European Journal of Operational Research 309 (2023) 188–201

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Robust maximum likelihood estimation of stochastic frontier models

Alexander D. Stead[a,1,*], Phill Wheat[a], William H. Greene[b]

[a] Institute for Transport Studies, University of Leeds, Leeds, LS2 9LJ, UK
[b] Department of Economics, Stern School of Business, New York University, United States

ARTICLE INFO

Article history:
Received 1 April 2022
Accepted 22 December 2022
Available online 24 December 2022

Keywords:
Robustness and sensitivity analysis
Stochastic frontier analysis
Outliers

ABSTRACT

When analysing the efficiency of decision-making units, the robustness of efficiency scores to changes in the data is desirable, especially in the context of managerial or regulatory benchmarking. However, the robustness of maximum likelihood estimation of stochastic frontier models remains underexplored. We examine the behaviour of the influence function of the estimator in a stochastic frontier context, and derive some sufficient conditions for robust maximum likelihood estimation in terms of the properties of the marginal distributions of the error components and, in cases where they are dependent, the copula density. We find that the canonical distributional assumptions do not satisfy these conditions. The Student's t noise distribution is found to have some particularly attractive properties which means it can be paired with a broad class of inefficiency distributions while still satisfying our conditions under independence. We show that parameter estimates and efficiency predictions from robust specifications are significantly less sensitive to contaminating observations than those from non-robust specifications.

1. Introduction

A regulator or manager of a group of firms or other decision-making units has a clear interest in measuring or estimating their relative efficiency in order to benchmark them against one another, and to provide a basis upon which to set targets for efficiency gains. This constitutes a form of yardstick competition in which decision-making units are incentivised to improve efficiency (Bogetoft, 1997; Shleifer, 1985). The use of efficiency analysis methods such as data envelopment analysis (DEA) (Charnes, Cooper, & Rhodes, 1978) and stochastic frontier (SF) modelling (Aigner, Lovell, & Schmidt, 1977) in such contexts is well-established – for examples and discussion of the application of these methods in economic regulation see e.g. Thanassoulis (2000), Jamasb & Pollitt (2001), Haney & Pollitt (2009), and Agrell & Bogetoft (2017).

It is clearly desirable that efficiency analysis, when used to inform important managerial or regulatory decisions in this way, should be robust in some sense to small changes in the data – e.g. ideally our results should not be unduly sensitive to the addition or removal a single observation, or to small changes in an observation. This is particularly desirable when such decisions are open to challenge by firms which have an incentive to game the analysis to their own advantage.

In the case of DEA, adding, removing, or altering an observation will influence the efficiency score of other firms only if the firm in question belongs to the set of firms that define the frontier. However, DEA efficiency scores are potentially very sensitive to changes affecting the observations that define the frontier, and that sensitivity is compounded by the fact that DEA attributes all departures from the frontier to inefficiency. Approaches to detecting and handling outliers in DEA are discussed by Ondrich & Ruggiero (2002), Simar (2003), Banker & Chang (2006), Tran, Shively, & Preckel (2010), and Bellini (2012).

On the other hand, in the case of econometric methods such as SF modelling, adding, changing, or removing any observation will have some impact on the estimated parameter vector, and therefore the predicted efficiency scores for every observation in the sample. Furthermore, experience shows that the sensitivity of results to even a single observation can be very considerable – for example, several commonly-employed SF specifications are known to suffer from a 'wrong skewness' issue that arises when the least squares residuals are skewed in the opposite direction compared to the theoretical skewness of the model's error – see Waldman (1982), Simar & Wilson (2010), and Horrace & Wright (2020) and

* Corresponding author.
E-mail address: a.d.stead@leeds.ac.uk (A.D. Stead).

[1] The authors gratefully acknowledge funding from the CQC Efficiency Network (see https://nhtnetwork.org/nht_product/cqc-efficiency/).

https://doi.org/10.1016/j.ejor.2022.12.033
0377-2217/

others for discussion of this phenomenon – which can force the predicted inefficiency for all observations to zero.

Such sensitivity to small changes in the data is highly problematic from the perspective of a regulator or manager seeking robust efficiency scores. Following Jondrow, Lovell, Materov, & Schmidt (1982), the usual approach to efficiency prediction in SF modelling is to base efficiency predictions on the distribution of $u_i|\varepsilon_i$. Since the conditional distribution of inefficiency is unknown, in practice plug-in estimators are used, which means that efficiency predictions are influenced by contaminating outliers via their effect on parameter estimates. This suggests that the use of robust estimators should reduce the sensitivity of efficiency predictions to contaminating outliers. A strand of recent literature – discussed in Section 3 – has explored different SF specifications or estimation methods which may offer greater robustness to outliers. Despite this, there has been little explicit discussion of the robustness properties of different estimators.

In this paper, we focus on maximum likelihood (ML) estimation, which is the standard approach to estimation in the SF literature. We derive a set of sufficient conditions for robust ML estimation of cross-sectional parametric stochastic frontier (SF) models of the form described by Assumption 1 below.

Assumption 1.

$$y_i = g(\boldsymbol{x_i}, \boldsymbol{\theta}) + \varepsilon_i, \quad \varepsilon_i = v_i - h(\boldsymbol{\theta})u_i,$$

where $i = 1, \dots, I$ indexes the observation, y_i is the response variable, $g(\boldsymbol{x_i}, \boldsymbol{\theta})$ is some real-valued function, $\boldsymbol{x_i}$ is a covariate vector, $\boldsymbol{\theta}$ is a vector of parameters and v_i and $u_i \geq 0$ are random variables representing noise and inefficiency, respectively, and $h(\boldsymbol{\theta}) \geq 0$ is a non-negative scaling function. The error components v_i and u_i may be dependent or independent of one another, but are uncorrelated with $\boldsymbol{x_i}$.The inefficiency term $h(\boldsymbol{\theta})u_i$ is drawn from a distribution with density function

$$\frac{1}{h(\boldsymbol{\theta})} f_u\left(\frac{u_i}{h(\boldsymbol{\theta})}, \boldsymbol{\theta}\right),$$

where f_u is the density of u_i, which is a density without a scale parameter.

SF models were introduced by Aigner et al. (1977) and Meeusen & van Den Broeck (1977) under specific distributional assumptions – normally distributed v_i, and half-normally or exponentially distributed u_i. A number of alternative distributional assumptions have been proposed – see Stead, Wheat, & Greene (2019) for a recent review.

In recent years, several alternative distributions for v_i have been proposed which have heavier tails than the normal distribution and have loosely been described as more 'robust'. There has also been increasing interest in alternative estimation methods, such as the use of quantile regression to estimate SF models – see Section 3 for a discussion of robustness in the SF literature. Despite this, the robustness properties of estimators of SF models have not been fully explored. This paper aims to further our knowledge of the robustness properties of SF models estimated via ML. For a discussion of the robustness properties of alternative estimation methods such as quantile regression in the context of SF modelling, see Stead, Wheat, & Greene (Forthcoming).

The remainder of this paper is organised as follows: In Section 2, we discuss robust estimation generally, with a focus on influence functions and their use in analysing robustness. In Section 3, we discuss previous work on robustness in the SF literature. In Section 4 we derive some sufficient conditions for robust ML estimation of the SF model. These are derived without making specific distributional assumptions, and are based on simple properties of the marginal noise and inefficiency distributions, with some additional conditions relating to the copula density when the error components are not independent. We discuss specific distributional assumptions in light of these conditions, and show that a Student's t model with fixed degrees of freedom can satisfy our conditions for robust ML estimation under various distributional assumptions about u_i – including any one-parameter scale family of distributions, including e.g. the half normal and exponential distributions. In Section 5, we compare influence across different specifications in a simple application. We examine the influence of individual observations on our estimated parameter vector, and on efficiency predictions. We show that parameter estimates and efficiency predictions from specifications that satisfy our conditions for robust ML estimation are significantly less sensitive to individual observations in our application. Section 6 summarises and concludes.

2. Robust estimation

Discussion of robust estimation often concerns the influence of outlying observations; this is the sense in which we discuss robustness in the present study. Where F is the model's underlying distribution function for the variable of interest, Hampel, Ronchetti, Rousseeuw, & Stahel (1986) define the influence function of the functional $T(F)$ as

$$IF\Big(y, T(F)\Big) = \lim_{h \to 0} \frac{T\Big((1-h)F + h\delta_y\Big) - T(F)}{h}, \tag{1}$$

where δ_y is a point mass at y. The influence function therefore gives the influence on $T(F)$ of an infinitesimal perturbation of the data at y. An estimator $T(F)$ is said to be bias robust, or B-robust[2] if the influence function is bounded. Many estimators examined in the literature on robust estimation belong to the class of M-estimators. Following Huber (1964), an M-estimator $\hat{\boldsymbol{\theta}}$ is one that satisfies the definition

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{I} \rho(y_i, \boldsymbol{\theta}). \tag{2}$$

Equivalently, so long as the *loss function* $\rho(y_i, \boldsymbol{\theta})$ is continuous and differentiable with respect to $\boldsymbol{\theta}$,

$$\sum_{i=1}^{I} \boldsymbol{\psi}(y_i, \hat{\boldsymbol{\theta}}) = \mathbf{0}, \tag{3}$$

where $\boldsymbol{\psi}(y_i, \boldsymbol{\theta}) = \partial\rho(y_i, \boldsymbol{\theta})/\partial\boldsymbol{\theta}$. The class of M-estimators encompasses least squares and ML, and many others. It is well known in the literature on robust estimation that the ML estimator is often non-robust, and alternative robust M-estimators have been proposed. Examples include minimum Hellinger distance (MHD) estimation (Beran, 1977), minimum density power divergence (MDPD) estimation (Basu, Harris, Hjort, & Jones, 1998), maximum Ψ-likelihood (MΨL) estimation (Eguchi & Kano, 2001; Miyamura & Kano, 2006), and maximum L_q-likelihood (ML_qL) estimation (Ferrari & Yang, 2010). However, under certain distributional assumptions, ML estimation can be shown to be robust, and a vast literature explores the use of alternative distributional assumptions. Lucas (1997), Arslan & Genç (2009), and Çankaya & Arslan (2020) examine the robustness of ML estimation of the parameters of Student's t, skew generalised t, and skew exponential power distributions, respectively. From Hampel et al. (1986), the influence function of an M-estimator is given by

$$\boldsymbol{IF}(y_i, \hat{\boldsymbol{\theta}}) = -\left(\mathbb{E}\left(\frac{\partial\boldsymbol{\psi}(y_i, \hat{\boldsymbol{\theta}})}{\partial\hat{\boldsymbol{\theta}}'}\right)\right)^{-1} \boldsymbol{\psi}(y_i, \hat{\boldsymbol{\theta}}). \tag{4}$$

[2] Henceforth, when we refer to robustness, we mean B-robustness.

The influence function is therefore a linear transformation of $\boldsymbol{\psi}(y_i, \hat{\boldsymbol{\theta}})$. To show that an M-estimator is robust for a given model, it is sufficient to show that $\boldsymbol{\psi}(y_i, \hat{\boldsymbol{\theta}})$ is bounded. The boundedness of $\boldsymbol{\psi}(y_i, \hat{\boldsymbol{\theta}})$ follows from our distributional assumptions and the estimation method used.

3. Robustness in the stochastic frontier literature

The issue of robustness has received scant attention in the SF literature over the years. Reflecting the broader literature on robust estimation, contributions have tended to take one of several approaches – outlier detection, the use of robust estimation methods, or a change in distributional assumptions.

Early contributions by Janssens & van den Broeck (1993) and Seaver & Triantis (1995) focused on outlier detection and removal, identifying outliers on the basis of standardised residuals from least median of squares and least trimmed squares regressions – see Rousseeuw (1984). Such approaches have an obvious weakness in a SF setting, since they do not account for the asymmetry of the composed error, and instances of large u_i could be discarded as outliers, which defeats the purpose of the analysis. Recently, Henningsen (2020) proposed the use of a 'pseudo-Cook's distance' in which $\hat{y}_i$ is adjusted by removing the predicted value of $h(\boldsymbol{\theta})\mathbb{E}(u_i|\varepsilon_i)$. However, this adjustment is not completely satisfactory since the distribution of predictions of efficiency scores differs from the underlying efficiency distribution (Wang & Schmidt, 2009). A case-weights perturbation approach to identifying influential observations in the normal-half normal SF model is proposed by Zhuo (2018). There is no general consensus on appropriate cut-off points for outlier detection, and outlier removal typically causes greater loss of efficiency than the use of robust estimation methods – simulation evidence in Section 4.1 of Hampel et al. (1986) shows that robust estimators outperform the combination of non-robust classical estimators with various rejection rules in terms of both efficiency and robustness. Intuitively, it makes more sense to simply down-weight some observations – as robust estimators do – rather than reject them entirely; though robust estimators may well effectively discard particularly gross outliers by giving them zero weight.

The usual approaches to estimation are ML estimation and corrected ordinary least squares (COLS). COLS uses least squares to estimate the frontier parameters, with the parameters of the error distribution estimated based on moments of the least squares residuals. Both of these methods require explicit distributional assumptions to be made regarding the noise term v_i and the inefficiency term u_i. Alternatively, one may specify u_i as a deterministic, non-negative function of some vector of covariates, $\boldsymbol{z_i}$, such that

$$y_i = g(\boldsymbol{x_i}, \boldsymbol{\theta}) + v_i - u_i(\boldsymbol{z_i}, \boldsymbol{\theta}),$$

in which case we may estimate the model via nonlinear least squares, avoiding specific distributional assumptions. Note however that the score function in this case is

$$2\sum_{i=1}^{I}\left(y_i + u_i\left(\boldsymbol{z_i}, \hat{\boldsymbol{\theta}}\right) - g\left(\boldsymbol{x_i}, \hat{\boldsymbol{\theta}}\right)\right)\left(\frac{\partial u_i(\boldsymbol{z_i}, \hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} - \frac{\partial g(\boldsymbol{x_i}, \hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right),$$

which is unbounded. Least squares estimation of the SF model, whether in the context of COLS or non-linear least squares, is therefore non-robust to outliers. Robust analogues of these approaches are possible if we change the loss function. In this vein, there has been some recent interest in quantile regression as a means of estimating SF models – see, e.g. Behr (2010), Jradi & Ruggiero (2019), Jradi, Parmeter, & Ruggiero (2019), Tsionas (2020), Tsionas, Assaf, & Andrikopoulos (2020), Jradi, Parmeter, & Ruggiero (2021) and Zhao (2021).

An in-depth exploration of the robustness properties of quantile regression and other robust regression methods the context of SF modelling is beyond the scope of this paper – see Stead et al. (Forthcoming) for a recent discussion.

Approaches to robustification remaining within an ML type framework with specific distributional assumptions involve a change in either the loss function or our distributional assumptions such that the influence function is bounded. There has been increased interest in both approaches in recent years. Song, Oh, & Kang (2008) propose the use of MDPD estimation, and show that estimator is robust in the normal-half normal case, while the ML estimator is not; the influence function is unbounded. The authors find that the MDPD estimator of the normal-half normal model achieves robustness with relatively little loss of efficiency in estimation. Similar approaches are taken by Bernstein, Parmeter, & Wright (2021), who use ML_qL and MΨL estimation, and contrast the performance of these estimators in the normal-half normal case, with that of the ML estimator in applications to simulated datasets with Cauchy contamination.

A number of alternative distributional assumptions about v_i have also been proposed, with focus on distributions with excess kurtosis, which are better able than the normal distribution to accommodate outliers. An early contribution in this respect was the suggestion of an 'approximative t' distribution by Janssens & van den Broeck (1993). More recently, the use of Laplace (Nguyen, 2010; Horrace & Parmeter, 2018), logistic (Stead, Wheat, & Greene, 2018), generalised logistic (Bonnano, De Giovanni, & Domma, 2017), skew normal (Badunenko & Henderson, 2021), extended skew normal (Wei, Zhu, & Wang, 2021), Cauchy (Zulkarnain & Indahwati, 2019; Gupta & Nguyen, 2010, and Student's t (Tancredi, 2002; Wheat, Stead, & Greene, 2019) noise distributions have been explored, with results generally indicating a material impact on parameter estimates and efficiency predictions compared to the standard SF model. The Student's t distribution is attractive due to its flexibility, since the heaviness of its tails varies with a degrees of freedom parameter, such that the normal distribution is encompassed as a limiting case. Wheat et al. (2019) discuss testing against the standard model, and present simulation evidence that the model approximates the standard model when the true noise distribution is normal.

Despite this recent interest, the robustness of ML estimation of SF models remains largely unexplored. To our knowledge, only Song et al. (2008) have examined the properties of the influence function, and their attention was restricted to the normal-half normal case. Hence the robustness properties of the ML estimator of the SF model is generally unknown, under both standard distributional assumptions and proposed 'robust' alternatives. Hence it will be useful to examine the problem without making specific distributional assumptions, and consider conditions under which ML estimation of SF models is robust. We then consider potential distributional assumptions that satisfy these requirements.

4. Maximum likelihood estimation

In this section, we derive some sufficient conditions for robust ML estimation of parametric SF models. The ML estimator of parametric SF models of the form shown in Assumption 1 maximises the log-likelihood function or, equivalently, minimises the negative log-likelihood function, such that

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{I}\left(-\ln f_\varepsilon(\varepsilon_i, \boldsymbol{\theta})\right),$$

where $f_\varepsilon(\varepsilon_i, \boldsymbol{\theta})$ is the marginal density of ε_i. This is derived by solving the integral

$$f_\varepsilon(\varepsilon_i, \boldsymbol{\theta}) = \int_E f_{v,u}\left(\varepsilon_i + h(\boldsymbol{\theta})u_i, \boldsymbol{\theta}\right) d\mu(u_i), \tag{5}$$

where $f_{v,u}(v_i, u_i, \boldsymbol{\theta})$ is the joint density of v_i and u_i, and E is the subset of the real line that contains the support of $h(\boldsymbol{\theta})u_i \in E$. The first-order conditions can be expressed

$$-\sum_{i=1}^{I} \frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln \int_E f_{v,u}\left(\hat{\varepsilon}_i + h(\hat{\boldsymbol{\theta}})u_i, \hat{\boldsymbol{\theta}}\right) d\mu(u_i) = \mathbf{0},$$

where $\mathbf{0}$ is a column vector of zeros and $\hat{\varepsilon}_i = y_i - g(\boldsymbol{x_i}, \hat{\boldsymbol{\theta}})$. We can see that this is an M-estimator as defined in Eqs. (2) and (3), where

$$\rho(y_i, \boldsymbol{x_i}, \boldsymbol{\theta}) = -\ln \int_E f_{v,u}\left(\varepsilon_i + h(\boldsymbol{\theta})u_i, \boldsymbol{\theta}\right) d\mu(u_i),$$

$$\boldsymbol{\psi}(y_i, \boldsymbol{x_i}, \hat{\boldsymbol{\theta}}) = -\frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln \int_E f_{v,u}\left(\hat{\varepsilon}_i + h(\hat{\boldsymbol{\theta}})u_i, \hat{\boldsymbol{\theta}}\right) d\mu(u_i) = \mathbf{0}.$$

Following Eq. (4), the influence function is therefore

$$\boldsymbol{IF}(y_i, \boldsymbol{x_i}, \hat{\boldsymbol{\theta}}) = -\left(\mathcal{I}(\hat{\boldsymbol{\theta}})\right)^{-1} \left(\frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln \int_E f_{v,u}\left(\hat{\varepsilon}_i + h(\hat{\boldsymbol{\theta}})u_i, \hat{\boldsymbol{\theta}}\right) d\mu(u_i)\right). \tag{6}$$

In (6), $\mathcal{I}(\hat{\boldsymbol{\theta}})$ denotes the Fisher information matrix. Since $\mathcal{I}(\hat{\boldsymbol{\theta}})$ and its inverse are matrices of constants, we can draw two conclusions. First, as noted in Section 2, the influence function is simply a linear transformation of $\boldsymbol{\psi}(y_i, \boldsymbol{x_i}, \hat{\boldsymbol{\theta}})$; in this case, the score vector. Second, inspection of the off-diagonal elements of the Hessian for various SF specifications is enough to tell us that their expectations will generally be non-zero, and therefore any given element of the influence function – that is, the influence function for any given parameter – will generally depend on every element of the score vector.

4.1. Propositions and assumptions

With the preceding discussion in mind, the most we can say in general is that boundedness of every element of the score function is a necessary and sufficient condition for boundedness of every element of the influence function, and therefore robustness of the ML estimator of the parameter vector. That is, a sufficient condition for robust ML estimation of the SF model is given by Proposition 1.

Proposition 1. *Satisfaction of the inequality*

$$\left|\frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln f_{v,u}\left(\hat{\varepsilon}_i + h(\hat{\boldsymbol{\theta}})u_i, \hat{\boldsymbol{\theta}}\right)\right| \leq \boldsymbol{b}, \quad ||\boldsymbol{b}||_\infty \leq \infty. \tag{7}$$

is a sufficient condition for the boundedness of the influence function. A proof is shown in the Supplementary Materials (Online Appendices).

One could, in principle, check whether this inequality is satisfied on a case-by-case basis for each SF specification. A preferable approach, which significantly simplifies the process of identifying particular distributional assumptions consistent with robust ML estimation, is to derive some sufficient conditions for robust ML estimation based directly on our underlying distributional assumptions about the error components, v_i and u_i, embodied in the joint density. From Sklar's theorem, any joint density $f_{v,u}(v_i, u_i, \boldsymbol{\theta})$ can be expressed in terms of the product of the marginal densities of the two components and a copula density governing the dependency between them. This leads us to Assumption 2.

Assumption 2. The marginal densities of v_i and u_i are known so that, following Sklar's theorem, we express their joint density in terms of the products of their marginal densities and a copula density such that

$$f_{v,u}(v_i, u_i, \boldsymbol{\theta}) = f_v(v_i, \boldsymbol{\theta}) f_u(u_i, \boldsymbol{\theta}) c_{v,u}\left(F_v(v_i, \boldsymbol{\theta}), F_u(u_i, \boldsymbol{\theta}), \boldsymbol{\theta}\right),$$

where f_v and F_v are the marginal density and distribution functions of v_i, f_u and F_u are the marginal density and distribution functions of u_i, and $c_{v,u}$ is the copula density.

We therefore derive sufficient conditions for robust ML estimation in terms of properties of the marginal densities and the copula density. As we do so, it will be useful to consider certain important special cases. For instance, although various forms of dependence between v_i and u_i have been considered – see e.g. Smith (2008), El Mehdi & Hafner (1976), and Gómez-Déniz & Pérez-Rodríguez (2015) – we typically assume independence. In this case, our sufficient conditions simplify significantly so that they concern only the marginal densities.

Assumption 3. The error components v_i and u_i are independent. In terms of Assumption 2, the copula density is

$$c_{v,u}\left(F_v(v_i, \boldsymbol{\theta}), F_u(u_i, \boldsymbol{\theta}), \boldsymbol{\theta}\right) = 1.$$

Another special case is models in which satisfy Assumption 4 below. This encompasses all cases in which the inefficiency distribution comes from a one-parameter scale family of distributions, and generalisations of these that possess the 'scaling property' – see Alvarez, Amsler, Orea, & Schmidt (2006). This is an important case that applies to many of the simpler inefficiency distributions found in the SF literature. We derive an alternative set of sufficient conditions for these cases, which may be easier to satisfy.

Assumption 4. $h(\boldsymbol{\theta})u_i$ is drawn from a distribution with density function

$$\frac{1}{h(\boldsymbol{\theta})} f_u\left(\frac{u_i}{h(\boldsymbol{\theta})}\right), \quad \frac{\partial f_u(u_i)}{\partial \boldsymbol{\theta}} = \mathbf{0},$$

where f_u is the density of u_i.

Finally, we consider alternative assumptions about the support of $h(\boldsymbol{\theta})u_i \in E$.

Assumption 5. The support of the inefficiency term $h(\boldsymbol{\theta})u_i \in E$ is given by $E = \{u_i \in \mathbb{R} : u_i \geq 0\}$.

Assumption 6. The support of the inefficiency term $h(\boldsymbol{\theta})u_i \in E$ is given by $E = \{u_i \in \mathbb{R} : l(\boldsymbol{\theta}) \geq u_i \geq 0\}$.

Models with bounded inefficiency are discussed by Almanidis, Qian, & Sickles (2014). We find that even when the bound is a function of $\boldsymbol{\theta}$, our general results can be applied, and we show that bounding u_i from above at some constant could help to satisfy our sufficient conditions for robust ML estimation. These conditions make use of the following propositions.

Proposition 2. *Under Assumption 2, the logarithmic derivative of the joint density $f_{v,u}$ with respect to $\hat{\boldsymbol{\theta}}$ may be expressed*

$$\begin{aligned}\frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln f_{v,u}\left(\hat{\varepsilon}_i + h(\hat{\boldsymbol{\theta}})u_i, \hat{\boldsymbol{\theta}}\right) = & \frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln f_v\left(\hat{\varepsilon}_i + h(\hat{\boldsymbol{\theta}})u_i, \hat{\boldsymbol{\theta}}\right) \\ & + \frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln f_u(u_i, \hat{\boldsymbol{\theta}}) \\ & + \frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln c_{v,u}(F_v, F_u, \hat{\boldsymbol{\theta}}).\end{aligned}$$

Under Assumption 3, this simplifies to

$$\begin{aligned}\frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln f_{v,u}\left(\hat{\varepsilon}_i + h(\hat{\boldsymbol{\theta}})u_i, \hat{\boldsymbol{\theta}}\right) = & \frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln f_v\left(\hat{\varepsilon}_i + h(\hat{\boldsymbol{\theta}})u_i, \hat{\boldsymbol{\theta}}\right) \\ & + \frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln f_u(u_i, \hat{\boldsymbol{\theta}}).\end{aligned}$$

Proposition 3. *The logarithmic derivatives of f_v and $c_{v,u}$ with respect to $\hat{\boldsymbol{\theta}}$ are*

$$\begin{aligned}\frac{\partial}{\partial\hat{\boldsymbol{\theta}}}\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)=&D_1\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\\&\left(\frac{\partial}{\partial\hat{\boldsymbol{\theta}}}h(\hat{\boldsymbol{\theta}})u_i-\frac{\partial}{\partial\hat{\boldsymbol{\theta}}}g(\hat{\boldsymbol{\theta}})\right)\\&+D_2\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right),\\ \frac{\partial}{\partial\hat{\boldsymbol{\theta}}}\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})=&D_1\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\\&\times\left(f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\left(\frac{\partial}{\partial\hat{\boldsymbol{\theta}}}h(\hat{\boldsymbol{\theta}})u_i\right.\right.\\&\left.\left.-\frac{\partial}{\partial\hat{\boldsymbol{\theta}}}g(\hat{\boldsymbol{\theta}})\right)+D_2F_v\right)\\&+D_2\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})D_2F_u\\&+D_3\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}}),\end{aligned}$$

where D_j denotes the derivative of a function with respect to its jth argument.

4.2. Conditions for robust maximum likelihood estimation of stochastic frontier models

In this section, we consider conditions for robust ML estimation of the SF model. These follow quite straightforwardly from the assumptions and propositions in the previous subsection. Two sets of conditions are given. Conditions 1a to 4a concern robustness to contamination with respect to y_i only, while Conditions 1b to 4b concern robustness to contamination with respect to both y_i and $\boldsymbol{x_i}$. The first set of conditions is therefore when considering $\boldsymbol{x_i}$ fixed, while the second set is relevant when considering sensitivity not only simple outliers in y_i but also to outliers in $\boldsymbol{x_i}$, i.e. leverage points.

Within both sets, four alternative conditions are given. The first and second conditions relate to the cases in which there is no upper bound on u_i, and v_i and u_i are, respectively, dependent and independent. The third and fourth relate to the cases in which there is an upper bound on u_i, and v_i and u_i are, respectively, dependent and independent.

We note that Conditions 2a, 4a, 2b, and 4b, which apply under independence of v_i and u_i, are relatively simple, since we do not need to consider the properties of the copula density as in Conditions 1a, 3a, 1b, and 3b. Conditions 2a and 2b are of greatest interest, given that the majority of specifications proposed in the SF literature, and the vast majority of empirical applications, assume independence of the error components and do not place an upper bound on u_i.

These conditions will generally be difficult to satisfy. However, focusing on the most general case covered by Condition 1a, we can see that the first inequality will be satisfied by any log-convex noise density. Likewise, the third and fourth inequalities will be satisfied if the noise and inefficiency densities are log-convex in their parameters, and the fifth, seventh, and eight inequalities will be satisfied if the copula density is log-convex in all of its arguments. This leaves only the the second and sixth inequalities, which are stronger. Subsequent discussion will focus on Conditions 1a to 4a, i.e. conditions for robustness to outliers in y_i, though we will return to the case of outliers in y_i and $\boldsymbol{x_i}$ in Section 4.3, when discussing the robustness properties of models with Student's t distributed v_i.

Condition 1a. Under Assumptions 1, 2 and 5, a sufficient condition for robustness of the ML estimator to contamination in y_i is:

$$\begin{aligned}&\left|D_1\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\right|\leq\infty, &&\left|D_1\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)u_i\right|\leq\infty,\\&\left|D_2\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\right|\leq\boldsymbol{b_1}, &&\left|\tfrac{\partial}{\partial\hat{\boldsymbol{\theta}}}\ln f_u(u_i,\hat{\boldsymbol{\theta}})\right|\leq\boldsymbol{b_2},\\&\left|D_1\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})\right|\leq\infty, &&\left|D_1\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})u_i\right|\leq\infty,\\&\left|D_2\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})\right|\leq\infty, &&\left|D_3\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})\right|\leq\boldsymbol{b_3},\\&||\boldsymbol{b_1}||_\infty+||\boldsymbol{b_2}||_\infty+||\boldsymbol{b_3}||_\infty\leq\infty.\end{aligned}$$

Condition 2a. Under Assumptions 1,2,3 and 5, a sufficient condition for robustness of the ML estimator to contamination in y_i is:

$$\begin{aligned}&\left|D_1\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\right|\leq\infty, &&\left|D_1\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)u_i\right|\leq\infty\\&\left|D_2\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\right|\leq\boldsymbol{b_1}, &&\left|\tfrac{\partial}{\partial\hat{\boldsymbol{\theta}}}\ln f_u(u_i,\hat{\boldsymbol{\theta}})\right|\leq\boldsymbol{b_2},\\&||\boldsymbol{b_1}||_\infty+||\boldsymbol{b_2}||_\infty\leq\infty.\end{aligned}$$

Condition 3a. Under Assumptions 1,2 and 6, a sufficient condition for robustness of the ML estimator to contamination in y_i is:

$$\begin{aligned}&\left|D_1\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\right|\leq\infty, &&\left|D_2\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\right|\leq\boldsymbol{b_1},\\&\left|\tfrac{\partial}{\partial\hat{\boldsymbol{\theta}}}\ln f_u(u_i,\hat{\boldsymbol{\theta}})\right|\leq\boldsymbol{b_2}, &&\left|D_1\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})\right|\leq\infty,\\&\left|D_2\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})\right|\leq\infty, &&\left|D_3\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})\right|\leq\boldsymbol{b_3},\\&||\boldsymbol{b_1}||_\infty+||\boldsymbol{b_2}||_\infty+||\boldsymbol{b_3}||_\infty\leq\infty.\end{aligned}$$

Condition 4a. Under Assumptions 1,2,3 and 6, a sufficient condition for robustness of the ML estimator to contamination in y_i is:

$$\begin{aligned}&\left|D_1\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\right|\leq\infty, &&\left|D_2\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\right|\leq\boldsymbol{b_1},\\&\left|\frac{\partial}{\partial\hat{\boldsymbol{\theta}}}\ln f_u(u_i,\hat{\boldsymbol{\theta}})\right|\leq\boldsymbol{b_2}, &&||\boldsymbol{b_1}||_\infty+||\boldsymbol{b_2}||_\infty\leq\infty.\end{aligned}$$

Condition 1b. Under Assumptions 1,2 and 5, a sufficient condition for robustness of the ML estimator to contamination in y_i and $\boldsymbol{x_i}$ is:

$$\begin{aligned}&\left|D_1\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\tfrac{\partial g(\boldsymbol{x_i},\hat{\boldsymbol{\theta}})}{\partial\hat{\boldsymbol{\theta}}}\right|\leq\boldsymbol{b_1}, &&\left|D_1\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)u_i\right|\leq\infty,\\&\left|D_2\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\right|\leq\boldsymbol{b_2}, &&\left|\tfrac{\partial}{\partial\hat{\boldsymbol{\theta}}}\ln f_u(u_i,\hat{\boldsymbol{\theta}})\right|\leq\boldsymbol{b_3},\\&\left|D_1\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})\right|\leq\infty, &&\left|D_1\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})u_i\right|\leq\infty,\\&\left|D_2\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})\right|\leq\infty, &&\left|D_3\ln c_{v,u}(F_v,F_u,\hat{\boldsymbol{\theta}})\right|\leq\boldsymbol{b_4},\\&||\boldsymbol{b_1}||_\infty+||\boldsymbol{b_2}||_\infty+||\boldsymbol{b_3}||_\infty+||\boldsymbol{b_4}||_\infty\leq\infty.\end{aligned}$$

Condition 2b. Under Assumptions 1,2,3 and 5, a sufficient condition for robustness of the ML estimator to contamination in y_i and $\boldsymbol{x_i}$ is:

$$\begin{aligned}&\left|D_1\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\tfrac{\partial g(\boldsymbol{x_i},\hat{\boldsymbol{\theta}})}{\partial\hat{\boldsymbol{\theta}}}\right|\leq\boldsymbol{b_1}, &&\left|D_1\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)u_i\right|\leq\infty\\&\left|D_2\ln f_v\left(\hat{\varepsilon}_i+h(\hat{\boldsymbol{\theta}})u_i,\hat{\boldsymbol{\theta}}\right)\right|\leq\boldsymbol{b_2}, &&\left|\tfrac{\partial}{\partial\hat{\boldsymbol{\theta}}}\ln f_u(u_i,\hat{\boldsymbol{\theta}})\right|\leq\boldsymbol{b_3},\\&||\boldsymbol{b_1}||_\infty+||\boldsymbol{b_2}||_\infty+||\boldsymbol{b_3}||_\infty\leq\infty.\end{aligned}$$

Condition 3b. Under Assumptions 1,2 and 6, a sufficient condition for robustness of the ML estimator to contamination in y_i and $\boldsymbol{x_i}$

is:

$$\left|D_1 \ln f_v\left(\hat{\varepsilon}_i + h(\hat{\boldsymbol{\theta}})u_i, \hat{\boldsymbol{\theta}}\right)\frac{\partial g(\boldsymbol{x_i}, \hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right| \le \boldsymbol{b_1}, \quad \left|D_2 \ln f_v\left(\hat{\varepsilon}_i + h(\hat{\boldsymbol{\theta}})u_i, \hat{\boldsymbol{\theta}}\right)\right| \le \boldsymbol{b_2},$$
$$\left|\frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln f_u(u_i, \hat{\boldsymbol{\theta}})\right| \le \boldsymbol{b_3}, \quad \left|D_1 \ln c_{v,u}(F_v, F_u, \hat{\boldsymbol{\theta}})\right| \le \infty,$$
$$\left|D_2 \ln c_{v,u}(F_v, F_u, \hat{\boldsymbol{\theta}})\right| \le \infty, \quad \left|D_3 \ln c_{v,u}(F_v, F_u, \hat{\boldsymbol{\theta}})\right| \le \boldsymbol{b_4},$$
$$||\boldsymbol{b_1}||_\infty + ||\boldsymbol{b_2}||_\infty + ||\boldsymbol{b_3}||_\infty + ||\boldsymbol{b_4}||_\infty \le \infty.$$

Condition 4b. Under Assumptions 1,2,3 and 6, a sufficient condition for robustness of the ML estimator to contamination in y_i and $\boldsymbol{x_i}$ is:

$$\left|D_1 \ln f_v\left(\hat{\varepsilon}_i + h(\hat{\boldsymbol{\theta}})u_i, \hat{\boldsymbol{\theta}}\right)\frac{\partial g(\boldsymbol{x_i}, \hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right| \le \boldsymbol{b_1}, \quad \left|D_2 \ln f_v\left(\hat{\varepsilon}_i + h(\hat{\boldsymbol{\theta}})u_i, \hat{\boldsymbol{\theta}}\right)\right| \le \boldsymbol{b_2},$$
$$\left|\frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln f_u(u_i, \hat{\boldsymbol{\theta}})\right| \le \boldsymbol{b_3}, \quad ||\boldsymbol{b_1}||_\infty + ||\boldsymbol{b_2}||_\infty + ||\boldsymbol{b_3}||_\infty \le \infty.$$

In the SF literature, it is common to specify one or more of the original parameters of the error distribution as a function of some vector of covariates, $\boldsymbol{z}$, and parameters $\boldsymbol{\vartheta}$. It is useful to note that this does not fundamentally affect the robustness properties of the model, since for some parameter $\theta_k(\boldsymbol{z}, \boldsymbol{\vartheta})$

$$\frac{\partial \ln f_\varepsilon\left(\varepsilon_i, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\vartheta}} = \frac{\partial \ln f_\varepsilon\left(\varepsilon_i, \boldsymbol{\theta}(\boldsymbol{z_i}, \boldsymbol{\vartheta})\right)}{\partial \theta_k(\boldsymbol{z_i}, \boldsymbol{\vartheta})}\frac{\partial \theta_k(\boldsymbol{z_i}, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}},$$

and $\theta_k(\boldsymbol{z}, \boldsymbol{\vartheta})$ and its derivative are constant with respect to ε_i, unless we are considering contamination with respect to $\boldsymbol{z_i}$.

Following the discussion and conditions above, log-convex distributions for noise and inefficiency, and log-convex copula densities are a good place to look, since they are log-convex in their location and scale parameters. This suggests the Student's t distribution and its Cauchy special case as candidate noise distributions. In terms of inefficiency distributions, note that f_u has no scale parameter. This means that any inefficiency distribution which is log-convex in its non-scale parameters can satisfy our conditions for robust estimation. This is a useful result, since it encompasses any one-parameter scale family, such as the half normal or exponential distributions.

The online supplement to this paper explores the properties of several candidate noise, inefficiency, and copula densities in detail. These results reinforce our discussion here. We find that, of several noise distributions considered, only the Student's t and its Cauchy special case satisfy our conditions. Note that the log derivative of the Student's t density with respect to its shape parameter does not appear to be bounded. However, this can be remedied either by fixing the shape parameter – treating it as a tuning parameter – or placing some upper bound upon it. We also find that, aside from one parameter scale families, none of the inefficiency distributions proposed in the SF literature satisfy our condition, and nor do the bivariate copula densities found in the literature.

To summarise, the majority of noise distributions found in the literature are not compatible with our conditions for robust ML estimation, but the Student's t and Cauchy distributions have good robustness properties which mean that they can be paired with any one-parameter scale family of distributions for inefficiency, under the assumption of independence between v_i and u_i.

4.3. Robustness properties of the Student's t model

Discussion in Section 4.2 and the online supplement suggests a model with Student's t distributed v_i, paired with a one-parameter scale family for u_i, for robust ML estimation. It is worth exploring the robustness properties of this estimator in more detail, since its robustness properties appear particularly strong, not only compared to ML estimation under other distributional assumptions considered, but also compared to the quantile regression estimator.

For the sake of simplicity, we consider a simple case in which the frontier is linear in its parameters, v_i is Student's t distributed, and $h(\hat{\boldsymbol{\theta}})u_i$ is drawn from a one-parameter scale family. That is, we begin with Assumptions 1,2, and 3, and additionally assume that

$$\hat{\boldsymbol{\theta}}' = \begin{pmatrix}\hat{\beta} & \hat{\sigma}_v & \hat{\sigma}_u\end{pmatrix}, \quad g(\boldsymbol{x_i}, \hat{\boldsymbol{\theta}}) = \boldsymbol{x_i'}\hat{\boldsymbol{\beta}},$$
$$h(\hat{\boldsymbol{\theta}}) = \hat{\sigma}_u, \quad \frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \ln f_u(u_i) = \boldsymbol{0},$$

where $\boldsymbol{0}$ denotes a vector of zeros, and that

$$\begin{aligned}
&\ln f_v(y_i - \boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + \hat{\sigma}_u u_i, \hat{\sigma}_v, \alpha_v)\\
&\quad = \ln\Gamma\left(\frac{\alpha_v+1}{2}\right) - \ln\Gamma\left(\frac{\alpha_v}{2}\right)\\
&\quad -\frac{1}{2}\ln\pi - \frac{1}{2}\ln\alpha_v - \ln\hat{\sigma}_v\\
&\quad -\frac{\alpha_v+1}{2}\ln\left(1 + \frac{1}{\alpha_v}\left(\frac{y_i - \boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + \hat{\sigma}_u u_i}{\hat{\sigma}_v}\right)^2\right).
\end{aligned}$$

Note that, in line with the preceding discussion (see the online supplement for further explanation) we are treating the degrees of freedom parameter α_v as a fixed tuning parameter, rather than estimating it via ML. Since we may disregard f_u, the relevant derivatives are

$$\frac{\partial \ln f_v(y_i - \boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + \hat{\sigma}_u u_i, \hat{\sigma}_v, \alpha)}{\partial \hat{\boldsymbol{\beta}}} = (\alpha_v+1)\frac{y_i - \boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + \hat{\sigma}_u u_i}{\alpha_v\hat{\sigma}_v + (y_i - \boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + \hat{\sigma}_u u_i)^2}\boldsymbol{x_i},$$
$$\frac{\partial \ln f_v(y_i - \boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + \hat{\sigma}_u u_i, \hat{\sigma}_v, \alpha)}{\partial \hat{\sigma}_v} = -\frac{1}{\hat{\sigma}_v} + \frac{\alpha_v+1}{\hat{\sigma}_v}\frac{(y_i - \boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + \hat{\sigma}_u u_i)^2}{\alpha_v\hat{\sigma}_v + (y_i - \boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + \hat{\sigma}_u u_i)^2},$$
$$\frac{\partial \ln f_v(y_i - \boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + \hat{\sigma}_u u_i, \hat{\sigma}_v, \alpha)}{\partial \hat{\sigma}_u} = -(\alpha_v+1)\frac{y_i - \boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + \hat{\sigma}_u u_i}{\alpha_v\hat{\sigma}_v + (y_i - \boldsymbol{x_i'}\hat{\boldsymbol{\beta}} + \hat{\sigma}_u u_i)^2}u_i,$$

which are all bounded so long as $\alpha_v < \infty, \hat{\sigma}_v > 0$. From Condition 2a, ML estimation under these assumptions is robust. What is notable about this case is that the estimator is robust not only to contamination in y_i, treating $\boldsymbol{x_i}$ as fixed, but also when we allow for contamination in both y_i and $\boldsymbol{x_i}$. In other words, the estimator is robust not only to outliers, but also to leverage points.

This contrasts with ML estimation under standard distributional assumptions, which is robust to neither outliers nor leverage points, but also with the quantile regression estimator, which is robust to outliers but not to leverage points – see Stead et al. (Forthcoming) for a discussion of the influence function of the quantile regression estimator and its robustness properties generally in an SF context.

A corollary of this is that the Student's t model, estimated via ML, has a breakdown point greater than $1/n$; since the influence function is bounded, it would take more than a single contaminating observation to force the estimator to take on arbitrary values. This is true whether we are considering the *conditional* or *finite sample* breakdown point, which considers only contamination in y_i, or the *unconditional* or *ordinary* breakdown point which also considers contamination in $\boldsymbol{x_i}$. This compares favourably with the quantile regression estimator, which has also has a finite sample breakdown point greater than $1/n$, but an ordinary breakdown point of $1/n$. A more in-depth investigation of the breakdown points of the estimator is beyond the scope of this paper – see again Stead et al. (Forthcoming) for further discussion is the context of SF modelling.

5. Empirical application

In this section, we provide a simple empirical application whereby we estimate cross-sectional SF models under several different distributional assumptions and compare the influence of the

Table 1
Parameter estimates.

	$u_i \sim N^+(0,\sigma_v^2)$			$u_i \sim$ Exponential$(0,\sigma_v)$		
	(1)	(2)	(3)	(4)	(5)	(6)
β_1 ($\ln q$)	0.966***	0.961***	0.959***	0.966***	0.962***	0.961***
	(0.013)	(0.008)	(0.011)	(0.012)	(0.011)	(0.012)
β_2 ($\ln^2 q$)	0.030***	0.027***	0.027***	0.029***	0.027***	0.028***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
β_3 ($\ln w$)	0.261***	0.321***	0.324***	0.270***	0.343***	0.324***
	(0.066)	(0.039)	(0.061)	(0.063)	(0.065)	(0.067)
β_4 ($\ln r$)	0.055	0.024	0.036	0.033	0.037	0.034
	(0.062)	(0.046)	(0.053)	(0.059)	(0.053)	(0.053)
β_0	3.735***	3.714***	3.749***	3.764***	3.756***	3.766***
	(0.035)	(0.014)	(0.047)	(0.020)	(0.017)	(0.033)
σ_v	0.109	0.023	0.072	0.104	0.037	0.077
	(0.023)	(0.007)	(0.030)	(0.014)	(0.010)	(0.017)
σ_u	0.149	0.178	0.130	0.097	0.101	0.094
	(0.049)	(0.013)	(0.062)	(0.022)	(0.016)	(0.042)
a	∞	1.000	2.695	∞	1.000	3.557
	-	-	(1.231)	-	-	(2.220)

Standard errors in parentheses N.B. where $a \to \infty, v_i \sim N(0,\sigma_v^2)$ and where $a = 1, v_i \sim$ Cauchy$(0,\sigma_v)$. In these models, a is fixed * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

observations on various model parameters. We use the dataset of Christensen & Greene (1976) on the costs, output, and input prices of a cross-section of US electricity generating firms in 1970. We estimate a generalised Cobb-Douglas stochastic cost frontier specified as follows:

$$\ln\left(\frac{c_i}{e_i}\right) = \beta_0 + \beta_1 \ln q_i + \beta_2 \ln^2 q_i + \beta_3 \ln\left(\frac{w_i}{e_i}\right) + \beta_4 \ln\left(\frac{r_i}{e_i}\right) + v_i + u_i, \quad u_i \geq 0,$$

where c_i is total cost, q_i is output in millions of kilowatt-hours generated, w_i is the price of labour, r_i is the price of capital, and e_i is the price of fuel. The cost and input price variables have all been divided through by the fuel price in order to impose linear homogeneity of degree one in input prices, and the output variable has been normalised by the sample mean for ease of interpretation of the first-order coefficient. We have opted for a relatively simple functional form, as opposed to the translog specification used by Christensen & Greene (1976), in order to keep the number of frontier parameters manageable given the focus here on comparing influence on parameter estimates between error specifications.

Our models differ only in their assumptions about the distributions of v_i and u_i. Models 1–3 assume that u_i follows a half normal distribution, while Models 4–6 assume an exponential distribution for u_i. Models 1 and 4 assume that v_i is normally distributed, Models 2 and 5 assume that v_i follows a Cauchy distribution, and Models 3 and 6 assume that v_i follows a Student's t distribution.

The models we estimate differ with respect to their robustness properties. To simplify matters, all of our models assume that v_i and u_i are independently distributed. Both of our assumed distributions for u_i, half normal and exponential, belong to the class of one-parameter scale families. However, the normal distribution for v_i assumed in Models 1 and 4 does not satisfy our conditions for ML estimation. Nor does the Student's t distribution for v_i assumed for Models 3 and 6, since we estimate the degrees of freedom parameter rather than treat it as a fixed tuning parameter. Models 2 and 5 assume Cauchy distributed v_i; that is, a Student's t distribution with degrees of freedom fixed at 1. Following the discussion in Section 4, we therefore expect that parameter estimates and efficiency predictions from Models (2) and (5) to display less sensitivity to individual observations than those from Models (1), (3), (4), and (6).

Models 1 and 4 were estimated via ML. Models 2, 3, 5, and 6 were estimated via maximum simulated likelihood using 250 Halton draws per observation to approximate the integral in Eq. (5). Parameter estimates and standard errors are shown in Table 1. The estimated frontier parameters are broadly comparable across all six error specifications. The most significant differences can be seen in the estimated parameters of the error distributions. This is highly significant in SF modelling, since the estimated error distributions are critical to efficiency prediction. Across Models 1 to 3, we can obtain an estimate of mean cost efficiency via

$$\mathbb{E}\Big(\exp(-u_i)\Big) = \exp\left(\frac{\hat{\sigma}_u^2}{2}\right)\left(1 - \mathrm{erf}\Big(\frac{\hat{\sigma}_u}{\sqrt{2}}\Big)\right), \quad u \sim N^+(0,\hat{\sigma}_u^2),$$

$$\mathbb{E}\Big(\exp(-u_i)\Big) = \frac{1}{\hat{\sigma}_u^2 + 1}, \quad u \sim \text{Exponential}(0,\hat{\sigma}_u),$$

and significant differences in the estimates parameters of the inefficiency distribution can therefore lead to significantly varying pictures of mean cost efficiency. When considering observation-specific efficiency, predicted using $\mathbb{E}\big(\exp(-u_i)|\varepsilon_i\big)$, our assumptions about the distribution of v_i and our estimates of $\boldsymbol{\theta_v}$ have a significant impact, especially for large values of $|\hat{\varepsilon}_i|$, as shown by Stead et al. (2018) and Wheat et al. (2019) who compare efficiency predictions from the normal-half normal model with those from logistic-half normal and Student's t-half normal models, respectively.

5.1. Calculating influence functions for parameter estimates

For each model, we evaluate the influence function for each observation. As can be seen from Eq. (4), this involves evaluation of the Fisher information matrix. For the models in question, many of the expectations involved will lack convenient analytical solutions. One possible approach would be to use simulation. Instead, in order to simplify the problem, we use the observed information matrix; i.e., the negative of the Hessian. This is a consistent estimator of the Fisher information matrix, and is easy to obtain following model estimation. The vectors of derivatives of the log-likelihood functions with respect to their parameter vectors are likewise easily obtained; for convenience's sake, we use numerical rather than analytical derivatives.

Figs. 1, 2, 3, 4 compare the influences of observations on the various model parameters between models. Note that, since the models include several covariates, there is no straightforward re-

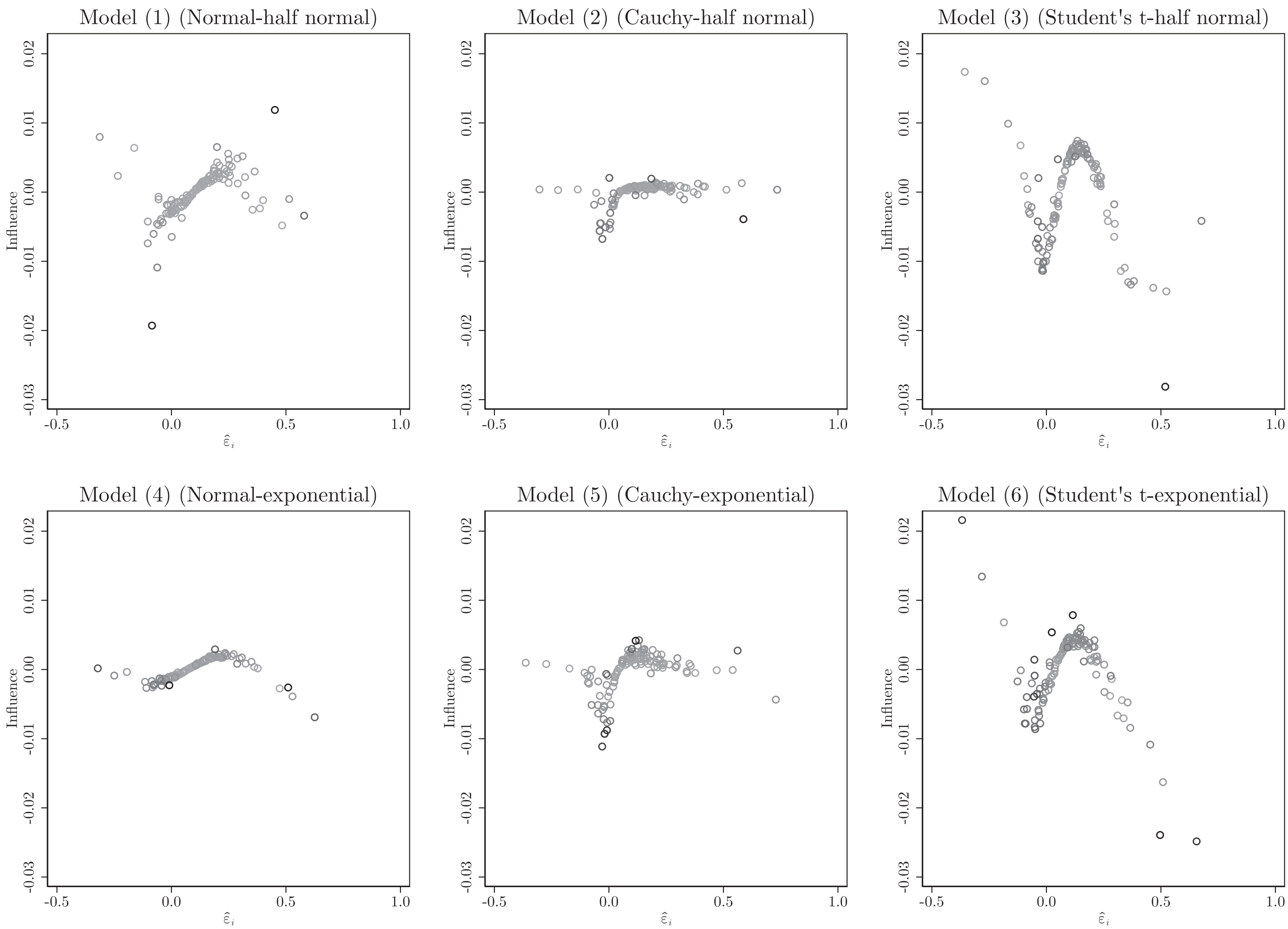

Fig. 1. Influence on $\hat{\beta}_0$.

lationship between influence and any one of the covariates. We therefore plot influence on the vertical axes against residual values on the horizontal while varying the darkness of the markers according to the absolute value of the leverage of the observation in question - i.e. the influence of observations on their own fitted values of the dependent variable.

Fig. 1 compares the influences of observations on the estimated intercept in the cost frontier. Except for the normal-half normal case, a relatively clear relationship can be seen between influence and the value of the residual. Around the centre of the distribution of $\hat{\varepsilon}_i$, we can see in each case a positive relationship between the value of the residual and influence, whereas in the tails the picture is more complicated. It is immediately noticeable that influence on the estimated intercept is generally lower in the models with Cauchy distributed v_i. Comparison of the normal and student's t models is less straightforward; the most extreme values are in fact found in the Student's t models, but this seems to be offset by the relatively large influence of observations in the centre of the distribution.

Our main interest here is in the influence of observations on the estimated parameters of the distributions of v_i and u_i, given that these appear in Table 1 to be the most sensitive to changes in our distributional assumptions, and the fact that they are crucial to the estimation of mean cost efficiency and for the prediction of firm-specific efficiencies. Figs. 2 and 3 compare the influences of observations on $\hat{\sigma}_v$ and $\hat{\sigma}_u$. It should be noted that the scale parameter σ_v is not directly comparable across specifications; when v_i is normally distributed, σ_v is the standard deviation, whereas for the Cauchy distribution this is undefined, and when v_i has a Student's t distribution, the standard deviation is $\sigma_v\sqrt{\alpha_v/(\alpha_v-1)}$. However, we can see that observations appear to be generally more influential in the the normal models than in the Cauchy models. The results for the Student's t models are interesting in that they suggest that observations become more influential as $|\hat{\varepsilon}_i|$ increases, but only to a point. For the most extreme values of $|\hat{\varepsilon}_i|$, influence appears to be declining in magnitude.

Finally, Fig. 4 compares influence on $\hat{\alpha}_v$ – the degrees of freedom or shape parameter of the Student's t distribution – when u_i is assumed to follow half normal and exponential distributions. This figure is of less direct interest, given that we expect neither Model 3 nor Model 6 to be robust to outliers. Most outlying observations appear to exert a positive influence, which is intuitive given that this parameter governs the heaviness of the tails of the noise distribution; note that $\hat{\alpha}_v$ is higher in the Student's t-exponential model than in the Student's t-half normal model. It is interesting to note that parameter seems much more sensitive in the exponential case; contrast this with the sensitivity of $\hat{\sigma}_v$ between the two models – it appears that the scale parameter of the noise distribution is more sensitive to outliers in Model (3) than in Model (6), while the reverse is true with respect to the shape parameter.

Overall, the Cauchy and Student's t models do appear to differ substantially from the normal models in terms of the influence of observations on parameter estimates. The overall picture is that the

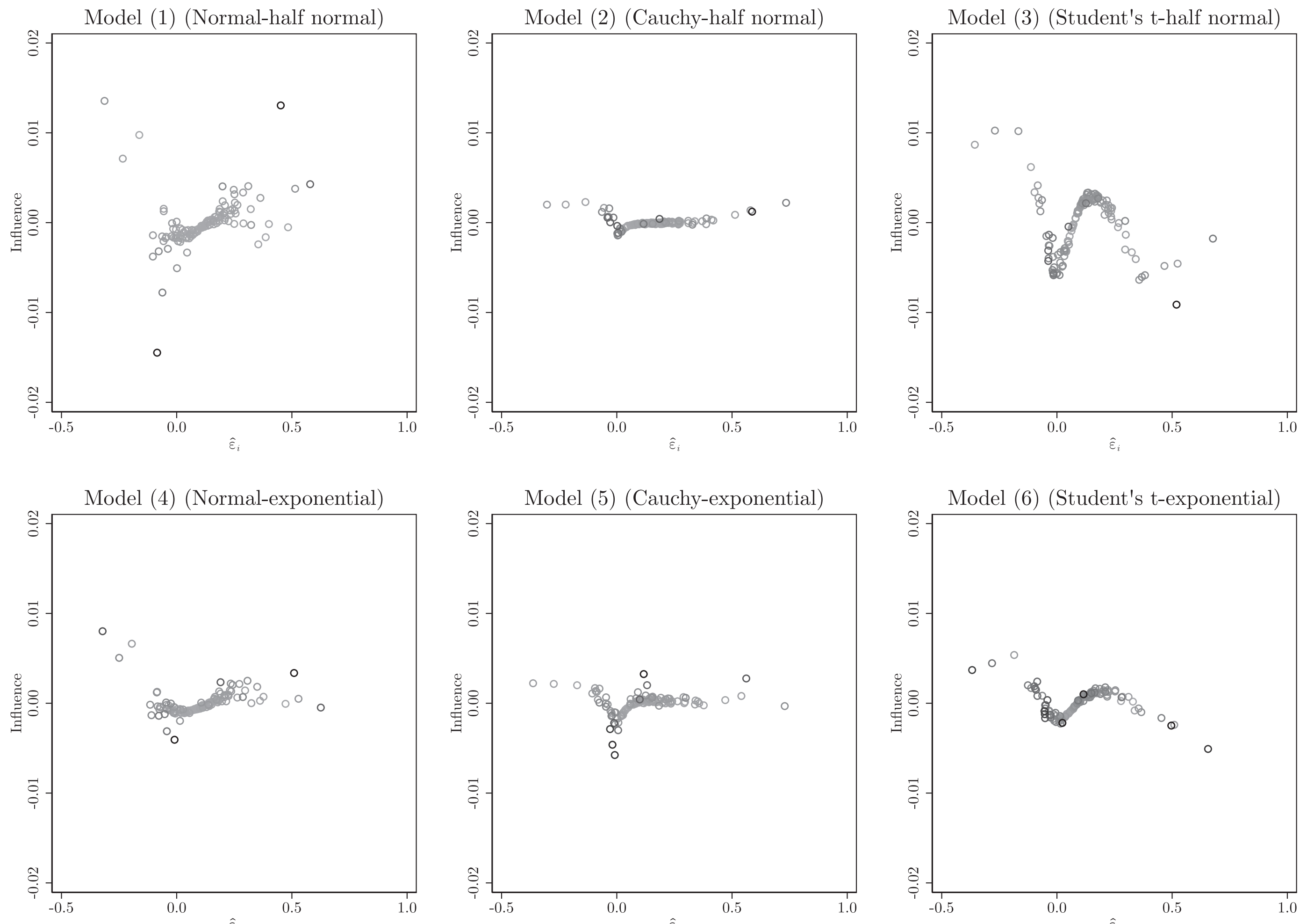

Fig. 2. Influence on $\hat{\sigma}_v$.

influence of observations in the shoulders of the distribution of $|\hat{\varepsilon}_i|$ relative to the influence of observations in the tails is higher than in the normal models. There is some indication also of the magnitude of influence diminishing for the most extreme values of $|\hat{\varepsilon}_i|$. Differences in influence on the estimated coefficients on the covariates in the frontier are relatively hard to discern. However, for the estimated frontier intercept and the parameters of the noise and inefficiency distributions, it is noticeable that the relationship between influence and the value of $\hat{\varepsilon}_i$ appears to be much clearer in the Student's t and Cauchy cases than in the normal case. This is linked to the fact that the leverage of observations is generally lower in magnitude in these models.

5.2. Calculating influence on efficiency predictions

We now consider the influence of observations on predicted efficiency scores. As discussed in Section 1, efficiency prediction is based on the conditional distribution of efficiency. Following Battese & Coelli (1988) we use the conditional expectation $\mathbb{E}\big(\exp(-u_i)|\varepsilon_i\big)$. In practice, since $\boldsymbol{\theta}$ is unknown, we use the plug-in predictor

$$\mathbb{E}\Big(\exp(-u_i)\Big|\hat{\varepsilon}_i\Big) = \frac{1}{f_\varepsilon(\hat{\varepsilon}_i, \hat{\boldsymbol{\theta}})} \int_E \exp(-u_i) f_{v,u}\Big(\hat{\varepsilon}_i + h(\hat{\boldsymbol{\theta}})u_i, \hat{\boldsymbol{\theta}}\Big) d\mu(u_i),$$

which is a statistical functional for which we can derive an influence function. Rather than derive the influence function directly, it is easier in cases such as this to exploit the fact that influence functions are derivatives (specifically, limiting cases of Gâteaux derivatives), and that we may apply an *influence function chain rule* to derive influence of observation j on the efficiency prediction for firm i as

$$IF\left(y_j, \boldsymbol{x_j}, \mathbb{E}\Big(\exp(-u_i)\Big|\hat{\varepsilon}_i\Big)\right) = \frac{\partial \mathbb{E}\Big(\exp(-u_i)\Big|\hat{\varepsilon}_i\Big)}{\partial \hat{\boldsymbol{\theta}}'} \boldsymbol{IF}(y_j, \boldsymbol{x_j}, \hat{\boldsymbol{\theta}}). \quad (8)$$

It is clear from Eq. (8) that if our parameter estimates are non-robust, our efficiency predictions will also be non-robust and hence sensitive to outliers in the data. On the other hand, a robust estimator implies robust efficiency prediction. We therefore expect that our efficiency predictions from Models (2) and (5) should display less sensitivity to contaminating outliers than our efficiency predictions from other specifications.

In principle, we are interested in calculating Eq. (8) across all i and all j. If we define

$$\boldsymbol{IF}(\boldsymbol{y}, \boldsymbol{X}, \hat{\boldsymbol{\theta}}) = \begin{pmatrix} (\boldsymbol{IF}(y_1, \boldsymbol{x_1}, \hat{\boldsymbol{\theta}}))' \\ \vdots \\ (\boldsymbol{IF}(y_i, \boldsymbol{x_i}, \hat{\boldsymbol{\theta}}))' \\ \vdots \\ (\boldsymbol{IF}(y_I, \boldsymbol{x_I}, \hat{\boldsymbol{\theta}}))' \end{pmatrix},$$

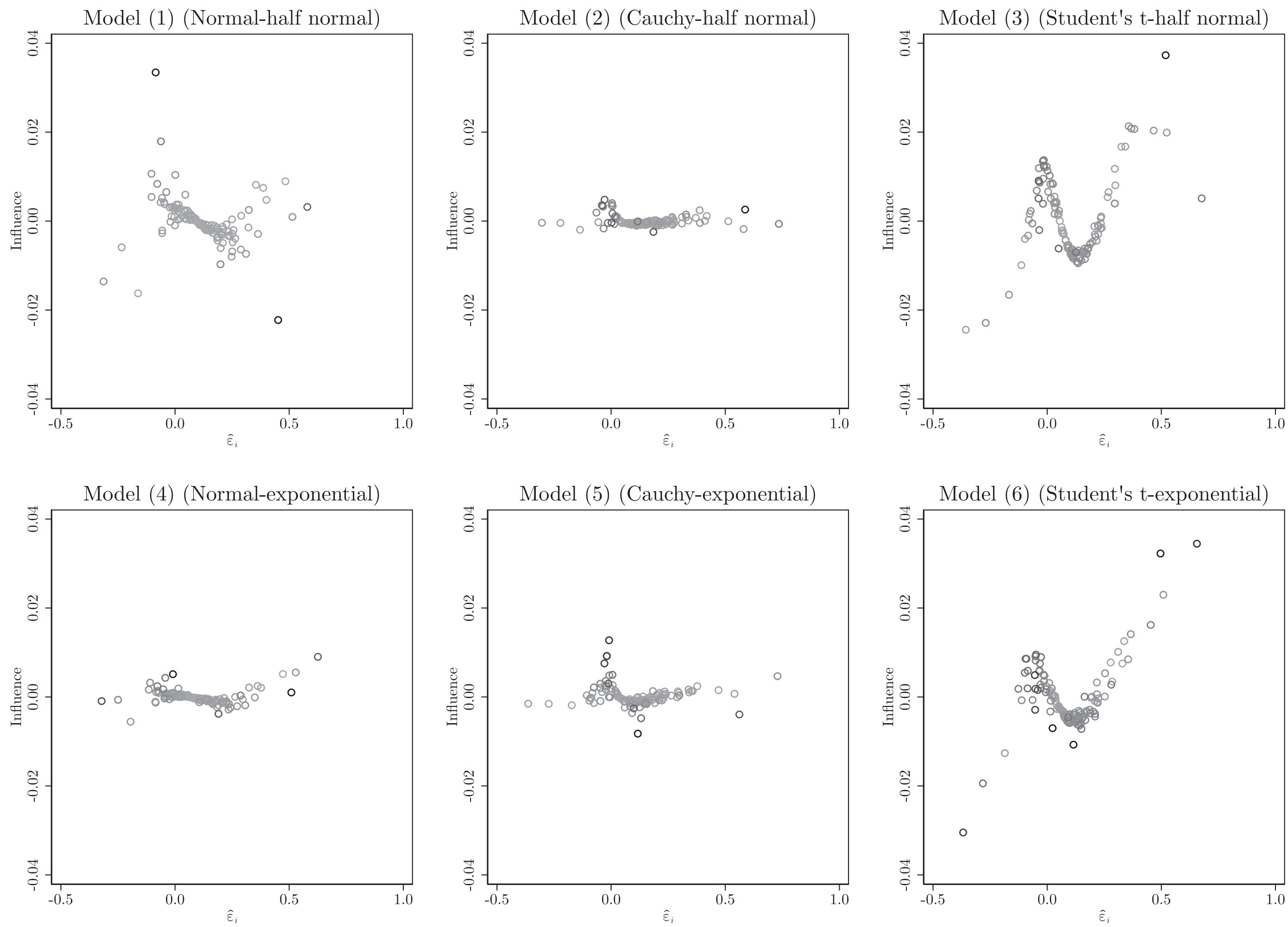

Fig. 3. Influence on $\hat{\sigma}_u$.

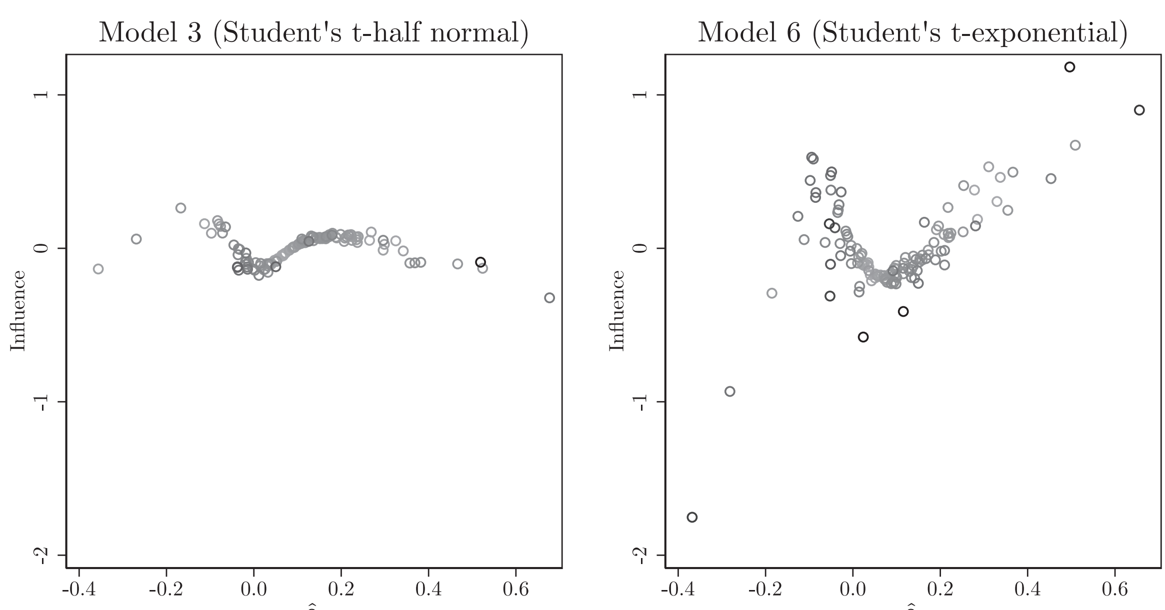

Fig. 4. Influence on $\hat{\alpha}$.

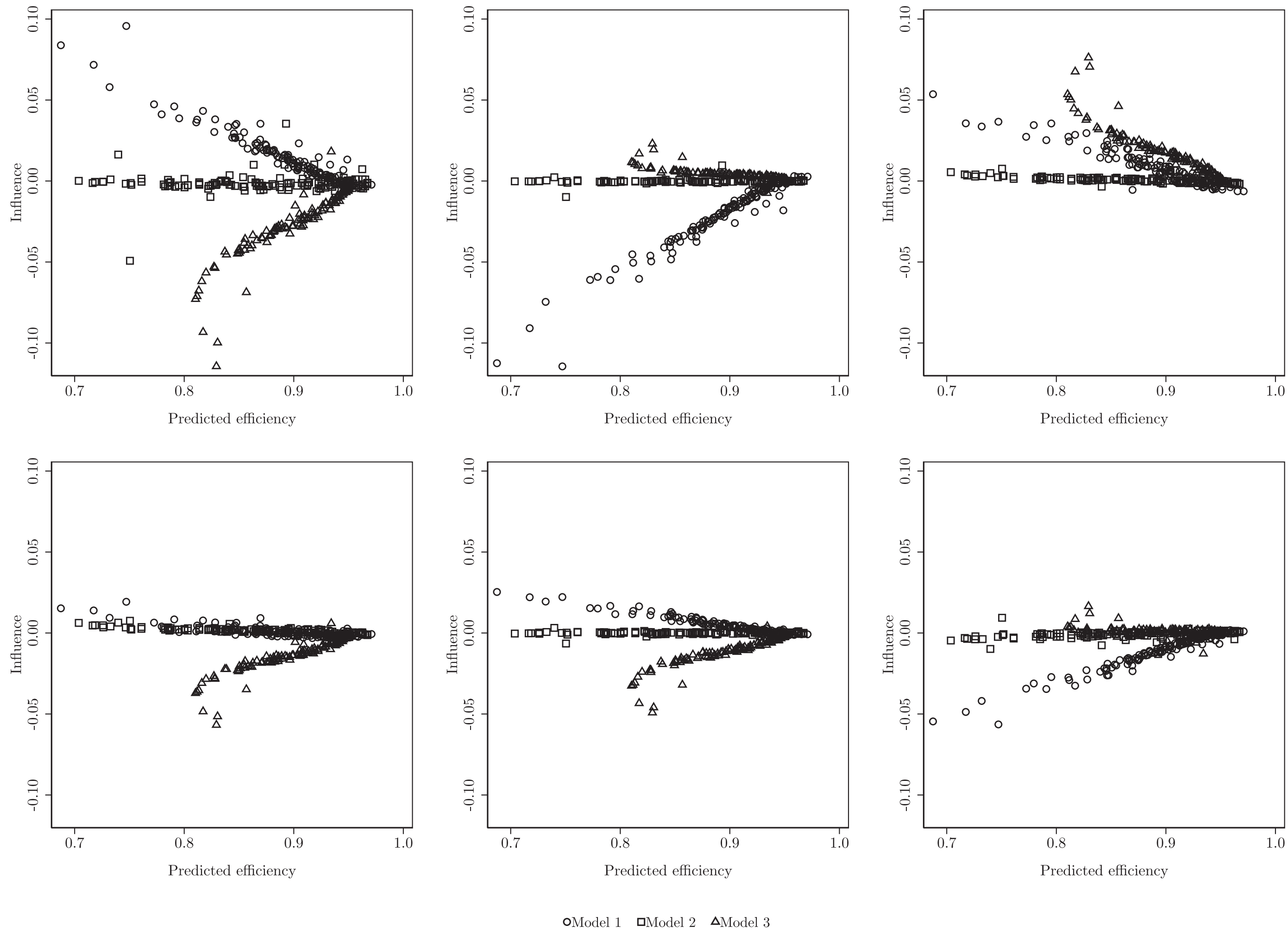

Fig. 5. Influence on efficiency predictions, models (1)–(3).

$$\mathbb{E}\Big(\mathbf{exp}(-\boldsymbol{u})\Big|\hat{\boldsymbol{\varepsilon}}\Big) = \begin{pmatrix} \mathbb{E}\Big(\exp(-u_1)\Big|\hat{\varepsilon}_1\Big) \\ \vdots \\ \mathbb{E}\Big(\exp(-u_i)\Big|\hat{\varepsilon}_i\Big) \\ \vdots \\ \mathbb{E}\Big(\exp(-u_I)\Big|\hat{\varepsilon}_I\Big) \end{pmatrix},$$

then we may define an $I \times I$ matrix of influences of observations on efficiency predictions

$$\boldsymbol{IF}\Big(\boldsymbol{y}, \boldsymbol{X}, \mathbb{E}\Big(\mathbf{exp}(-\boldsymbol{u})\Big|\hat{\boldsymbol{\varepsilon}}\Big)\Big) = \frac{\mathbb{E}\Big(\mathbf{exp}(-\boldsymbol{u})\Big|\hat{\boldsymbol{\varepsilon}}\Big)}{\partial\hat{\boldsymbol{\theta}}'}\big(\boldsymbol{IF}(\boldsymbol{y}, \boldsymbol{X}, \hat{\boldsymbol{\theta}})\big)',$$

of which the element belonging to column i and row j is the influence of observation j on the efficiency prediction for firm i, as given by Eq. (8). As with our calculation of the influence function for the parameter vector, we use numerical derivatives of the efficiency predictors in practice for convenience's sake. Rather than present the full 123×123 matrices for each of our models, we limit our attention to the observations with the greatest influence on predicted efficiency scores.

Fig. 5 plots influence on predicted efficiency against predicted efficiency for Models (1)–(3) (i.e. models in which $u_i \sim N^+(0, \sigma_u^2)$). Each plot corresponds to a specific observation, and shows the influence of that observation on the efficiency predictions for all 123 observations. The six most influential observations, defined in terms of the range of influences, are shown. We can see that these observations have a substantial impact on many efficiency predictions from Models (1) and (3). Efficiency predictions from Model (2) appear to be much less sensitive, with observations having no material impact on efficiency scores for the most part.

Fig. 6 plots influence on predicted efficiency against predicted efficiency for Models (4)-(6) (in which $u_i \sim \text{Exponential}(0, \sigma_u)$). Here the picture is somewhat different, with the highest influence seen in Model (6). However the general picture remains that the observations do not have a material impact on efficiency scores from Model (5) in which $v_i \sim \text{Cauchy}(0, \sigma_v)$, with the exception of one observation shown in the bottom left plot.

Overall, the influences of observations on efficiency predictions are in line with our expectation that efficiency predictions from Models (2) and (5) ought to be less sensitive than those from Models (1), (3), (4), and (6), given the robustness of the parameter estimates in the former cases. This demonstrates the practical importance of outlier robustness in SF modelling. When parameter estimation is not robust, the most influential observations are shown in this case to influence some efficiency predictions by five to ten percentage points in either direction. Furthermore, it is clear that the lowest efficiency predictions tend to be the most heavily influenced. From a regulatory or managerial benchmarking perspective, this sensitivity is clearly not ideal. Our findings suggest that robust alternatives, such as the Student's t model with fixed degrees of freedom, considerably reduce the sensitivity of efficiency predictions to outliers.

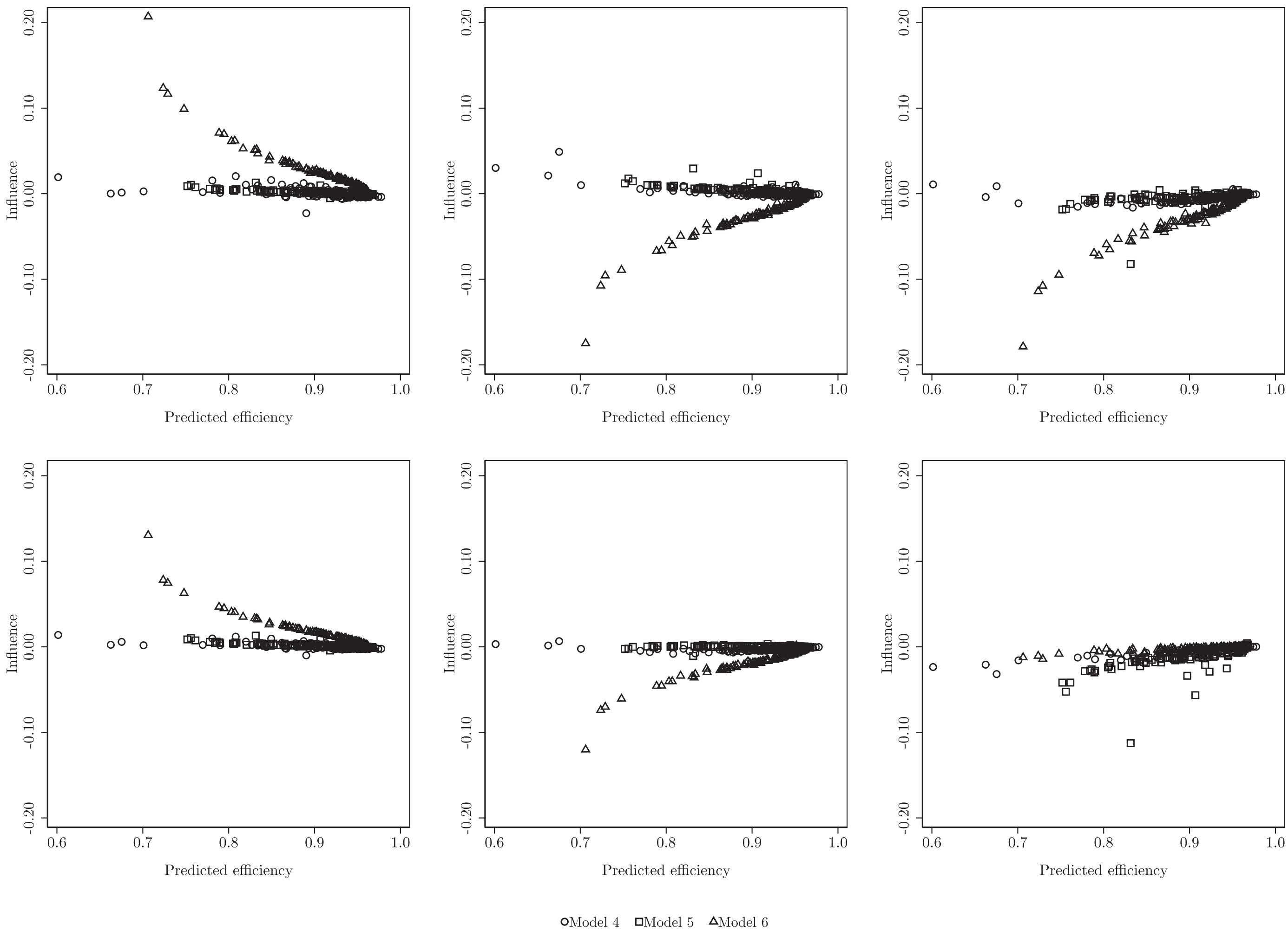

Fig. 6. Influence on efficiency predictions, models (4)–(6).

6. Discussion and conclusions

For managers or regulators applying efficiency analysis techniques to benchmark firms or other decision making units, robustness of efficiency scores to changes in the data is desirable. However, the estimation of standard stochastic frontier models via maximum likelihood is known to be potentially very sensitive to even small changes in the sample. This in turn means that efficiency predictions derived from standard SF models estimated in this way are not robust to outliers and other contaminating observations in the data.

Robustness has received increased attention in the stochastic frontier literature in recent years. Alternative distributional assumptions and estimation methods aimed at robustification have been put forward, but without a direct examination of the robustness properties of either these alternatives or existing models and methods. We fill this gap by considering the robustness of maximum likelihood estimation of stochastic frontier models in terms of the behaviour of the influence function, and deriving conditions under which maximum likelihood estimation can be considered robust. We derive sufficient conditions for robust maximum likelihood estimation of the stochastic frontier model based on relatively simple and easily checked properties of the marginal densities of the error components and, in the case of dependence, the copula density. This provides a convenient method of checking the robustness properties of various specifications without the need to consider each set of distributional assumptions on a case-by-case basis.

We find that most of the canonical distributional assumptions – e.g. normally distributed noise – and recently proposed ‘robust’ alternatives do not satisfy our conditions for robust maximum likelihood estimation. However, the Student’s t distribution satisfies our conditions, subject to some innocuous upper bound being placed on the degrees of freedom parameter, and has a particularly attractive property such that it can be paired with a broad class of inefficiency distributions – including any one parameter scale family such as the half normal or exponential distributions – under independence, and maximum likelihood estimation will remain robust. Finding more flexible inefficiency distributions, or allowing for dependence between error terms, in such a way that our conditions for robust maximum likelihood estimation are satisfied is more challenging.

The use of a Student’s t distribution for noise is therefore a viable approach to robust maximum likelihood estimation of stochastic frontier models in many cases. This proposal has other attractive features. Since the model with normally distributed noise is recovered as the degrees of freedom parameter increases, testing against the standard model is possible (Wheat et al., 2019). Relative to alternative approaches of modifying the loss function, we are able to retain the simplicity and attractive properties of maximum likelihood estimation, and avoid loss of efficiency in estimation and the need to specify a tuning parameter governing the trade-off between robustness and efficiency that comes with such approaches. Alternatively, if we believe that the ‘true’ model before contamination by outliers has normally distributed noise, we can view the Student’s t model as a pseudo-maximum likelihood estimator, analogous to changing the loss function, where the degrees of freedom parameter controls the trade-off between robustness and efficiency (see Stead et al. (Forthcoming) for further discussion). Wheat et al. (2019) provide simulation evidence that the model performs well in terms of recovering the standard model when the data generating process involves normally distributed noise.

We demonstrate the calculation of influence functions in an empirical application, and derive expressions for influence on efficiency predictions. Evidence from our empirical application appears to confirm that the parameter estimates and efficiency predictions from robust specifications are significantly less sensitive to individual observations than those from non-robust specifications. These findings are therefore of practical value to applied researchers, managers, regulators, and others who are concerned with the robust efficiency prediction.

Several possible avenues for future research are apparent. The identification of additional marginal and copula densities consistent with our conditions for robust maximum likelihood estimation, particularly copulas that allow for stronger dependence than the Farlie-Gumbel-Morgenstern – the Student’s t copula may be a natural candidate – and extension to a panel data setting would be of interest. As noted in Section 2, an analysis of the resistance and breakdown points of stochastic frontier models along similar lines would also add to our understanding of the impact of outliers.

Declaration of Competing Interest

The authors report there are no competing interests to declare.

Supplementary material

Appendices: Appendices containing discussion of the properties of particular distributions, and proofs of results. (.pdf file)

Replication materials: This Stata.do file[3] loads the Christensen and Greene (1976) data and the *rfrontier* (version 1.1.0) package[4], and uses these to replicate the analysis in Section 5, reproducing Table 1 and Figs. 1–6.

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2022.12.033.

References

Agrell, P. J., & Bogetoft, P. (2017). Regulatory benchmarking: Models, analyses and applications. *Data Envelopment Analysis Journal, 3*(1–2), 49–91.

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

Almanidis, P., Qian, J., & Sickles, R. C. (2014). Stochastic frontier models with bounded inefficiency. In R. C. Sickles, & W. C. Horrace (Eds.), *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications* (pp. 47–81). New York, NY: Springer. chapter 3

Alvarez, A., Amsler, C., Orea, L., & Schmidt, P. (2006). Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. *Journal of Productivity Analysis, 25*(1), 201–212.

Arslan, O., & Genç, A. İ. (2009). The skew generalized t distribution as the scale mixture of a skew exponential power distribution and its applications in robust estimation. *Statistics, 43*(5), 481–498.

Badunenko, O., & Henderson, D. J. (2021). Production analysis with asymmetric noise. MPRA Paper No. 110888.

[3] https://raw.githubusercontent.com/AlexStead/replication/main/robustMLEofSFM/code/influence.do

[4] See https://github.com/AlexStead/rfrontier for further information and the lastest version.

Banker, R. D., & Chang, H. (2006). The super-efficiency procedure for outlier identification, not for ranking efficient units. *European Journal of Operational Research, 175*(2), 1311–1320.

Basu, A., Harris, I. R., Hjort, N. L., & Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika, 85*(3), 549–559.

Battese, G. E., & Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics, 38*(3), 387–399.

Behr, A. (2010). Quantile regression for robust bank efficiency score estimation. *European Journal of Operational Research, 200*(2), 568–581.

Bellini, T. (2012). Forward search outlier detection in data envelopment analysis. *European Journal of Operational Research, 216*(1), 200–207.

Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *Annals of Statistics, 5*(3), 445–463.

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

Bogetoft, P. (1997). DEA-based yardstick competition: The optimality of best-practice regulation. *Annals of Operations Research, 73*, 277–298.

Bonnano, G., De Giovanni, D., & Domma, F. (2017). The ‘wrong skewness’ problem: A re-specification of stochastic frontiers. *Journal of Productivity Analysis, 47*(1), 49–64.

Çankaya, M. N., & Arslan, O. (2020). On the robustness properties for maximum likelihood estimators of parameters in exponential power and generalized t distributions. *Communications in Statistics - Theory and Methods, 49*(3), 607–630.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444.

Christensen, L. R., & Greene, W. H. (1976). Economies of scale in U.S. electric power generation. *Journal of Political Economy, 84*(4), 655–676.

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

El Mehdi, R., & Hafner, C. M. (1976). Inference in stochastic frontier analysis with dependent error terms. *Mathematics and Computers in Simulation, 102*(Supplement C), 104–116.

Ferrari, D., & Yang, Y. (2010). Maximum L_q-likelihood estimation. *Annals of Statistics, 38*(2), 753–783.

Gómez-Déniz, E., & Pérez-Rodríguez, J. V. (2015). Closed-form solution for a bivariate distribution in stochastic frontier models with dependent error. *Journal of Productivity Analysis, 43*(2), 215–223.

Gupta, A. K., & 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., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. New York, NY: John Wiley.

Haney, A., & Pollitt, M. (2009). Efficiency analysis of energy networks: An international survey of regulators. *Energy Policy, 37*(12), 5814–5830.

Henningsen, A. (2020). Influential observations in stochastic frontier analysis. Working Paper, Konkurrence-og Forbrugstyrelsen.

Horrace, W. C., & Parmeter, C. F. (2018). A Laplace stochastic frontier model. *Econometric Reviews, 37*(3), 260–280.

Horrace, W. C., & Wright, I. A. (2020). Stationary points for parametric stochastic frontier models. *Journal of Business & Economic Statistics, 38*(3), 516–526.

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

Jamasb, T., & Pollitt, M. (2001). Benchmarking and regulation: International electricity experience. *Utilities Policy, 9*(3), 107–130.

Janssens, G. K., & van den Broeck, J. (1993). Outliers, sample size and robust estimation of stochastic frontier production models. *Journal of Information and Optimization Sciences, 14*(3), 257–274.

Jondrow, J., Lovell, C. A. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics, 19*(2), 233–238.

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

Jradi, S., Parmeter, C. F., & Ruggiero, J. (2021). Quantile estimation of stochastic frontiers with the normal-exponential specification. *European Journal of Operational Research, 295*(2), 475–483.

Jradi, S., & Ruggiero, J. (2019). Stochastic data envelopment analysis: A quantile regression approach to estimate the production frontier. *European Journal of Operational Research, 278*(2), 385–393.

Lucas, A. (1997). Robustness of the student t based m-estimator. *Communications in Statistics - Theory and Methods, 26*(5), 1165–1182.

Meeusen, W., & van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review, 18*(2), 435–444.

Miyamura, M., & Kano, Y. (2006). Robust Gaussian graphical modeling. *Journal of Multivariate Analysis, 97*(7), 1525–1550.

Nguyen, N. (2010). *Estimation of technical efficiency in stochastic frontier analysis*. Bowling Green State University Ph.D. thesis..

Ondrich, J., & Ruggiero, J. (2002). Outlier detection in data envelopment analysis: An analysis of jackknifing. *Journal of the Operational Research Society, 53*(3), 342–346.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association, 79*(388), 871–880.

Seaver, B. L., & Triantis, K. P. (1995). The impact of outliers and leverage points for technical efficiency measurement using high breakdown procedures. *Management Science, 41*(6), 937–956.

Shleifer, A. (1985). A theory of yardstick competition. *Rand Journal of Economics, 16*(3), 319–327.

Simar, L. (2003). Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis, 20*(3), 391–424.

Simar, L., & Wilson, P. W. (2010). Inferences from cross-sectional, stochastic frontier models. *Econometric Reviews, 29*(1), 62–98.

Smith, M. D. (2008). Stochastic frontier models with dependent error components. *Econometrics Journal, 11*(1), 172–192.

Song, J., Oh, D., & Kang, J. (2008). Robust estimation in stochastic frontier models. *Computational Statistics and Data Analysis, 105*, 243–267.

Stead, A. D., Wheat, P., & Greene, W. H. (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, & M. Voia (Eds.), *Productivity and Inequality* (pp. 1–19). Cham, Switzerland: Springer. chapter 1

Stead, A. D., Wheat, P., & Greene, W. H. (2019). Distributional forms in stochastic frontier analysis. In T. ten Raa, & W. H. Greene (Eds.), *Palgrave Handbook of Economic Performance Analysis* (pp. 225–274). Cham, Switzerland: Palgrave Macmillan. chapter 8

Stead, A. D., Wheat, P., & Greene, W. H. Forthcoming. Robustness in stochastic frontier analysis. In P. Macedo, V.F. Moutinho, & M. Madaleno (Eds.), *Advanced Mathematical Methods in Economic Efficiency Analysis*. Springer.

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

Thanassoulis, E. (2000). DEA and its use in the regulation of water companies. *European Journal of Operational Research, 127*(1), 1–13.

Tran, N. A., Shively, G., & Preckel, P. (2010). A new method for detecting outliers in data envelopment analysis. *Applied Economics Letters, 17*(4), 313–316.

Tsionas, M. G. (2020). Quantile stochastic frontiers. *European Journal of Operational Research, 282*(3), 1177–1184.

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

Waldman, D. M. (1982). A stationary point for the stochastic frontier likelihood. *Journal of Econometrics, 18*(2), 275-279.

Wang, W. S., & Schmidt, P. (2009). On the distribution of estimated technical efficiency in stochastic frontier models. *Journal of Econometrics, 148*(1), 36–45.

Wei, Z., Zhu, X., & Wang, T. (2021). The extended skew-normal-based stochastic frontier model with a solution to 'wrong skewness' problem. *Statistics, 55*(6), 1387–1406.

Wheat, P., Stead, A. D., & Greene, W. H. (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.

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

Zhuo, S. (2018). Local influence analysis of stochastic frontier estimation: A case-weights perturbation approach. *Economics Letters, 164*, 79–81.

Zulkarnain, R., & Indahwati, I. (2019). 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.