2}\Big(h'(z_i)-1\Big)\boldsymbol{x_i} \\ \left(\dfrac{\hat{\sigma}_u}{\sqrt{\hat{\sigma}_v^2+\hat{\sigma}_u^2}}\right)^3\left(h(z_i) - sz_i - sz_i\hat{\sigma}_v\left(2\dfrac{\hat{\sigma}_v}{\hat{\sigma}_u}+\dfrac{\hat{\sigma}_u}{\hat{\sigma}_v}\right)\big(h'(z_i)-1\big)\right) \\ \left(\dfrac{\hat{\sigma}_v}{\sqrt{\hat{\sigma}_v^2+\hat{\sigma}_u^2}}\right)^3\left(h(z_i) - sz_i + sz_i\dfrac{\hat{\sigma}_u}{\hat{\sigma}_v}\big(h'(z_i)-1\big)\right) \end{pmatrix},$$

$$h'(z_i) = h(z_i)\big(h(z_i) - sz_i\big) \qquad h(z_i) = \frac{\phi(sz_i)}{1-\Phi(sz_i)}, \qquad z_i = \frac{y_i - \boldsymbol{x_i'\hat{\beta}}}{\sqrt{\hat{\sigma}_v^2+\hat{\sigma}_u^2}}\frac{\hat{\sigma}_u}{\hat{\sigma}_v},$$

which is clearly unbounded. Therefore, even if our estimator $\hat{\boldsymbol{\theta}}$ is robust, contaminating outliers could have an arbitrarily large impact on efficiency predictions for some observations. Additionally, in the N-HN case the tail behaviour of the conditional mean predictor is such that

$$\lim_{s\varepsilon\to\infty}\mathbb{E}(u_i|y_i - \boldsymbol{x_i'\beta}) = \infty, \qquad \lim_{s\varepsilon\to-\infty}\mathbb{E}(u_i|y_i - \boldsymbol{x_i'\beta}) = 0,$$

so that efficiency predictions can approach zero or one as the magnitude of the estimated residual is increased, depending on the sign. Under the assumption of independence of v_i and u_i, Ondrich and Ruggiero (2001) show that, when the distribution of v_i is log-concave, the conditional mean predictor decreases (increases) monotonically as the residual increases (decreases) in a production (cost) frontier setting. Their result implies that this monotonicity is strong when the log-concavity is strong, weak where the log-concavity is weak, and that in the case of log-convex v_i, the direction of the relationship may be reversed. The distribution of v_i is therefore crucial in determining whether or not the efficiency predictions may approach zero or one when the residual is sufficiently large in magnitude. Since the normal distribution is strongly log-concave everywhere, the Ondrich and Ruggiero (2001) result implies that this is the case not only in the N-HN model, but whenever $v_i \sim \mathcal{N}(0, \sigma_v^2)$ and v_i and u_i are independent. It follows from this that the conditional mean cannot be considered robust in these cases.

Under alternative distributional assumptions, the conditional mean may well be bounded or even non-monotonic; Horrace and Parmeter (2018) show that, in the L-TL and L-EXP cases, the predictor is only weakly monotonic, being constant when the estimated residual is positive (negative) for a production (cost) frontier. This is in accordance with the Ondrich and Ruggiero (2001) result, since the Laplace distribution is only weakly log-concave everywhere. The logistic distribution is strongly log-concave everywhere but approaches weak log-concavity at the tails, and accordingly Stead et al. (2018) show that the conditional mean predictor in the Log-HN case appears to approach finite, non-zero limits at the tails. Under the CN-HN specification explored by Stead et al. (2023), the predictor is non-monotonic, changing direction at the shoulders of the distribution where the scale-contaminated normal distribution is strongly log-convex, but since the tails of the distribution are Gaussian, efficiency predictions nevertheless approach zero or one as the magnitude of the residual becomes large. By contrast, the Student's t distribution being strongly log-convex in its tails, Wheat et al. (2019) show that conditional mean predictor in the T-HN case is also non-monotonic, but that the change in direction is sustained as the residual becomes large in magnitude; in fact, the conditional mean appears to approach the unconditional mean $\mathbb{E}(u_i)$ in both directions as $|\hat{\varepsilon}_i| \to \infty$.

Tancredi (2002) sheds additional light on the latter result, contrasting the limiting behaviour of $f_{u|\varepsilon}$ under the N-HN and Student's t-half t (T-HT) cases; in the former case, $f_{u|\varepsilon}$ becomes increasingly concentrated as $|\varepsilon_i| \to \infty$, while in the latter case $f_{u|\varepsilon}$ becomes increasingly flat. This points to an important qualitative difference in the way the two models handles efficiency prediction for outlying observations – in the N-HN case, prediction uncertainty decreases as $|\hat{\varepsilon}_i| \to \infty$, while in the T-HT case the prediction uncertainty increases. A similar result appears to hold in the T-HN case. This differing tail behaviour of the conditional mean predictor and the conditional distribution generally clearly have important implications in terms of comparing efficiency predictions between firms, and especially in identifying the most and least efficient firms in a sample.

Overall, there appears to be a link between log-convexity of the distribution of v_i and robustness of the conditional mean predictor. This apparent link is interesting, since the results of Stead et al. (2023) indicate a link between log-convexity and robustness in estimation. This suggests that distributional assumptions are key, and that under appropriate distributional assumptions, robust estimation and robust efficiency prediction coincide.

5.2 Conditional Mode

As an alternative to the conditional mean, Jondrow et al. (1982) also suggested using the mode of the conditional distribution. The conditional mode predictor is

$$\mathrm{M}(u_i|y_i - \boldsymbol{x}_i'\boldsymbol{\beta})\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \arg\min_{u_i} \big(- f_{u|\varepsilon}(u_i|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}, \boldsymbol{\theta})\big)\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{32}$$

which, as Jondrow et al. (1982) noted, is analogous to a maximum likelihood estimator. This suggests that the conditional mode predictor will not generally be robust to contaminating outliers. It also suggests possible approaches to robust prediction, drawing on the literature on robust M-estimation. For example, we could apply a

Box-Cox transformation to $f_{u|\varepsilon}$ in Eq. 32 for an approach analogous to ML_qLE. The conditional mode has not received as much attention in the SF literature as the conditional mean, but in the N-HN case is given by

$$\mathrm{M}(u_i|\varepsilon_i)\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -s\hat{\sigma}_v\hat{\sigma}_u z_i \mathbb{I}_{sz_i \leq 0}.$$

The conditional mode therefore results in an efficiency prediction of 1 whenever $sz_i > 0$, while approaching zero monotonically as $sz_i \to -\infty$. A similar result holds in the N-EXP model – see Jondrow et al. (1982) for a discussion of the behaviour of the conditional mode in both cases. Although, as discussed, we may not regard an efficiency score of 1 as problematic in some cases, the conditional mode is no more robust than the conditional mean in the opposite direction. Likewise, we can see that the derivative of the predictor in the N-HN case

$$\frac{\partial \mathrm{M}(u_i|y_i - \boldsymbol{x_i'}\boldsymbol{\beta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -s\hat{\sigma}_v \mathbb{I}_{sz_i \leq 0} \begin{pmatrix} -\frac{\hat{\sigma}_u^3}{\sqrt{\hat{\sigma}_v^2 + \hat{\sigma}_u^2}}\boldsymbol{x_i} \\ \frac{\hat{\sigma}_u^2}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2}\frac{\hat{\sigma}_u}{\hat{\sigma}_v} z_i \\ \frac{3\hat{\sigma}_v^2 + 2\hat{\sigma}_u^2}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2} z_i \end{pmatrix} - \hat{\sigma}_v \delta(sz_i) \begin{pmatrix} \frac{\hat{\sigma}_u^3}{\sqrt{\hat{\sigma}_v^2 + \hat{\sigma}_u^2}} z_i \boldsymbol{x_i} \\ \frac{\hat{\sigma}_v\hat{\sigma}_u}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2} z_i^2 \\ -\frac{2\hat{\sigma}_v^2 + \hat{\sigma}_u^2}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2} z_i^2 \end{pmatrix},$$

is unbounded. Thus a contaminating outlier may have an arbitrarily large influence on not only its own efficiency prediction, but efficiency predictions for other observations using the conditional mode in the N-HN case.

The properties of the conditional mode under alternative distributional assumptions do not seem to have been investigated in detail, though findings on the behaviour of the conditional distribution generally imply that its behaviour is similar to that of the conditional mean – Horrace and Parmeter (2018) find that $f_{u|\varepsilon}$ is constant for $s\varepsilon_i \geq 0$, but varies with ε_i when $s\varepsilon_i < 0$. The behaviour of the conditional mode predictor in other cases, such as the Log-HN or T-HN models, is worth exploring further.

5.3 Conditional Median

Yet another potential efficiency predictor is the median of the conditional efficiency distribution, as proposed by Tsukuda and Miyakoshi (2003). This is given by

$$\mathrm{Median}(u_i|y_i - \boldsymbol{x_i'}\boldsymbol{\beta})\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = F_{u|\varepsilon}^{-1}\left(\frac{1}{2}\bigg|y_i - \boldsymbol{x_i'}\boldsymbol{\beta}, \boldsymbol{\theta}\right)\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

where $F_{u|\varepsilon}^{-1}$ denotes the quantile function of the conditional distribution. In the N-HN case this has a convenient expression since, as Jondrow et al. (1982) note, the conditional distribution is simply that of a truncated normal random variable.

Substituting in the relevant parameters, this is given by

$$\text{Median}(u_i|y_i - \boldsymbol{x}_i'\boldsymbol{\beta})\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -s\hat{\sigma}_v\hat{\sigma}_u z_i + \frac{\hat{\sigma}_v\hat{\sigma}_u}{\sqrt{\hat{\sigma}_v^2+\hat{\sigma}_u^2}}\Phi^{-1}\left(\frac{1}{2}\Phi\left(s\sqrt{\hat{\sigma}_v^2+\hat{\sigma}_u^2}z_i\right)\right), \quad (33)$$

where Φ^{-1} denotes the standard normal quantile function[9]. Since the sample median is robust, in contrast to the sample mean, it may be tempting to intuit that the conditional median predictor ought to be a robust alternative to the conditional mean. However, comparisons shown by Tsukuda and Miyakoshi (2003) for the N-HN case indicate that the two predictors are very similar, and from Eq. 33 we can see that it is monotonic and shares the same limits as the conditional mean. In terms of deriving the influence function, the derivative of the predictor in the N-HN case is given by

$$\frac{\partial \text{Median}(u_i|y_i - \boldsymbol{x}_i'\boldsymbol{\beta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{1}{2}\frac{s\phi\left(s\sqrt{\hat{\sigma}_v^2+\hat{\sigma}_u^2}z_i\right)}{\phi\left(\Phi^{-1}\left(\frac{1}{2}\Phi\left(s\sqrt{\hat{\sigma}_v^2+\hat{\sigma}_u^2}z_i\right)\right)\right)}\begin{pmatrix}\frac{\hat{\sigma}_u^2}{\sqrt{\hat{\sigma}_v^2+\hat{\sigma}_u^2}}\boldsymbol{x}_i\\ \hat{\sigma}_v z_i\\ -\hat{\sigma}_u z_i\end{pmatrix}$$

$$+\Phi^{-1}\left(\frac{1}{2}\Phi\left(s\sqrt{\hat{\sigma}_v^2+\hat{\sigma}_u^2}z_i\right)\right)\begin{pmatrix}0\\ \left(\frac{\hat{\sigma}_u}{\sqrt{\hat{\sigma}_v^2+\hat{\sigma}_u^2}}\right)^3\\ \left(\frac{\hat{\sigma}_v}{\sqrt{\hat{\sigma}_v^2+\hat{\sigma}_u^2}}\right)^3\end{pmatrix} - sz_i\begin{pmatrix}0\\ \hat{\sigma}_u\\ \hat{\sigma}_v\end{pmatrix} + \frac{s\hat{\sigma}_v\hat{\sigma}_u^2}{\hat{\sigma}_v^2+\hat{\sigma}_u^2}\begin{pmatrix}\frac{\sqrt{\hat{\sigma}_v^2+\hat{\sigma}_u^2}}{\hat{\sigma}_v}\boldsymbol{x}_i\\ \frac{2\hat{\sigma}_v^2+\hat{\sigma}_u^2}{\hat{\sigma}_v\hat{\sigma}_u}z_i\\ -\left(\frac{\hat{\sigma}_v}{\hat{\sigma}_u}\right)^2 z_i\end{pmatrix},$$

which is unbounded. The conditional median therefore appears no more robust than the conditional mean in the N-HN case. Horrace and Parmeter (2018) derive the conditional median predictor for the L-TL and L-EXP models, which again exhibits similar behaviour to the conditional mean. To summarise, the choice of mean, median, or mode of the conditional distribution appears less important than the model's distributional assumptions with respect to robustness. However, further investigation of possible alternative predictors is needed.

6 Summary and Conclusions

The robustness of stochastic frontier (SF) modelling has been an understudied area, but has been given increased attention in recent years, with the use alternative estimators and distributional assumptions better able to accomodate contaminating outliers explored. Despite this there has been relatively little explicit discussion and comparison of the robustness properties of different estimators and model specifications. In

[9] An equivalent expression for inefficiency, defined as $1 - \exp(-\text{Median}(u_i|\varepsilon_i))$ is given by Tsukuda and Miyakoshi (2003).

this chapter we have aimed to address this gap, discussing the robustness properties of various approaches in terms of the influence functions, gross error sensitivities, and breakdown points of estimators. The discussion of influence is particularly useful, since the concept is easily extended to efficiency prediction, allowing discussion of the sensitivity of efficiency predictions to contaminating observations.

We show that the influence function for the maximum likelihood estimator is unbounded under standard distributional assumptions, although under alternative distributional assumptions the estimator may be robust. Recent results from Stead et al. (2023) give sufficient conditions for robust maximum likelihood estimation (MLE) of the SF model. Some recent proposals, such as the use of logistic or Laplace noise distributions – see Stead et al. (2018) and Horrace and Parmeter (2018), respectively – do not satisfy these conditions. On the other hand, the Student's t distribution for noise satisfying these conditions when paired with many inefficiency distributions.

This offers a route to achieving robust estimation while remaining within the framework of MLE, which is attractive for two main reasons. First, we would like to retain the efficiency of ML estimation. With robust estimation methods, there is generally a trade-off between robustness and efficiency. Second, a key objective of SF modelling is the deconvolution of ε_i into v_i and u_i. Derivation of the Jondrow et al. (1982) and Battese and Coelli (1988) efficiency predictors under alternative distributional assumptions is straightforward. As such, alternative estimation methods can be used to deal with outliers when it comes to estimation, but leave handling outliers in the efficiency prediction stage as a separate problem. Altering distributional assumptions offers a consistent way of dealing with outliers in both stages.

Alternative approaches generalising MLE by changing the loss function such that the influence function is unbounded, such as minimum density power divergence estimation, maximum L_q-likelihood estimation, and maximum Ψ-likelihood estimation, have recently been considered by Song et al. (2017) and Bernstein et al. (2021). Under these approaches, the loss function is transformed such that MLE is contained as a limiting case, and a tuning parameter controls the trade-off between robustness and efficiency. We note that MLE under the assumption of Student's t noise can be conceptualised in the same way, where the transformed loss function is derived directly from an explicit model of contamination, which is also reflected in efficiency prediction.

We show that the corrected ordinary least squares (COLS) approach to estimation is non-robust, though analogous 'corrected' robust regression methods could be considered. We also consider the application of quantile regression (QR) to SF modelling, which has gained attention recently. QR represents another possible approach to robust estimation of the SF model, though the robustness of the estimator is reduced when we choose extreme quantiles. The appropriate choice of quantile reflects underlying distributional assumptions – see Jradi and Ruggiero (2019) and Jradi et al. (2021), suggesting that as with MLE, the robustness of QR estimation of the SF model depends fundamentally on distributional assumptions.

To summarise, recent work on robust estimation of the SF model highlights three main approaches: MLE under appropriate distributional assumptions, alternative robust M-estimation methods, and QR estimation. Each of these belong to the general class of M-estimators, making the derivation and comparison of influence functions straightforward. For large samples, MLE and related approaches may offer a better trade-off between robustness and efficiency than QR when the 'true' model is correctly specified.

With respect to efficiency prediction, we discuss robustness in two related senses: the tail behavior of predictors, and the sensitivity of predictors to contaminating outliers via influence on parameter estimates. The former is relevant when considering the efficiency prediction with respect to gross outliers, and how such outliers may affect the identification of the highest and lowest ranking firms. We discuss derivation of influence functions for efficiency predictors, and their resulting properties. We note that the conditional mean, conditional mode, and conditional median are all non-robust under standard distributional assumptions. Results on the limiting behaviour of the conditional distribution of efficiency again suggest that distributional assumptions are key – for instance, the conditional mean predictor is robust in the Student's t-half normal case.

Our discussion highlights several interesting avenues for future research. With respect to robust estimation, direct comparison of the various robust estimators discussed in the context of SF modelling is needed to establish how they compare in terms of robustness, efficiency, and the trade-off between the two under various settings. In addition, while we identify estimators that are *robust* in the sense of having finite gross error sensitivity, a natural next step would be to investigate estimators that are *resistant* in terms of having high breakdown points. The distinction between the ordinary breakdown point and the finite sample breakdown point is crucial here; QR estimation has a finite sample breakdown point greater than $1/n$, but a $1/n$ ordinary breakdown point, while both are greater than $1/n$ for the the Student's t model. This suggests that the latter offers some robustness to leverage points as well as outliers, but the extent of this resistance needs further investigation. In terms of efficiency prediction, further investigation is needed into the link between distributional assumptions and the robustness of efficiency predictors, and also into possible robust approaches to prediction under standard distributional assumptions.

References

Aigner, D., C. A. K. Lovell and P. Schmidt (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics 6*(1), 21–37.

Almanidis, P., J. Qian and R. C. Sickles (2014). Stochastic frontier models with bounded inefficiency. In R. C. Sickles and W. C. Horrace (Eds.), *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications*, 47–81. New York, NY: Springer.

Andrews, D. F. and C. L. Mallows (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B (Methodological) 36*(1), 99–102.

Bassett, G. Jr. and R. Koenker (1978). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association 73*(363), 618–22.

Basu, A., I. R. Harris, N. L. Hjort and M. C. Jones (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika 85*(3), 549–59.

Battese, G. E. and T. J. Coelli (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics 38*(3), 387–99.

Behr, A. (2010). Quantile regression for robust bank efficiency score estimation. *European Journal of Operational Research 200*(2), 568–81.

Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *Annals of Statistics 5*(3), 445–63.

Bernini, C., F. Marzia and A. Gardini (2004). Quantile estimation of frontier production function. *Empirical Economics 29*(2), 373–81.

Bernstein, D. H., C. F. Parmeter and I. A. Wright (2021). Robust estimation of the stochastic frontier model. Working Paper, University of Miami.

Chernozhukov, V., I. Fernández-Val and B. Melly (2022). Fast algorithms for the quantile regression process. *Empirical Economics 62*(1), 7–33.

Coelli, T. J., D. S. P. Rao and G. E. Battese (2005). *An Introduction to Efficiency and Productivity Analysis*. New York, NY: Springer

Dielman, T. E. (2005). Least absolute value regression: recent contributions. *Journal of Statistical Computation and Simulation 75*(4), 263–86.

Donoho, D. L. and P. J. Huber (1977). The notion of breakdown point. In P. J. Bickel, K. Doksum and J. L. Hodges, Jr. (Eds.), *A Festschrift for Erich L. Lehmann*, 157–84. Belmont, CA.

Durio, A. and Isaia, E. D. (2011). The minimum density power divergence approach in building robust regression models. *Informatica 22*(1), 43–56.

Eguchi, S. and Y. Kano (2001). Robustifing maximum likelihood estimation by psi-divergence. Research Memorandum 802, Institute of Statistical Mathematics.

El Mehdi, R. and C. M. Hafner (2014) Inference in stochastic frontier analysis with dependent error terms. *Mathematics and Computers in Simulation 102*(Supplement C), 104–16.

Ferrari, D. and Y. Yang (2010). Maximum L_q-likelihood estimation. *Annals of Statistics 38*(2), 753–83.

Giloni, A., J. S. Simonoff and B. Sengupta (2006). Robust weighted LAD regression. *Computational Statistics & Data Analysis 50*(11), 3124–40.

Giloni, A. and M. Padberg (2004). The finite sample breakdown point of ℓ_1-regression. *SIAM Journal on Optimization 14*(4), 1028–42.

Greene, W. H. (1980). Maximum likelihood estimation of econometric frontier functions. *Journal of Econometrics 13*(1), 27–56.

Greene, W. H. (1990). A gamma-distributed stochastic frontier model. *Journal of Econometrics 46*(1), 141–63.

Greene, W. H. (2008). The econometric approach to efficiency analysis. In H. O. Fried, C. A. K. Lovell and S. S. Schmidt (Eds.), *The Measurement of Productive Efficiency and Productivity Growth*, 92–159. Oxford, UK: Oxford University Press.

Gupta, A. K. and Nguyen, N. (2010) Stochastic frontier analysis with fat-tailed error models applied to WHO health data. *International Journal of Innovative Management, Information and Production 1*(1), 43–48.

Hampel, F. R. (1968). Contributions to the theory of robust estimation. PhD Thesis, University of California, Berkeley.

Hampel, F. R. (1971). A general qualitative definition of robustness. *Annals of Mathematical Statistics 42*(6), 1887–96.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York, NY: John Wiley

He, X., J. Jurečková, R. Koenker and S. Portnoy (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica 58*(5), 1195–214.

Horrace, W. C. and C. F. Parmeter (2018). A Laplace stochastic frontier model. *Econometric Reviews 37*(3), 260–80.

Horrace, W. C. and Wright, I. A. (2020). Stationary points for parametric stochastic frontier models. *Journal of Business & Economic Statistics 38*(3), 516–26.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics 35*(1), 73–101.

Huber, P. J. (1981). *Robust Statistics*. New York, NY: John Wiley

Janssens, G. K. and J. van den Broeck (1993). Outliers, sample size and robust estimation of stochastic frontier production models. *Journal of Information and Optimization Sciences 14*(3), 257–74.

Jondrow, J., C. A. K. Lovell, I. S. Materov and P. Schmidt (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics 19*(2), 233–8.

Jradi, S. and J. Ruggiero (2019). Stochastic data envelopment analysis: A quantile regression approach to estimate the production frontier. *European Journal of Operational Research 278*(2), 385–93.

Jradi, S., C. F. Parmeter and J. Ruggiero (2019). Quantile estimation of the stochastic frontier model. *Economics Letters 182*, 15–18.

Jradi, S., C. F. Parmeter and J. Ruggiero (2021). Quantile estimation of stochastic frontiers with the normal-exponential specification. *European Journal of Operational Research 295*,(2) 475–83.

Knox, K. J., E. C. Blankmeyer and J. R. Stutzman (2007). Technical efficiency in Texas nursing facilities: A stochastic production frontier approach. *Journal of Economics and Finance 31*(1), 75–86.

Koenker, R. (2005). *Quantile Regression*. Cambridge, UK: Cambridge University Press

Koenker, R. and G. Jr. Bassett (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Kumbhakar, S. C. and C. A. K. Lovell (2000). *Stochastic Frontier Analysis*. Cambridge, UK: Cambridge University Press

Lind, J. C., K. L. Mehra and J. N. Sheahan (1977). Asymmetric errors in linear models: Estimation-theory and Monte Carlo. *Statistics: A Journal of Theoretical and Applied Statistics 23*(4), 305–20.

Liu, C., A. Laporte and B. S. Ferguson (2008). The quantile regression approach to efficiency measurement: Insights from Monte Carlo simulations. *Health Economics 17*(9), 1073–87.

Meeusen, W. and J. van Den Broeck (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review 18*(2), 435–44.

Miyamura, M. and Y. Kano (2006). Robust Gaussian graphical modeling. *Journal of Multivariate Analysis 97*(7), 1525–50.

Mizera, I. and C. H. Müller (2001). The influence of the design on the breakdown points of ℓ_1-type M-estimators In A. Atkinson, P. Hackl and C. H. Müller (Eds.), *MODA6—Advances in Model-Oriented Design and Analysis*, 193–200. Heidelberg, Germany: Physica-Verlag.

Murillo-Zamorano, L. R. (2004). Economic efficiency and frontier techniques. *Journal of Economic Surveys 18*(1), 33–77.

Nguyen, N. (2010). Estimation of technical efficiency in stochastic frontier analysis. PhD Thesis, Bowling Green State University.

Olson, J. A., P. Schmidt and D. A. Waldman (1980). A Monte Carlo study of estimators of stochastic frontier production functions. *Journal of Econometrics 13*(1), 67–82.

Ondrich, J. and J. Ruggiero (2001). Efficiency measurement in the stochastic frontier model. *European Journal of Operational Research 129*(2), 434–42.

Papadopoulos, A. and C. F. Parmeter (2022). Quantile methods for stochastic frontier analysis. *Foundations and Trends in Econometrics 12(1), 1–120*.

Parmeter, C. F. (2001). Is it MOLS or COLS? Efficiency Series Paper 04/21, Departmento de Economía, Universidad de Oviedo.

Parmeter, C. F. and S. C. Kumbhakar (2014). Efficiency analysis: A primer on recent advances. *Foundations and Trends in Econometrics 7*(3–4), 191–385.

Portnoy, S. and R. Koenker (1997). The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statistical Science 12*(4), 279–300.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory 7*(2), 186–99.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association 79*(388), 871–80.

Seaver, B. L. and K. P. Triantis (2006). The impact of outliers and leverage points for technical efficiency measurement using high breakdown procedures. *Management Science 41*(6), 937–56.

Siegel, A. F. (1982). Robust regression using repeated medians. *Biometrika 69*(1), 242–4.

Song, J., D. Oh and J. Kang (2017). Robust estimation in stochastic frontier models. *Computational Statistics and Data Analysis 105*, 243–67.

Stead, A. D., P. Wheat and W. H. Greene (2018). Estimating efficiency in the presence of extreme outliers: A logistic-half normal stochastic frontier model with application to highway maintenance costs in England. In W. H. Greene, L.

Khalaf, P. Makdissi, R. C. Sickles, M. Veall and M. Voia (Eds.), *Productivity and Inequality*, 1–19. Cham, Switzerland: Springer.

Wheat, P., A. D. Stead and W. H. Greene (2017). Allowing for outliers in stochastic frontier models: A mixture noise distribution approach. *Journal of Productivity Analysis*.

Stead, A. D., P. Wheat and W. H. Greene (2019). Distributional forms in stochastic frontier analysis In T. ten Raa T and W. H. Greene (Eds.), *Palgrave Handbook of Economic Performance Analysis*, 225–74. Cham, Switzerland: Palgrave Macmillan.

Stead, A. D., P. Wheat and W. H. Greene (2023). Robust maximum likelihood estimation of stochastic frontier models. *European Journal of Operational Research 309*(1), 188-201.

Stefanski, L. A. (1991). A normal scale mixture representation of the logistic distribution. *Statistics & Probability Letters 11*(1), 69–70.

Stevenson, R. E. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics 13*(1), 57–66.

Tancredi, A. (2002). Accounting for heavy tails in stochastic frontier models. Working Paper 2002.16, University of Padua.

Tsionas, M. G. (2020). Quantile stochastic frontiers. *European Journal of Operational Research 282*(3), 1177–84.

Tsionas, M. G., A. G. Assaf and A. Andrikopoulos (2020). Quantile stochastic frontier models with endogeneity. *Economics Letters 188*, 108964.

Tsukuda, Y. and T. Miyakoshi (2003). An alternative method for predicting technical inefficiency in stochastic frontier models. *Applied Economics Letters 10*, 667-70.

Waldman, D. M. (1982). A Stationary Point for the Stochastic Frontier Likelihood. *Journal of Econometrics 18*(2), 275–9.

Wheat, P., A. D. Stead and W. H. Greene (2019). Robust stochastic frontier analysis: a Student's t-half normal model with application to highway maintenance costs in England. *Journal of Productivity Analysis 51*(1), 21–38.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*(1), 1–25.

Yu, C. and W. Yao (2017). Robust linear regression: A review and comparison. *Communications in Statistics - Simulation and Computation 46*(8), 6261–82.

Zhao, S. (2021). Quantile estimation of stochastic frontier models with the normal–half normal specification: A cumulative distribution function approach. *Economics Letters 206*, 109998.

Zulkarnain, R. and I. Indahwati (2021). Robust stochastic frontier using Cauchy distribution for noise component to measure efficiency of rice farming in East Java. *Journal of Physics: Conference Series 1863*, 012031.