CrossMark

Robust stochastic frontier analysis: a Student's *t*-half normal model with application to highway maintenance costs in England

Phill Wheat [1] · Alexander D. Stead [1] · William H. Greene[2]

Abstract
The presence of outliers in the data has implications for stochastic frontier analysis, and indeed any performance analysis methodology, because they may lead to imprecise parameter estimates and, crucially, lead to an exaggerated spread of efficiency predictions. In this paper we replace the normal distribution for the noise term in the standard stochastic frontier model with a Student's *t* distribution, which generalises the normal distribution by adding a shape parameter governing the degree of kurtosis. This has the advantages of introducing flexibility in the heaviness of the tails, which can be determined by the data, as well as containing the normal distribution as a limiting case, and we outline how to test against the standard model. Monte Carlo simulation results for the maximum simulated likelihood estimator confirm that the model recovers appropriate frontier and distributional parameter estimates under various values of the true shape parameter. The simulation results also indicate the influence of a phenomenon we term 'wrong kurtosis' in the case of small samples, which is analogous to the issue of 'wrong skewness' previously identified in the literature. We apply a Student's *t*-half normal cost frontier to data for highways authorities in England, and this formulation is found to be preferred by statistical testing to the comparator normal-half normal cost frontier model. The model yields a significantly narrower range of efficiency predictions, which are non-monotonic at the tails of the residual distribution.

Keywords Stochastic frontier analysis · Robust analysis · Student's *t* · Outliers · Wrong Kurtosis

JEL Codes C12 · C13 · C44 · C46 · D24 · R42 · R53

1 Introduction

Frontier analysis is concerned with the measurement of efficiency relative to an estimated production or cost frontier. The presence of noise in the sample is potentially problematic in two ways: it affects the position of each decision making unit (DMU) relative to the frontier, and it affects the precision of the estimates of the shape of the frontier itself. The magnitudes of these effects vary from one method to another: deterministic methods, such as data envelopment analysis (DEA) (Charnes et al. 1978) and corrected ordinary least squares (COLS) are particularly sensitive, given that they make no allowance for noise. In contrast, Stochastic Frontier (SF) Analysis explicitly controls for noise, mitigating its impact on the estimated frontier and on individual efficiency scores.

The range of efficiency scores may still be very large in the presence of data with outlying observations. In terms of empirical motivation for this paper, this paper is in response to the authors' finding of an implausibly wide range of efficiency scores in our work studying cost drivers and cost efficiency in the highways maintenance operations of local authorities in England, which utilises detailed data on operating and capital expenditure provided by each authority. This can be narrowly explained as being due to a combination of under-reporting and over-reporting of expenditure, unobserved investment cycle effects, and extreme weather events. However we

✉ Phill Wheat
  p.e.wheat@its.leeds.ac.uk

1 Institute for Transport Studies, University of Leeds, Leeds LS2 9LJ, UK

2 Department of Economics, Stern School of Business, New York University, New York, NY 10012, USA

also came across this issue in a number of other datasets, across numerous sectors including hotels in Taiwan (Chen 2007), container ports (Cullinane et al. 2006), regional iron and steel production in China (Lin and Wang 2014), crop farms in Poland (Latruffe et al. 2004), French insurance companies (Fecher et al. 1993), Belgian municipalities (De Borger and Kerstens 1996), US banks (Bauer et al. 1998), and police forces in England and Wales (Drake and Simper 2003). In general, as observed by Gupta and Nguyen (2010), financial data are often heavy tailed. For example Berger and Humphrey (1991) observe heavy tailed distribution of costs in banking, an industry to which efficiency analysis is often applied. As such the issue in this paper has broad applicability across the performance literature.

In this paper we outline and discuss the merits and empirical application of a new stochastic frontier model which accommodates the influence of outlying observations. This is the stochastic frontier model with a Student's *t* distribution for the noise term. The advantages of this model over previous proposals lie in its flexibility, since the degree of kurtosis is no longer fixed but allowed to vary with the degrees of freedom parameter. In fact, the Student's *t* distribution nests (or more precisely, contains as a limiting case) the normal distribution as the degrees of freedom parameter approaches infinity. For any given distribution of u_i, our model encompasses the standard SF model. This enables testing down against the standard model, in contrast to previous proposals which utilise non-nested specifications, and lets the data determine the extent to which outlying observations influence the kurtosis of the noise error term. Thus our model is an original and significant contribution to the literature, not just in being able to better accommodate outlying observations in efficiency analysis relative to the standard SF model, but it is the first contribution to contain as a testable limiting case the standard SF model. As such our model provides a natural extension to the tools of practitioners in the field.

The structure of this paper is as follows: Section 2 reviews existing methods available in the literature to handle a large number of outliers in frontier analysis. Section 3 introduces a *t*-truncated normal stochastic frontier model for dealing with heavy-tailed noise and discusses efficiency prediction. Testing of the normal distribution hypothesis is considered in Section 4. Section 5 presents Monte Carlo evidence on the performance of the maximum simulated likelihood estimator of the model. Section 6 applies the Student's *t*-half normal model to data on highways maintenance costs in England and compares the results to those obtained from normal-half normal model, and Section 7 gives our summary and conclusions. Some technical results appear in the Appendices.

2 Approaches to outliers in stochastic frontier analysis

The standard SF model, as developed by Aigner et al. (1977) and Meeusen and van Den Broeck (1977), is as follows:

$$y_i = f(x_i, \beta) + \varepsilon_i, \quad \varepsilon_i = v_i - su_i \tag{1}$$

Where the subscript i denotes the observation, y_i is the dependent variable, x_i is a vector of independent variables, and ε_i is an error term with two components. The noise component, v_i, follows a symmetric distribution centred at zero, and the inefficiency component, u_i, is drawn from a one-sided distribution. In the production case $s = 1$, while in the cost case $s = -1$. Many alternative distributions have been proposed for u, for example the half normal or exponential (Aigner et al. 1977), truncated normal (Stevenson 1980), or gamma (Greene 2003) distributions. In comparison, v_i is almost always assumed to follow a normal distribution, although both distributional assumptions are crucial with regards to the both the robustness of the parameter estimates to outliers and the decomposition of the estimated residual into noise and inefficiency.

Concerning robustness to outliers, it is worth pointing out that one of the original motivations behind SF analysis as an alternative to ordinary least squares (OLS)—which after all yields unbiased estimates of the frontier parameters apart from the intercept—was to obtain estimates that are more robust to skewness of the residuals implied by inefficiency. It was not until the Jondrow et al. (1982) paper that a method of obtaining observation-specific predictions of efficiency was introduced. On the decomposition of the residuals, obtaining these observation-specific efficiency scores proceeds by making some prediction of u_i based on the conditional distribution of $u_i|\varepsilon_i$ Dropping the subscript i from here on, this is given by

$$f_{u|\varepsilon}(u|\varepsilon) = \frac{f_v(\varepsilon + su)f_u(u)}{f_\varepsilon(\varepsilon)} \tag{2}$$

where f_v and f_u are the probability density functions of v and u respectively, and f_ε is the probability density function of the composed error derived as the convolution of the two error components. The usual approach to efficiency prediction is to use the mean of this distribution according to $\exp[-E(u|\varepsilon)]$ (Jondrow et al. 1982) or $\exp[E(-u|\varepsilon)]$ (Battese and Coelli 1988). Clearly, any efficiency prediction derived in this way will depend upon the distribution not only of u but also of v: Wang and Schmidt (2009) derive the distribution of $E(u|\varepsilon)$ in the normal-half normal case, and show that $E(u|\varepsilon) \rightarrow u$ only as $VAR(v) \rightarrow 0$, while $E(u|\varepsilon) \rightarrow E(u)$ as $VAR(v) \rightarrow \infty$. The conditional mean predictor is therefore a shrinkage of u towards its mean with the degree of shrinkage depending upon the distribution of v.

Given that the presence of outliers, where these are attributed to noise rather than inefficiency, implies a leptokurtic[1] noise distribution, the normal distribution usually assumed for v is inadequate since it is mesokurtic for any given parameter values. Intuitively, we may therefore expect outliers in the data to result in an exaggerated spread of efficiency scores for two reasons: first, because of an inflated estimate of the scale of the distribution of u, and second because of insufficient shrinkage of u towards its mean, especially at the extremes. That is to say, if leptokurtosis in the noise term due to outliers is not taken into account, residuals will be attributed disproportionately to inefficiency rather than noise, particularly in outlying observations. This motivates the development of alternative SF models that can accommodate outliers.

A review of the literature in this area provided by Stead et al. (2018) covers several different approaches to outliers: the use of alternative efficiency predictors, thick frontier analysis, heteroskedastic SF models, and the use of alternative noise distributions. Other possible approaches include right truncation of the inefficiency distribution to restrict the range of possible efficiency predictions, and detecting and removing outliers.

The detection and removal (or reweighting) of outliers is a common approach to dealing with outliers in regression analysis, with outliers identified on the basis of the extent to which an observation has a disproportionate effect on parameter estimates, captured by Cook's distance (Cook 1977), Mahalanobis distance (Mahalanobis 1936), and similar measures. However, existing methods may not be appropriate in the case of skewness of the composed error. For example, Cook's Distance explicitly assumes normally distributed errors, while in SF analysis the composed error has a skew normal distribution (Azzalini 1985), and in the normal-exponential case it follows what is known as an exponentially modified Gaussian distribution (Grushka 1972). One approach to outliers could be to derive measures of influence and leverage appropriate for SF models, although these would depend upon the particular specification used.

Finally, we can account for outliers by assuming an appropriate distribution for v. In principle, any distribution that is symmetric, centred around zero and unimodal is an appropriate candidate for the distribution of v. Although far more attention has been paid in the literature to the distribution of u, several such alternatives have been suggested. Outside of the SF literature, Lange et al. (1989) suggest the use of the Student's t distribution for the error

terms as a robust alternative to OLS. Lange and Sinsheimer (1993) discuss estimation when errors are drawn from the logistic, slash, t, and contaminated normal distributions. Note that the latter is simply the mixture of two normal distributions with the same mean but differing variances. All of these distributions have heavier tails than the normal distribution, and therefore offer greater robustness to outliers. Reviews of the SF literature by Greene (2008) and Parmeter and Kumbhakar (2014) include some discussion of alternative noise distributions.

In the context of SF analysis, Tancredi (2002) proposes a model in which v is drawn from a Student's t distribution and u from a half t distribution. Tchumtchoua and Dey (2007) study the Student's t-half t model in a Bayesian setting. Griffin and Steel (2007) and Hajargasht and Griffiths (2018) also discuss estimation of Bayesian SF models with Student's t noise using Markov Chain Monte Carlo (MCMC) and variational Bayes methods, respectively. Nguyen (2010) introduces two additional alternative heavy tailed distributions for v: the Laplace distribution and the Cauchy distribution – note that the latter is a Student's t distribution with degrees of freedom equal to one – and derive formulae for Cauchy-truncated Cauchy and Cauchy-half Cauchy SF models. Applications are shown in Gupta and Nguyen (2010). Horrace and Parmeter (2018) study Laplace-truncated Laplace and Laplace-exponential[2] SF models. Noting that the Laplace distribution is ordinary smooth, in contrast to the normal distribution, which is supersmooth, they conjecture that the Laplace distribution may be advantageous with respect to the deconvolution of ε into v and u, since optimal convergence rates for deconvolution problems decrease with the smoothness of the noise distribution, being in particular much slower when the noise distribution is supersmooth rather than ordinary smooth (Fan 1991; 1992).

The performance of these models in the presence of outliers is interesting, but under-explored. Following the previous discussion, given the use of heavy-tailed distributions for v, we might expect these models to be more robust to outliers in terms of parameter estimation, and in terms of yielding less extreme efficiency predictions for outlying observations. Indeed, Tancredi (2002) shows that as $s\varepsilon_i \to \infty$, $f_{u|\varepsilon}(u|\varepsilon)$ becomes completely flat in the Student's t-half t model, in contrast to how $f_{u|\varepsilon}(u|\varepsilon)$ becomes a degenerate distribution at zero as $s\varepsilon \to \infty$ in the normal-half normal model. Similarly, Horrace and Parmeter (2018) show that $f_{u|\varepsilon}(u|\varepsilon)$ – and hence $E(u|\varepsilon)$ – is constant for positive values of $s\varepsilon$ in the Laplace-exponential case. Focusing more explicitly on outliers, Stead et al. (2018) propose a model in which v follows a logistic distribution and u follows a half normal

---

[1] The terms leptokurtic, mesokurtic, and platykurtic are used to denote distributions with positive, zero, and negative excess kurtosis. That is, kurtosis greater than, equal to, or less than that of the normal distribution.

[2] Note that the left truncation of a Laplace distribution at or above zero results in an exponential distribution.

distribution, and show that the model results in a smaller estimate for $VAR(u)$ and yields a considerably narrower spread of efficiency scores than the normal-half normal model, with little change in $\exp[E(-u|\varepsilon)]$ for large $|s\varepsilon|$. Each of these cases suggest that the choice of a heavy tailed distribution for v significantly affects the prediction of efficiency, and the uncertainty in prediction, at one or both tails, producing less extreme efficiency predictions.

3 The Student's *t*-truncated normal SF model

In this section, we propose a robust SF model in which v follows a Student's t distribution and u follows a truncated normal distribution. The Student's t distribution is for v has a number of attractive properties in the present context. First, it is a heavy tailed distribution, and thus more robust to the presence of outliers. Second, it avoids the arbitrary assumptions on the degree of kurtosis in v embedded in existing models—the normal, logistic, and Laplace distributions have excess kurtosis of 0, 1.2, and 3, respectively, regardless of parameter values. Ultimately, the kurtosis of v is an empirical question, and therefore we should ideally use a distribution for v for which kurtosis is flexible. The kurtosis of the Student's t distribution depends upon the degrees of freedom parameter. Third, the Student's t distribution nests the Cauchy distribution when the degrees of freedom parameter is equal to one and the normal distribution as it approaches infinity. Therefore an SF model with a Student's t distribution for v encompasses a model in which v follows a Cauchy or normal distribution for any given distribution of u. This enables testing against these alternatives. In the latter case, we are testing against the standard SF model, which could be interpreted as a test of robustness to outliers.

Below, we derive simulated log likelihood functions and efficiency predictors for the Student's t-truncated normal SF model, and discuss estimation and hypothesis testing. Results for the Student's t-half normal and Student's t-exponential models can be obtained via some simple modifications. Extensions to other distributions of u are straightforward if the quantile function of that distribution has a closed form, while in many other cases—as with the Student's t-gamma—the simulated log likelihood function becomes slightly more complex.

3.1 Formulation and estimation

In SF models, the error ε (noting again that we drop the subscript i for notational simplicity) is composed of a symmetric noise component v and an inefficiency component u which is drawn from a one-sided distribution:

$$\varepsilon = v - su \tag{3}$$

In this study, we assume that v is drawn from a non-standardized Student's t distribution—which includes a scale parameter σ_v—and that u follows a truncated normal distribution, so that the probability density functions f_v and f_u are given by

$$f_v(v) = \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\sqrt{\pi a}\sigma_v}\left[1 + \frac{1}{a}\left(\frac{v}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}} \tag{4}$$

$$f_u(u) = \begin{cases} \dfrac{\frac{1}{\sigma_u}\phi\left(\frac{u-\mu}{\sigma_u}\right)}{\Phi\left(\frac{\mu}{\sigma_u}\right)}, & u >; 0 \\ 0, & u \leq 0 \end{cases} \tag{5}$$

where μ and σ_u are the mean and standard deviation of the pre-truncation distribution of u, respectively, a is a shape parameter that determines the kurtosis of the Student's t distribution, and Γ is the gamma function. If μ is set to zero, then the model reduces to a Student's t-half normal model as applied to our data in Section 6. As noted previously, as $a \to \infty$ the Student's t distribution approaches the normal distribution. Thus, our model encompasses as a limiting case the normal-truncated normal model. Similarly, when $a = 1$ we have a Cauchy distributed noise component.

The joint density of ε and u is given by

$$f_{u,\varepsilon}(u, \varepsilon) = \begin{cases} \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\sqrt{\pi a}\sigma_v}\left[1 + \frac{1}{a}\left(\frac{\varepsilon+su}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}}\frac{\frac{1}{\sigma_u}\phi\left(\frac{u-\mu}{\sigma_u}\right)}{\Phi\left(\frac{\mu}{\sigma_u}\right)}, & u\,gt;0 \\ 0, & u \leq 0 \end{cases} \tag{6}$$

and the marginal density of ε is given by the convolution

$$f_\varepsilon(\varepsilon) = \int_0^\infty \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\sqrt{\pi a}\sigma_v}\left[1 + \frac{1}{a}\left(\frac{\varepsilon+su}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}}\frac{\frac{1}{\sigma_u}\phi\left(\frac{u-\mu}{\sigma_u}\right)}{\Phi\left(\frac{\mu}{\sigma_u}\right)}du \tag{7}$$

which lacks a convenient closed form, hindering ML estimation. As an alternative, we may use simulation to approximate the integral and arrive at a simulated log likelihood function—see Train (2009) for an introduction to maximum simulated likelihood—as proposed by Greene (2003) for the normal-gamma model and Stead et al. (2018) for the logistic-half normal model. We begin by noting that (7) is the expectation of $f_v(\varepsilon + su)$, where u is drawn from a truncated-normal distribution,

$$h(u) = E[f_v(\varepsilon + su)|u \geq 0], \quad u \sim N(\mu, \sigma_u) \tag{8}$$

that can be estimated by

$$\widehat{h} = \frac{1}{Q}\sum_{q=1}^{Q} f_v(\varepsilon + su_q) \qquad (9)$$

where u_q is a draw from a truncated normal distribution[3]. This gives us the simulated probability density function for ε

$$\widehat{f_\varepsilon(\varepsilon)} = \frac{1}{Q}\frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\sqrt{\pi a}\sigma_v}\sum_{q=1}^{Q}\left[1 + \frac{1}{a}\left(\frac{\varepsilon + su_q}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}} \qquad (10)$$

and the simulated log-likelihood function is

$$\ln SL = -N\ln Q + N\ln\left[\Gamma\left(\frac{a+1}{2}\right)\right] - N\ln\left[\Gamma\left(\frac{a}{2}\right)\right]$$
$$-\frac{N}{2}(\ln\pi + \ln a) - N\ln\sigma_v + \sum_{i=1}^{N}\ln\sum_{q=1}^{Q}\left[1 + \frac{1}{a}\left(\frac{\varepsilon + su_q}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}} \qquad (11)$$

This may be maximised like any conventional log-likelihood function, provided we have our draws from the truncated normal distribution. From Geweke et al. (1997) and Greene (2003), we have the quantile function of a truncated normal random variable:

$$u_q = \mu + \sigma_u\Phi^{-1}\left\{\left[\Phi\left(\frac{t_r - \mu}{\sigma_u}\right) - \Phi\left(\frac{t_l - \mu}{\sigma_u}\right)\right]F_q + \Phi\left(\frac{t_l - \mu}{\sigma_u}\right)\right\} \qquad (12)$$

Where t_l and t_r are the left and right truncation points, respectively, and F_q is draw q from the standard uniform distribution. Since we know that $t_l = 0$, $t_r = \infty$, this simplifies to

$$u_q = \mu + \sigma_u\Phi^{-1}\left[1 + \Phi\left(\frac{\mu}{\sigma_u}\right)(F_q - 1)\right] \qquad (13)$$

In order to modify the model so that the one-sided error follows some other distribution, we need only change u_q such that we instead obtain draws from the chosen distribution. The most obvious choices are the exponential and gamma distributions. For the t-exponential case, we have the quantile function

$$u_q = -\sigma_u\ln F_q \qquad (14)$$

And by substituting (13) or (14) into (11), we have the simulated log-likelihood function for the Student's t-truncated-normal and Student's t-exponential models respectively. Other proposed distributions for u, such as the Weibull (Tsionas 2007) and Rayleigh (Hajargasht 2015) distributions also have closed form quantile functions. However, even in cases in which the quantile distribution has no analytical expression, such as the gamma distribution, forming the simulated log-likelihood is possible—see Greene (2003).

Proceeding with the Student's t-truncated normal variant of the model, (11) and (13) give the simulated log-likelihood function. First order conditions for maximisation are given in Appendix B. One remaining issue is the method of taking random draws: we prefer to use Halton draws, which aim for good coverage of the unit interval rather than randomness: this significantly reduces the number of draws needed to approximate the integral (see Greene (2003) for a fuller discussion).

Recalling the discussion of Horrace and Parmeter (2018) around convergence rates and the smoothness property of the noise distribution in deconvolution problems, a further attraction of the Student's t distribution is the potential advantage in terms of estimating f_u. Fan (1991) discusses the smoothness properties of several distributions. The normal distribution is supersmooth of order 2, while the Cauchy distribution is supersmooth of order 1. The smoothness of the Student's t distribution is not explicitly discussed, but may be determined as follows. From Fan (1991), the distribution of a random variable is said to be supersmooth of order β if its characteristic function, $\varphi(t)$, satisfies

$$d_0|t|^{\beta_0}\exp\left(-\frac{|t|^\beta}{\gamma}\right) \leq \varphi(t) \leq d_1|t|^{\beta_1}\exp\left(-\frac{|t|^\beta}{\gamma}\right), \quad t \to \infty \qquad (15)$$

where d_0, d_1, β, and γ are positive constants, and β_0 and β_1 are constants. Expression 4.8 in Hurst (1995) gives the following result

$$\varphi_v(t) \cong \frac{\sqrt{\pi}}{\Gamma\left(\frac{a}{2}\right)2^{\frac{a-1}{2}}}\left(\sqrt{a}|t|\right)^{\frac{a-1}{2}}\exp\left(-\sqrt{a}|t|\right), \quad t \to \infty \qquad (16)$$

where $\varphi_v(t)$ denotes the characteristic function of a Student's t distribution with degrees of freedom parameter a. Substituting (16) into (15), we see that the Student's t distribution is supersmooth of order 1 when a is finite—which encompasses the Cauchy case where $a = 1$. From the results of Fan (1991), this implies that optimal convergence rates are faster when noise is Student's t distributed than when noise is normally distributed. It should be noted however that ordinary smooth distributions, such as the

---

[3] An alternative approach is to note that the Student's t distribution is a scale mixture of normal distributions where the mixing distribution is inverse gamma distributed. To estimate a Student's t-half normal SF model, we could approximate a scale mixture of skew normal distributions with draws from an inverse gamma distribution. It is unclear if this approach has any advantage over that used in this paper, however it is slightly less convenient in terms of implementation, particularly when it comes to changing the distribution of u.

Laplace distribution, have convergence rates that are faster still, being polynomial functions of N, as opposed to logarithmic functions of N as in the case supersmooth distributions (Fan 1992). For more discussion of the differences between smooth and supersmooth noise distributions in stochastic frontier analysis, see Horrace and Parmeter (2011; 2018).

3.2 Parameter identification

To demonstrate that the model parameters are identified, we show that consistent method of moments estimators exist. For the sake of tractability, we restrict attention to the case where $\mu = 0$. As in the standard SF model, the OLS estimator $\widehat{\beta}$ is a consistent estimate of β, with the exception that $\widehat{\beta}_0$ yields a consistent estimator of $\beta_0 - sE(u)$. Since v and u are independent, the first four moments of ε are given by

$$E(\varepsilon) = E(v) - sE(u) = -s\sigma_u\sqrt{2/\pi}, \quad a >; 1 \tag{17}$$

$$E(\varepsilon^2) = E(v^2) + E(u^2) = \sigma_v^2\frac{a}{a-2} + \sigma_u^2\frac{\pi-2}{\pi}, \quad a > 2 \tag{18}$$

$$E(\varepsilon^3) = E(v^3) - sE(u^3) = -s\sqrt{2/\pi}\frac{(4-\pi)}{\pi}\sigma_u^3, \quad a > 3 \tag{19}$$

$$E(\varepsilon^4) = E(v^4) + E(u^4) + 6E(v^2)E(u^2) = \\ \frac{6}{a-4} + 3 + \frac{3\pi^2 - 4\pi - 12}{\pi^2}\sigma_u^4 + 6\sigma_v^2\frac{a}{a-2}\sigma_u^2\frac{\pi-2}{\pi}, \quad a > 4 \tag{20}$$

Setting these equal to the corresponding raw sample moments μ_2, μ_3, and μ_4 then rearranging gives the following method of moments estimators—note the correction to the estimated intercept.

$$\widetilde{\sigma}_u = \left[-s\mu_3\sqrt{\pi/2}\frac{\pi}{(4-\pi)}\right]^{\frac{1}{3}}, \quad a > 3 \tag{21}$$

$$\widetilde{a} = \frac{6}{\mu_4 - 3 - \widetilde{\sigma}_u^4\frac{20\pi - 3\pi^2 - 36}{\pi^2} - 6\mu_2\widetilde{\sigma}_u^2\frac{\pi-2}{\pi}} + 4, \quad a > 4 \tag{22}$$

$$\widetilde{\sigma}_v = \sqrt{\left(\mu_2 - \widetilde{\sigma}_u^2\frac{\pi-2}{\pi}\right)\frac{\widetilde{a}-2}{\widetilde{a}}}, \quad a > 2 \tag{23}$$

$$\widetilde{\beta}_0 = \widehat{\beta}_0 - s\widetilde{\sigma}_u\sqrt{2/\pi}, \quad \widetilde{\beta}_j = \widehat{\beta}_j, \quad j = 1, \dots, k, \quad a > 1 \tag{24}$$

Note that since the moments of the Student's t distribution are variously undefined or infinite for small a, this is also the case for moments of ε, so these estimators exist only when a is sufficiently large. Further, just as the method

of moments approach to estimation of the standard SF model breaks down when $\mu_3 \leq 0$, μ_4 must be sufficiently large that the denominator in (22) is positive for method of moments to be used in this case.[4] In Section 5, we undertake Monte Carlo simulations to analyse the performance of the model for a variety of values of $\alpha = 2$, $a = 4$, $a = 5$, $a = 10$, and $a \to \infty$, the latter corresponding to $v \sim N(0,1)$.

Well-known results from Waldman (1982) show that the OLS estimator is a stable stationary point of the log-likelihood function for the standard model, and that the identification of the σ_u parameter hinges on the skewness of the OLS residuals. Horrace and Parmeter (2018) show that, while these results do not apply to the SF model with Laplace noise since OLS is not the limiting estimator as Var$(u) \to 0$, an analogous result applies: the least absolute deviations (LAD) estimator is a stationary point of the log-likelihood[5]. It is trivial to show that a similar result applies to our model, i.e. that the ML estimator of a regression model with Student's t errors is a stationary point in our log-likelihood function. Horrace and Wright (forthcoming) show that such a stationary point exists under very weak distributional assumptions about v and u. This suggests that identification of σ_u in our model depends upon the skewness of the residuals from the Student's t regression model. However, while Horrace and Parmeter (2018) point out that in the Laplace case the LAD stationary point is not stable due to non-differentiability of the log-likelihood function in the limiting case, this does not appear to apply in the Student's t case, in which the log-likelihood function is everywhere differentiable.

3.3 Efficiency prediction

As discussed in previous sections, the usual approach to generating observation-specific efficiency scores is to predict values based on the distribution of $u|\varepsilon$, which is given by:

$$f_{u|\varepsilon}(u|\varepsilon) = \frac{f_v(\varepsilon + su)f_u(u)}{f_\varepsilon(\varepsilon)} \tag{25}$$

The most widely used predictors are the mean of this conditional distribution according to $\exp[-E(u|\varepsilon)]$:

$$\exp[-E(u|\varepsilon)] = \exp\left[-\int_0^\infty u\frac{f_v(\varepsilon + su)f_u(u)}{f_\varepsilon(\varepsilon)}du\right] \tag{26}$$

---

[4] Note also that the estimator (21) is substituted into (22) and (23), and (22) into (23), in place of the unknown true values. However, if $2 < a \leq 4$, the method of moments estimator for $\widetilde{\sigma}_v$ exists but depends on a, for which the method of moments estimator does not exist. An alternative approach could be to treat a as a fixed tuning parameter, and substitute its value into (23) for $\widetilde{\sigma}_v$.

[5] It is well known that, just as OLS is the ML estimator for a regression model with normal errors, LAD is the ML estimator for a regression model with Laplace errors.

Or, $E[\exp(-u|\varepsilon)]$:

$$E[\exp(-u|\varepsilon)] = \int_0^\infty \exp(-u)\frac{f_v(\varepsilon + su)f_u(u)}{f_\varepsilon(\varepsilon)}du \qquad (27)$$

Note that in both cases, since $f_\varepsilon(\varepsilon)$ is not a function of u, it can be moved outside the integral. Bearing in mind that $f_\varepsilon(\varepsilon)$ is the convolution of $f_v(\varepsilon+su)$ and $f_u(u)$, we therefore have

$$\exp[-E(u|\varepsilon)] = \exp\left[-\frac{\int_0^\infty uf_v(\varepsilon + su)f_u(u)du}{\int_0^\infty f_v(\varepsilon + su)f_u(u)du}\right] \qquad (28)$$

And

$$E[\exp(-u|\varepsilon)] = \frac{\int_0^\infty \exp(-u)f_v(\varepsilon + su)f_u(u)du}{\int_0^\infty f_v(\varepsilon + su)f_u(u)du} \qquad (29)$$

For the model we are considering, none of these integrals have closed form solutions, so we approximate them via simulation. The simulated integral in the denominators of both formulae are given by (10), and the integrals in the numerators are the expectations of $uf_v(\varepsilon+su)$ and $\exp(-u)f_v(\varepsilon+su)$ given that u is a random variable with the probability density function $f_u(u)$. We therefore have, with some simplifying and rearranging:

$$\exp\left[-\widehat{E}(u|\varepsilon)\right] = \exp\left\{-\frac{\sum_{q=1}^Q u_q\left[1 + \frac{1}{a}\left(\frac{\varepsilon+su_q}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}}}{\sum_{q=1}^Q \left[1 + \frac{1}{a}\left(\frac{\varepsilon+su_q}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}}}\right\} \qquad (30)$$

$$\widehat{E}[\exp(-u|\varepsilon)] = \frac{\sum_{q=1}^Q \exp(-u_q)\left[1 + \frac{1}{a}\left(\frac{\varepsilon+su_q}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}}}{\sum_{q=1}^Q \left[1 + \frac{1}{a}\left(\frac{\varepsilon+su_q}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}}} \qquad (31)$$

Where u_q is given by (13) in the Student's t-truncated normal and (14) in the Student's t-exponential models, for example.

4 Hypothesis testing concerning the noise distribution

As discussed previously, an attraction of the Student's t distribution in the current context is that the normal distribution is the limiting case as $a \to \infty$. Therefore a stochastic frontier model in which v follows a Student's t distribution encompasses a model in which v follows a normal distribution. This allows us to test down from a

model with Student's t noise to a standard SF model, which could be interpreted as a testing for heavy tails—or the significance of outliers in the data—and used for model selection. For this purpose, the likelihood ratio test statistic is an obvious choice. This is defined as

$$LR = -2(\ln L_A - \ln L_0) \qquad (32)$$

Where $\ln L_A$ is the simulated log-likelihood from the Student's t model and $\ln L_0$ is the log-likelihood from the null model with normally distributed v. The standard result that this statistic has a limiting χ^2 distribution with degrees of freedom equal to the difference in dimensionality between the alternative and null models does not apply, since under the null hypothesis the degrees of freedom parameter a lies on the boundary of the parameter space. Case 5 in Self and Liang (1987)—see also Case 2 in Chen and Liang (2010)—shows that the likelihood ratio statistic follows an asymptotic 50:50 mixture of χ_0^2 and χ_1^2 distributions, denoted $\chi_{1:0}^2$. Economou (2011) applies this result to an analogous problem in survival analysis of testing down from a three parameter Burr XII distribution to a two parameter Weibull distribution. A further analogue in stochastic frontier analysis is testing for the presence of an upper bound on u, since an SF model with a tail truncated distribution for u nests the standard SF model as the tail truncation point $B \to \infty$.

We present simulation evidence regarding the distribution of the LR, under the null hypothesis that v is normally distributed, in Appendix C. We do this to verify that the results of Chen and Liang (2010) do apply in this case. Note that we could reparameterise the model to include an inverse degrees of freedom parameter, $1/a$, in which case the standard model is the limiting case as $1/a \to 0$; however, this boundary remains an open boundary. In particular, we note that in this setting, and the two other examples appearing in the literature discussed above (Weibull and upper truncation point in the inefficiency distribution), the null hypothesis corresponds to a parameter value at an open (and not closed) boundary. Chen and Liang (2010) state that their result still holds when the boundary is an open boundary, but we wish to verify this. The evidence presented in Appendix B lends support to the idea that $LR \sim \chi_{1:0}^2$ under the null hypothesis. We therefore consider this test to be appropriate for this purpose.

5 Monte Carlo simulations

In this section, we examine the performance of the Student's t-half normal model under various data generating processes (DGPs). We are primarily interested in how the model performs in two different scenarios: when v is Student's t distributed, and when v is normally distributed. In

Table 1 Normal-half normal vs. Student's t-half normal model under various data generating processes—1000 replications

		N = 100						N = 200					
		Normal-half normal			Student's t-half normal			Normal-half normal			Student's t-half normal		
DGP 1		Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.
β_0	1	0.765	1.429	1.641	1.241	1.427	0.589	0.824	1.528	1.876	1.237	1.239	0.514
β_1	1	1.001	0.983	0.547	0.991	0.993	0.166	0.974	0.991	0.547	0.994	0.998	0.116
σ_v	1	2.243	1.974	2.209	0.995	1.032	0.295	2.537	2.209	2.207	1.025	1.053	0.219
σ_u	1	1.326	0.004	2.948	0.847	0.433	4.397	1.211	0.004	2.795	0.922	0.718	4.936
a	2	∞	∞	–	4.1E+04	2.073	7.0E+05	∞	∞	–	2.233	2.118	0.689
DGP 2		Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.
β_0	1	1.132	1.057	0.636	1.246	1.484	0.583	1.156	0.983	0.577	1.260	1.377	0.522
β_1	1	0.996	0.999	0.163	0.997	1.000	0.146	0.997	1.001	0.110	0.998	1.000	0.098
σ_v	1	1.301	1.316	0.307	1.008	1.056	0.262	1.339	1.353	0.242	1.032	1.066	0.207
σ_u	1	0.831	0.990	0.839	0.706	0.333	0.755	0.810	1.057	0.771	0.687	0.578	0.674
a	4	∞	∞	–	3.3E+07	4.545	8.3E+08	∞	∞	–	8.4E+05	4.473	2.6E+07
DGP 3		Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.
β_0	1	1.143	1.010	0.603	1.221	1.383	0.595	1.174	0.985	0.534	1.229	1.197	0.527
β_1	1	1.005	1.008	0.150	1.004	1.004	0.137	1.006	1.008	0.100	1.004	1.004	0.094
σ_v	1	1.203	1.218	0.265	0.982	1.027	0.263	1.247	1.250	0.194	1.023	1.057	0.205
σ_u	1	0.829	1.010	0.779	0.749	0.580	0.759	0.784	1.030	0.700	0.723	0.764	0.680
a	5	∞	∞	–	2.8E+08	5.759	4.6E+09	∞	∞	–	3.0E+07	5.652	8.3E+08
DGP 4		Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.
β_0	1	1.120	0.987	0.519	1.158	1.078	0.571	1.155	1.005	0.464	1.166	1.016	0.509
β_1	1	0.999	1.001	0.121	1.001	1.000	0.122	1.002	1.003	0.092	1.003	1.004	0.092
σ_v	1	1.060	1.082	0.215	0.936	0.985	0.249	1.096	1.095	0.161	0.983	1.001	0.191
σ_u	1	0.856	1.050	0.660	0.821	0.957	0.730	0.814	1.017	0.592	0.808	1.027	0.652
a	10	∞	∞	–	2.9E+08	10.983	2.6E+09	∞	∞	–	2.1E+08	11.015	2.4E+09
DGP 5		Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.
β_0	1	1.075	0.959	0.478	1.067	0.897	0.546	1.078	0.955	0.404	1.041	0.892	0.453
β_1	1	1.000	0.997	0.124	1.000	0.996	0.127	1.003	1.001	0.084	1.004	1.001	0.085
σ_v	1	0.936	0.944	0.178	0.864	0.889	0.218	0.973	0.972	0.137	0.914	0.914	0.165
σ_u	1	0.902	1.062	0.593	0.923	1.159	0.682	0.900	1.047	0.501	0.956	1.156	0.568
a	∞	∞	∞	–	5.9E+08	361.929	7.6E+09	∞	∞	–	2.8E+08	152.462	2.2E+09

		N = 1000					
		Normal-half normal			Student's t-half normal		
DGP 1		Mean	Median	St. Dev.	Mean	Median	St. Dev.
β_0	1	1.157	1.706	2.206	1.220	1.091	0.400
β_1	1	0.988	0.997	0.491	0.999	0.998	0.049
σ_v	1	3.115	2.607	3.412	1.047	1.052	0.125
σ_u	1	0.759	0.003	2.733	0.724	0.876	0.517
a	2	∞	∞	–	2.118	2.107	0.274
DGP 2		Mean	Median	St. Dev.	Mean	Median	St. Dev.
β_0	1	1.183	0.941	0.497	1.188	1.043	0.401
β_1	1	1.000	1.000	0.049	0.999	1.001	0.045
σ_v	1	1.390	1.378	0.146	1.033	1.032	0.130
σ_u	1	0.774	1.084	0.645	0.772	0.957	0.519
a	4	∞	∞	–	4.359	4.222	0.971

Table 1 (continued)

		N = 1000					
		Normal-half normal			Student's *t*-half normal		
DGP 3		Mean	Median	St. Dev.	Mean	Median	St. Dev.
β_0	1	1.160	0.946	0.455	1.191	1.044	0.408
β_1	1	1.000	0.998	0.044	1.000	0.999	0.042
σ_v	1	1.281	1.264	0.122	1.031	1.024	0.132
σ_u	1	0.805	1.087	0.588	0.771	0.966	0.531
a	5	∞	∞	–	5.582	5.262	1.692
DGP 4		Mean	Median	St. Dev.	Mean	Median	St. Dev.
β_0	1	1.099	1.002	0.326	1.070	0.953	0.352
β_1	1	1.000	0.998	0.040	1.000	0.998	0.040
σ_v	1	1.120	1.112	0.090	0.999	0.990	0.118
σ_u	1	0.878	1.005	0.414	0.925	1.074	0.450
a	10	∞	∞	–	3.7E+04	10.266	5.4E+05
DGP 5		Mean	Median	St. Dev.	Mean	Median	St. Dev.
β_0	1	1.034	0.990	0.213	0.971	0.922	0.232
β_1	1	1.000	0.999	0.036	1.000	0.999	0.036
σ_v	1	0.999	0.996	0.071	0.959	0.952	0.084
σ_u	1	0.954	1.008	0.264	1.046	1.104	0.292
a	∞	∞	∞	–	6.1E+08	124.929	1. 3E+10

the former case, we wish to see how the model performs for various values of a, the degrees of freedom parameter. In the Student's t case, the model should be able to estimate the parameters of the DGP with a reasonable degree of accuracy for a range of values of a, and the model should also be able to approximate the case where v is normally distributed. The DGP in each case is as follows:

$$y = 1 + x + v + u,$$
$$x \sim N(0,1), \quad u = |w|, \quad w \sim N(0,1) \tag{33}$$

And v is a standard Student's t random variable with degrees of freedom parameter $a = 2$, $a = 5$, $a = 10$, and $a \to \infty$, respectively. Note that in the latter case, v is normally distributed, so we draw from the standard normal distribution. For each of these DGPs, we then vary the number of observations used in each replication. We consider the cases $N = 100$, $N = 200$, and $N = 1000$. This will give some insight into the small sample performance of the model.

We estimate a Student's t-half normal stochastic cost frontier model using the simulated data generated for each of one thousand replications per DGP. As a point of comparison, we also estimate a normal-half normal stochastic cost frontier model using the data from each replication. The comparison between the results from the two models helps to put the results from our model into context, allowing us to judge its performance relative to that of the

standard model. When comparing the results from the two models, it should be kept in mind that the interpretation of the σ_v parameter differs. In the normal-half normal model, σ_v^2 is the variance of the noise term v, while in the Student's t-half normal model, the variance of v also depends on the degrees of freedom parameter, a. Specifically, the variance of v in our model is given by

$$\mathrm{Var}(v) = \sigma_v^2 \frac{a}{a-2}, \quad a > 2 \tag{34}$$

And is undefined for $a \le 2$. It is therefore to be expected that, when the true DGP involves a Student's t distributed v, the estimated σ_v parameter from the normal-half normal model will tend to be greater than the value used in the DGP. Table 1 summarises the results from each simulation. The mean, median, and standard deviation of the estimates for each parameter across the 1000 replications are shown from both the normal-half normal and Student's t-half normal models. The first two columns show the parameters and corresponding values in each DGP.

Across each of the four DGPs, and regardless of the number of observations, the mean parameter estimates from the Student's t-half normal model are similar to those from the normal-half normal model, and both tend to be reasonably close to the true values. An exception to this is that in the $N = 100$ and $N = 200$ cases for the DGPs with finite a, the mean estimates of the degrees of freedom parameter are many times greater than the true values from the DGP.

Table 2 Summary of within-replication differences in parameter estimates

		N = 100		N = 200		N = 1000	
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
DGP 1 (a = 2)	β_0	0.477	1.634	0.413	1.854	0.062	2.185
	β_1	−0.010	0.524	0.021	0.534	0.011	0.488
	σ_v	−1.248	2.263	−1.512	2.262	−2.068	3.432
	σ_u	−0.479	5.254	−0.289	5.659	−0.035	2.710
	$\sqrt{\text{Var}(v)}$	–	–	–	–	–	–
DGP 2 (a = 4)	β_0	0.114	0.599	0.105	0.550	0.004	0.500
	β_1	0.001	0.086	0.001	0.053	0.000	0.022
	σ_v	−0.293	0.321	−0.307	0.256	−0.357	0.169
	σ_u	−0.125	0.798	−0.124	0.724	−0.002	0.649
	$\sqrt{\text{Var}(v)}$	0.199	0.685	0.131	0.343	0.051	0.127
DGP 3 (a = 5)	β_0	0.078	0.515	0.055	0.483	0.031	0.420
	β_1	−0.001	0.061	−0.001	0.035	0.000	0.015
	σ_v	−0.221	0.273	−0.224	0.197	−0.249	0.135
	σ_u	−0.080	0.678	−0.061	0.632	−0.034	0.544
	$\sqrt{\text{Var}(v)}$	0.111	0.341	0.067	0.211	0.037	0.103
DGP 4 (a = 10)	β_0	0.038	0.370	0.012	0.284	−0.029	0.237
	β_1	0.002	0.031	0.001	0.020	0.000	0.007
	σ_v	−0.123	0.172	−0.113	0.126	−0.121	0.088
	σ_u	−0.036	0.484	−0.005	0.366	0.047	0.307
	$\sqrt{\text{Var}(v)}$	0.033	0.164	0.011	0.094	−0.002	0.063
DGP 5 (a → ∞)	β_0	−0.009	0.249	−0.037	0.186	−0.064	0.075
	β_1	0.000	0.022	0.001	0.010	0.000	0.002
	σ_v	−0.072	0.110	−0.059	0.084	−0.040	0.033
	σ_u	0.021	0.322	0.057	0.240	0.092	0.096
	$\sqrt{\text{Var}(v)}$	−0.003	0.104	−0.013	0.066	−0.018	0.026

However, the median estimates of a are much closer to the true values. The reason for the large difference between the mean and median estimates of a is that the distributions of the estimates are skewed by a small number extremely large values. We find that these very large values occur in repetitions in which the kurtosis of the OLS residuals is low, and specifically where the excess kurtosis— i.e. kurtosis over and above that of the normal distribution—is near to or less than zero. This is significant since, as with the discussion in Section 3.23.2, it further suggests a link between excess kurtosis and identification of the degrees of freedom parameter a, analogous to the link between skewness and identification of the σ_u parameter in the standard model.

In the $N = 1000$ repetitions, this 'wrong kurtosis' issue does not arise in the $a = 4$ or $a = 5$ cases, but does in the $a = 10$ case. In the $\alpha = 2$ case, 'wrong kurtosis' does not arise in either the $N = 200$ or $N = 1000$ repetitions. Intuitively, this is due to the fact that the Student's t distribution with $a = 10$ already closely resembles the normal distribution, and further increases in a yield relatively small changes to the shape of the distribution. Just as the probability of 'wrong skew' arising in the standard model decreases as the number

of observations or the signal to noise ratio, or both, increase (Simar and Wilson 2010), it appears that the probability of 'wrong kurtosis' is negatively related to sample size and positively related to the degrees of freedom parameter.

Finally, on the subject of 'wrong kurtosis', it is important to note that whilst this phenomenon is similar to the 'wrong skew' problem in terms of potentially arising due to the luck of small sample draws from a 'correct kurtosis' distribution, the practical implications of this occurrence are not as severe as in the wrong skew case. In the wrong kurtosis case, the estimation of the Student's t-half normal stochastic frontier models recovers the normal-half normal model estimates, whilst under 'wrong skew' the estimate of the variance of the inefficiency distribution approaches zero, indicating no evidence of inefficiency.

Estimates of σ_v from the normal-half normal model for DGPs 1–4 tend to be greater than 1, as expected. For DGPs 2–4, in which VAR(v) is defined, the standard deviations of v are 1.414, 1.291, and 1.118, respectively. Interestingly, the mean estimates of σ_v from the normal-half normal model correspond closely to these values. This suggests that, although the normal-half normal model cannot capture the

kurtosis of v, it approximates the variance of the noise term relatively well.

Looking at the standard deviations of the parameter estimates from our model, we can see that the sampling variation is generally less than that from the normal-half normal for the $a = 2$, $a = 4$ and $a = 5$ cases. This reflects the greater sensitivity of the standard model to outlying observations, and suggests that our model is indeed more robust to the presence of heavy tails. For the $a = 10$ and $v \sim N(0,1)$ cases, the estimates from the normal-half normal model generally have lower standard deviations than those from our model.

The final DGP corresponds to a model in which $v \sim N(0,1)$. In this case, our model performs well when the true DGP is the normal-half normal model, as evidenced by the very large mean and median estimates of a across all sample sizes, and by the similarity between the frontier and scale parameter estimates produced by the model and those produced by the normal-half normal model. In this case, we are also interested in how the results typically differ from those of the normal-half normal model. Table 2 summarises the within-replication differences between the estimates of β_0, β_1, σ_v, and σ_u from the Student's t-half normal and normal-half normal models, respectively, showing the mean difference and the standard deviation of the differences. A positive (negative) mean indicates that the Student's t-half normal model tended to produce a higher (lower) estimate of a given parameter than the normal-half normal model.

From the above, we can see that, as expected following the discussion in Section 2, the Student's t-half normal model tends to produce lower estimates of σ_u for a given replication. This holds for DGPs 1, 2, and 3, in which the kurtosis of v is greatest. Also, for DGPs 1–3, we see that the normal-half normal model yields higher estimates of σ_v; however, given the discussion around the differing interpretations of σ_v in the two models, we have also looked at the differences in $\sqrt{\mathrm{Var}(v)}$ directly. Table 2 shows that, while the Student's t-half normal model tends to yield lower estimates of σ_v when a is small, its estimates of the standard deviation of v in fact tend to be higher. This supports the expectation, following the discussion in Section 2 and in Stead et al. (2018), that a model with a heavy-tailed distribution in v should attribute more of the overall error variance to noise than to inefficiency.

6 Application to highways maintenance costs in England

In this section, we apply a Student's t-half normal model to the dataset on highway maintenance costs in England used by Stead et al. (2018). In England, responsibility for road maintenance is divided between Highways England—until

2015 the Highways Agency—a publicly-owned company responsible for the strategic 'trunk road network', and the county councils and unitary authorities, who maintain the non-trunk roads within their boundaries. Our data are from the CQC Efficiency Network[6] and consist of costs and cost drivers associated with local authorities' highway maintenance activities.

The majority of previous studies of road maintenance costs have focussed on the issue of marginal costs, and their implications for road pricing, rather than relative efficiency. These use data on motorways and canton roads in Switzerland (Schreyer et al. 2002), motorways in Austria (Sedlacek and Herry 2002), roads in Sweden (Haraldsson 2006; Jonsson and Haraldsson 2008), trunk roads in Poland (Bak et al. 2006; Bak and Borkowski 2009), and motorways and federal roads in Germany (Link 2006; 2009; 2014).

With respect to efficiency studies, Wheat (2017) undertakes the first study of local road maintenance costs in Britain and utilised a forerunner of the dataset under consideration in this paper. The author considered the optimal scale of operation as well as evidence for the cost efficiency of local highway authorities. A normal-half normal stochastic frontier model was used. A further study relating to efficiency in road maintenance is that of Fallah-Fini et al. (2009), which applies DEA to data on eight counties of the US state of Virginia, with expenditure, traffic and equivalent single axle loads as inputs, road area and quality indicators as outputs, and climate factors as non-discretionary variables.

We used an unbalanced panel consisting of data on the 70 English unitary authorities and county councils that were members of the CQC Efficiency Network in 2015–16 and supplied data for at least one year from 2009–10 to that year; this gives us 327 observations in total. Cost data were supplied to the network by each authority according to definitions agreed by a working group of network members, and relate to carriageway maintenance activities only, i.e. they exclude costs associated with related activities such as winter service and footway maintenance. The dataset is updated annually for a new round of analysis. In this study we use the dataset from the 2015–16 round, which was the first year that the network ran. We observe large differences in unit costs, with a large number of extreme outliers in both directions, which are clearly the result of reporting errors. As a result, standard SF models yield a wide range of efficiency predictions, motivating the development of robust SF methods.

In line with the previous literature mentioned above—see Link (2014) for a summary—we use road length and traffic variables as output variables. Detailed breakdowns of local authorities' total road lengths into urban and rural and by

---

[6] See http://www.nhtnetwork.org/

Table 3 Outputs from the Student's *t*-half normal and normal-half normal models

	Student's *t*-half normal			Normal-half normal		
	Estimate	s.e.	Sig	Estimate	s.e.	Sig
β_0	16.058	0.092	***	16.035	0.145	***
β_1 (ln *URL*)	0.149	0.107		0.127	0.171	
β_2 (ln *RRL*)	0.895	0.113	***	0.917	0.179	***
β_3 (ln *URL*^{2})	0.236	0.043	***	0.241	0.063	***
β_4 (ln *RRL*^{2})	0.082	0.010	***	0.085	0.016	***
β_5 (ln *URL* ln*RRL*)	−0.064	0.028	**	−0.081	0.044	*
β_6 (ln *TRAFFIC*)	0.366	0.099	***	0.415	0.154	***
β_7 (*RDCA*)	0.432	0.094	***	0.464	0.144	***
β_8 (*RDCBC*)	−0.071	0.026	***	−0.071	0.039	*
β_9 (*RDCU*)	−0.004	0.003		−0.005	0.005	
β_{10} (*PROP$_{UA}$*)	8.690	1.941	***	7.810	3.241	**
β_{11} (*PROP$_{UB}$*)	1.642	2.279		0.662	3.869	
β_{12} (*PROP$_{UC}$*)	0.273	1.196		0.448	2.054	
β_{13} (*PROP$_{UU}$*)	0.969	0.547	*	1.090	0.835	
β_{14} (*PROP$_{RA}$*)	2.612	1.045	**	2.120	1.571	
β_{15} (*PROP$_{RB}$*)	2.417	1.056	**	2.678	1.544	*
β_{16} (*PROP$_{RC}$*)	1.015	0.641		0.983	0.988	
β_{17} (*YEAR*)	0.038	0.011	***	0.045	0.017	***
β_{18} (ln*WAGE*)	0.779	0.223	***	0.891	0.340	***
$(1-\beta_{18})$ (ln *ROCOSM*)	0.221	–	–	0.109	–	–
σ_u	0.535	0.046	–	0.568	0.015	–
σ_v	0.233	0.016	–	0.276	0.030	–
a	5.198	1.510	–	∞	–	–
Log Likelihood	−187.06	–	–	−189.14	–	–

Statistical significance at the: * 10% level, ** 5% level, *** 1% level

Table 4 Estimated error variances

	Student's *t*-half normal	Normal-half normal
VAR (*u*)	0.104	0.117
VAR(*v*)	0.088	0.076
VAR(*ε*)	0.192	0.194

classification are publicly available from the Department for Transport (DfT). Roads in the UK are classified as: M (motorways), A, B, classified unnumbered, or unclassified; we hereafter refer to the latter two as C and U, respectively. The trunk road network, maintained in England by Highways England, consists of motorways and trunk A roads, leaving non-trunk A roads and all B, C, and U roads as the responsibility of local authorities. Our road length data therefore exclude motorways and trunk A roads, and likewise we use traffic data supplied by DfT which relate only to local authority maintained roads.

We separate total network length into urban and rural road lengths, *URL* and *RRL*, and also include a set of proportion variables relating to each road classification. To control for network quality, we include three road condition indicators, also available from the DfT: *RDCA*, *RDCBC*, and *RDCU* relating to the proportion of A roads, B and C roads, and U roads respectively where maintenance should be considered; we weight these by the corresponding share

of their road classifications in total network length. Our input price variables are: *WAGE*, the median hourly wage in civil engineering by NUTS1 region from the Annual Survey of Hours and Earnings (ASHE), published by the Office for National Statistics (ONS), and a national index of materials prices in road construction that were published by the Department of Business, Innovation and Skills (BIS), *ROCOSM*. We use a modified Cobb-Douglas functional form including second-order terms relating to urban and rural road length. We estimate the following cost frontier:

$$
\begin{aligned}
\ln TOTEX = {} & \beta_0 + \beta_1 \ln URL + \beta_2 \ln RRL + \beta_3 \ln URL^2 \\
& + \beta_4 \ln RRL^2 + \beta_5 \ln URL \ln RRL \\
& + \beta_6 \ln TRAFFIC + \beta_7 RDCA \\
& + \beta_8 RDCBC + \beta_9 RDCU + \beta_{10} PROP_{UA} \\
& + \beta_{11} PROP_{UB} + \beta_{12} PROP_{UC} + \beta_{13} PROP_{UU} \\
& + \beta_{14} PROP_{RA} + \beta_{15} PROP_{RB} + \beta_{16} PROP_{RC} \\
& + \beta_{17} YEAR + \beta_{18} \ln WAGE + \beta_{19} \ln ROCOSM + \varepsilon
\end{aligned}
\tag{35}
$$

Where *TOTEX* is total expenditure, *PROP$_{UA}$* through to *PROP$_{RC}$* are urban A roads, urban B roads, etc. as proportions of the total network length, with the proportion of rural unclassified roads omitted, and *YEAR* is a time trend. All variables are mean-centred, and linear homogeneity in input prices is imposed.

Table 3 compares results from the Student's *t*-half normal and normal-half normal models, showing parameter estimates, standard errors and significance levels. The Student's *t*-half normal model was estimated using a developer's version of *LIMDEP*, along with formulae for efficiency predictions as discussed in the previous section, and the model was estimated based upon 100 Halton draws.

We can see from Table 3 that both models result in generally similar parameter estimates; the main difference is in the standard errors of these estimates, which are approximately a third smaller in the *t*-half normal model than in the normal-half normal model. While the σ_u parameters are comparable between the two models, as discussed in Section 5 the σ_v parameters are not. The variances of *u* is given in both cases by:

$$
\mathrm{VAR}(u) = \frac{\pi - 2}{\pi} \sigma_u^2
\tag{36}
$$

Table 5 Summary of efficiency scores

	Student's *t*-half normal	Normal-half normal
Mean	0.721	0.660
Minimum	0.527	0.225
Maximum	0.855	0.918

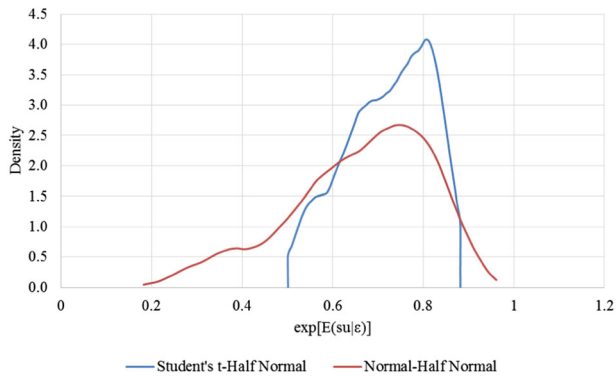**Fig. 1** Kernel densities of cost efficiency scores
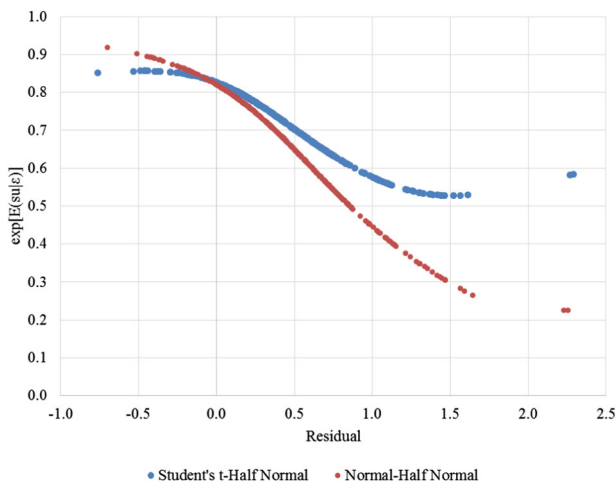

Fig. 2 Scatterplot of cost efficiency scores vs residuals

while VAR(*v*) is given by σ_v^2 in the normal-half normal case, and by (34) in the Student's *t*-half normal case. The estimates of these are compared in Table 4.

Table 4 shows that, in line with our expectations, and the findings of Stead et al. (2018) relating to the logistic-half normal model, that the Student's *t*-half normal model results in a lower estimated variance in inefficiency than the normal-half normal model, and that more of the total error variance is attributed to *v*. The overall error variance is also slightly lower, likewise mirroring the results from the logistic-half normal model. Following from this we expect to find a considerably narrower distribution of efficiency predictions from the Student's *t*-half normal model owing to both the lower estimated VAR(*u*) and the greater shrinkage

towards the mean resulting from the higher estimated VAR(*v*).

Table 5 compares the mean, minimum and maximum efficiency scores from the Student's *t*-half normal and normal-half normal models obtained via $\exp[-E(u|\varepsilon)]$. As expected, the minimum efficiency estimate is considerably higher, and the maximum also significantly lower, in the *t*-half normal model, and therefore the range of efficiency scores is remarkably smaller. In this case less than half. A more complete description is given by Fig. 1, which compares the kernel density estimates for the two sets of efficiency scores.

As well as the distribution of efficiency predictions, we are also particularly interested in predictions at the tails. Figure 2 compares the relationships between the estimated residuals and efficiency predictions from the Student's *t*-half normal and normal-half normal models.

Given the similarity of the frontier parameter estimates, as shown in Table 3, the residuals from the Student's *t*-half normal and normal-half normal models are highly correlated. However the relationships between the residuals and efficiency predictions is shown by Fig. 2 to be substantially different between the two models: for values of residuals between around −0.5 and 1.5, the slope of the function is considerably flatter in the Student's *t*-half normal case, so that a change in *ε* results in a much smaller change in $\exp[-E(u|\varepsilon)]$. However, the most striking difference is that the relationship is non-monotonic in the Student's *t*-half normal case, with $\exp[-E(u|\varepsilon)]$ beginning to decrease slightly for the smallest values of *ε* and increase very considerably for the largest values of *ε*. This is in contrast to the standard SF model, and also the logistic-half normal model as shown by Stead et al. (2018), in which the relationship is monotonic.

The explanation for the large reduction in the range of the efficiency predictions relative to those from the standard model is therefore twofold. First, the use of a heavy-tailed distribution for *v* results in greater shrinkage of efficiency predictions at the tails, as is the case in the normal-logistic model (Stead et al. 2018). Second, the non-monotonicity in $E(u|\varepsilon)$ means that the highest and lowest efficiency scores belong not to the most outlying observations in either direction, as in the standard model, but to observations with less extreme estimated residuals.

The intuitive explanation for this non-monotonicity in the Student's *t*-half normal case is that for outlying observations, the uncertainty associated with $\exp[-E(u|\varepsilon)]$ increases, and further increases in $|\varepsilon|$ are attributed to *v* to such an extent that there is a reduction in $\exp[-E(u|\varepsilon)]$. More formally, we know that the monotonicity of $E(u|\varepsilon)$ is linked to the log-concavity property of the distribution of *v*: specifically, $E(u|\varepsilon)$ is a weakly (strictly) monotonic function of *ε* for any weakly (strictly) log-concave f_v (Ondrich and

Ruggiero [2001]). Goldberger ([1983]) discusses the log-concavity of the (standard) normal, logistic, Laplace and Student's t distributions, and these results are easily extended to the nonstandardised cases with scaling σ_v. It is notable that each of these distributions are log-concave everywhere, with the sole exception of the Student's t distribution, which is log-concave where $v \leq \sqrt{\alpha}\sigma_v$ but log-convex where $v \geq \sqrt{\alpha}\sigma_v$, i.e. at the tails[7]. This explains why $E(u|\varepsilon)$ can be non-monotonic, changing direction at the tails, when v follows a Student's t distribution, whereas $E(u|\varepsilon)$ has been noted to be either weakly or strictly monotonic everywhere when v follows a normal, logistic, or Laplace distribution.

Generalising slightly outside this particular empirical application, the finding of non-monotonicity in efficiency predictions with respect to residuals is useful for applications in economic regulation (RPI-X regulation). This model has interesting incentive properties which may help overcome informational asymmetries in economic regulation between the regulated firms and the regulator. In particular the non-monotonicity property should help discourage firms from trying to game the process by submitting over-favourable data e.g. under reporting costs in the case of cost benchmarking.

As discussed in Section 4, we are interested in testing whether v is normally distributed, in which case the Student's t-half normal contains as a limiting case the normal-half normal model. In addition to this hypothesis we are also interested in testing the null hypothesis of no inefficiency. The likelihood ratio follows a $\chi^2_{1:0}$ distribution in both cases. Log-likelihoods are given in Table [3], and are used to calculate the likelihood ratio statistic as shown in (32). For the first null hypothesis, this gives a likelihood ratio of 4.155 and a corresponding p-value of 0.021. For our second null hypothesis, the Student's t-half normal model reduces to a regression model with Student's t errors: we do not report the results of this regression, except the log-likelihood, which is -189.140. Therefore the corresponding likelihood ratio statistic is 4.150 and the p-value is 0.021. We therefore reject the null hypotheses of normally distributed v and zero inefficiency, indicating that this model performs better than either the standard SF model or the Student's t regression model for this data.

Overall our empirical application demonstrates that the t-half normal SF model is indeed supported by our data relative to the more standard normal-half normal SF model.

Further the frontier parameter estimates are plausible and very similar in both models. In terms of the efficiency predictions, the Student's t-half normal SF model yields a much more plausible spread of efficiency predictions at the tails.

7 Summary and Conclusions

This paper proposes a new stochastic frontier model as a means to account for outlying observations in the context of stochastic frontier analysis. A failure to account for outliers in the standard stochastic frontier model can lead to an exaggerated wide range of efficiency predictions, with the efficiency predictions relating to the least efficient firms being implausibly low. This problem is apparent in our application to highway maintenance departments in England, and also in several other applications we identify in the literature, across a range of countries and industries.

We propose a model combining a Student's t distribution for noise, v, with a half normal distribution for inefficiency, u. Our model is an original and significant contribution to the literature, not just in being able to better accommodate outlying observations in efficiency analysis relative to the standard normal-half normal stochastic frontier (SF) model, but it is the first contribution to contain as a testable limiting case the standard SF model. As such our model provides a natural extension to the tools of practitioners in the field. The advantages of this distribution are that the kurtosis of v is determined by a degrees of freedom parameter which is freely estimated, and that it encompasses as a limiting case the normal distribution as this parameter approaches infinity. This means that the heaviness of the tails of v, reflecting the prevalence of outliers in the data, is flexible, and that testing down to the standard SF model is possible. We derive the log-likelihood and efficiency predictors for the t-truncated normal SF model, and discuss extension to other distributions for u, which is straightforward.

We consider how to test the null hypothesis of normally distributed noise against the alternative of a heavier tailed distribution. We show that the associated LR test statistic is distributed as a mixture chi-squared through appealing to results in Chen and Liang ([2010]), and provide simulation evidence that the test statistic does follow the proposed distribution under the null in large samples.

Simulation evidence is provided for the maximum simulated likelihood estimator of our model. As well as showing that the estimator performs well in recovering the true parameters of our DGP for small values of the Student's t shape (degrees of freedom) parameter, the simulations indicate that in the case of the true DGP being the standard SF (normal-half normal) model, the Student's t SF model recovers very similar estimates to the standard SF model. The combination of this finding, coupled with the

---

[7] Note that as $\alpha \to \infty$, i.e. the distribution of v approaches normality, we would therefore expect the non-monotonicity seen in Fig. [2] to become less prominent. Simulation evidence, not included here for brevity's sake, indicates that this is the case: the larger a is, the more closely $E(u|\varepsilon)$ corresponds to the normal-half normal case. For sufficiently large a, the relationship may be monotonic within the range of the estimated residuals.

ability to test the hypothesis of a normally distributed noise error against a heavier tailed distribution using the LR test statistic, provides reassurance as to the robustness of a modelling approach based on starting with the Student's t SF model and testing to see if a standard SF model is an appropriate simplify restriction.

The simulation results also highlight the possibility of 'wrong kurtosis', specifically where the excess kurtosis of the OLS residual distribution is less than zero. In this case the maximum simulated likelihood estimates approach those from the normal/half-normal model and as such 'wrong kurtosis' is similar to 'wrong skew' previously identified in the literature. Just as the probability of 'wrong skew' arising in the standard model decreases as the number of observations or the signal to noise ratio, or both, increase (Simar and Wilson 2010), it appears that the probability of 'wrong kurtosis' is negatively related to sample size and positively related to the degrees of freedom parameter. It is important to note that the practical implications of this 'wrong kurtosis' are not as severe as in the wrong skew case. In the wrong kurtosis case, the estimation of the Student's t-half normal stochastic frontier models recovers the normal-half normal model estimates, whilst under 'wrong skew' the estimate of the variance of the inefficiency distribution approaches zero, indicating no evidence of inefficiency.

We apply a Student's t-half normal model to estimate a cost frontier using a dataset on English local authorities' highway maintenance costs, and compare the model's outputs with those from the normal-half normal model. We find similar frontier parameter estimates from the two models, though with reduced standard errors in the t-half normal model. We implement testing against the standard SF model, and find that we are able to reject the null hypothesis of normally distributed v. This implies that it is important to account for heavy tails in our data.

The main empirical differences between the two models are firstly, a reduced estimate of $VAR(u)$ and an increased estimate of $VAR(v)$ and secondly a reduced range of efficiency predictions according to $\exp[-E(u|\varepsilon)]$. Thirdly, we find a non-monotonic relationship between residuals and $\exp[-E(u|\varepsilon)]$, in contrast to the standard SF model. This could be a very an important feature of the model in practical applications, for example in regulatory settings, where this could incentivise correct reporting of data by regulated firms, as the alternative 'gaming' behaviour of under reporting, say, costs, many actually reduce the firm's efficiency prediction.

Finally, we have integrated our model into the LIMDEP/NLOGIT econometric software (Greene 2016), and have also written a Stata package enabling estimation of the model[8], which means that it is ready for use by practitioners. An interesting avenue for future research would be to consider the impact of Student's t noise in the context of the various panel data SF specifications, for example models accounting for unobserved heterogeneity.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

8 Appendix A

First order conditions (reintroducing i subscripts for completeness) for maximisation in the t-truncated normal case are

$$\frac{\partial \ln SL}{\partial a} = \frac{N}{2}\left[\psi\left(\frac{a+1}{2}\right) - \psi\left(\frac{a}{2}\right) - \frac{1}{a}\right] + \sum_{i=1}^{N}\sum_{q=1}^{Q}\frac{\frac{\partial g(u_{iq})}{\partial a}}{\sum_{q=1}^{Q} g_i(u_{iq})} = 0 \tag{37}$$

$$\frac{\partial \ln SL}{\partial \beta} = \sum_{i=1}^{N}\frac{\sum_{q=1}^{Q}\frac{\partial g(u_{iq})}{\partial \varepsilon_i}\frac{\partial \varepsilon_i}{\partial \beta}}{\sum_{q=1}^{Q} g_i(u_{iq})} = 0 \tag{38}$$

$$\frac{\partial \ln SL}{\partial \sigma_v} = -\frac{N}{\sigma_v} + \sum_{i=1}^{N}\sum_{q=1}^{Q}\frac{\frac{\partial g(u_{iq})}{\partial \sigma_v}}{\sum_{q=1}^{Q} g_i(u_{iq})} = 0 \tag{39}$$

$$\frac{\partial \ln SL}{\partial \sigma_u} = \sum_{i=1}^{N}\sum_{q=1}^{Q}\frac{\frac{\partial g(u_{iq})}{\partial u_{iq}}\frac{\partial u_{iq}}{\partial \sigma_u}}{\sum_{q=1}^{Q} g_i(u_{iq})} = 0 \tag{40}$$

$$\frac{\partial \ln SL}{\partial \mu} = \sum_{i=1}^{N}\sum_{q=1}^{Q}\frac{\frac{\partial g(u_{iq})}{\partial u_{iq}}\frac{\partial u_{iq}}{\partial \mu}}{\sum_{q=1}^{Q} g_i(u_{iq})} = 0 \tag{41}$$

---

[8] The package *rfrontier*, which can be found at www.its.leeds.ac.uk/bear, enables estimation of stochastic frontier models in which v follows a Student's t, Cauchy, or logisitic distribution.

The expressions needed to complete the first order conditions above are as follows

$$g(u_{iq}) = \left[1 + \frac{1}{a}\left(\frac{\varepsilon_i + su_{iq}}{\sigma_v}\right)^2\right]^{-\frac{a+1}{2}} \tag{42}$$

$$\frac{\partial g(u_{iq})}{\partial a} = g(u_{iq})\frac{(a+1)\left(\frac{\varepsilon + su_{iq}}{\sigma_v}\right)^2}{2a^2\left[1 + \frac{1}{a}\left(\frac{\varepsilon + su_{iq}}{\sigma_v}\right)^2\right]} - \frac{1}{2}g(u_{iq})\ln\left[1 + \frac{1}{a}\left(\frac{\varepsilon + su_{iq}}{\sigma_v}\right)^2\right] \tag{43}$$

$$\frac{\partial g(u_{iq})}{\partial \varepsilon_i}\frac{\partial \varepsilon_i}{\partial \beta} = \left(\frac{a+1}{a}\right)\left(\frac{x_i}{\sigma_v}\right)\left[1 + \frac{1}{a}\left(\frac{\varepsilon + su_{iq}}{\sigma_v}\right)^2\right]^{-\frac{a+3}{2}}\left(\frac{\varepsilon + su_{iq}}{\sigma_v}\right) \tag{44}$$

$$\frac{\partial g(u_{iq})}{\partial \sigma_v} = \left(\frac{a+1}{a}\right)\left(\frac{1}{\sigma_v}\right)\left[1 + \frac{1}{a}\left(\frac{\varepsilon + su_{iq}}{\sigma_v}\right)^2\right]^{-\frac{a+3}{2}}\left(\frac{\varepsilon + su_{iq}}{\sigma_v}\right)^2 \tag{45}$$

$$\frac{\partial g(u_{iq})}{\partial u_{iq}} = -\left(\frac{a+1}{a}\right)\left(\frac{s}{\sigma_v}\right)\left[1 + \frac{1}{a}\left(\frac{\varepsilon + su_{iq}}{\sigma_v}\right)^2\right]^{-\frac{a+3}{2}}\left(\frac{\varepsilon + su_{iq}}{\sigma_v}\right) \tag{46}$$

$$\frac{\partial u_{iq}}{\partial \mu} = 1 + \frac{\phi\left(\frac{\mu}{\sigma_u}\right)(F_{iq} - 1)}{\phi\left\{\Phi^{-1}\left[1 + \Phi\left(\frac{\mu}{\sigma_u}\right)(F_{iq} - 1)\right]\right\}} \tag{47}$$

$$\frac{\partial u_{iq}}{\partial \sigma_u} = \Phi^{-1}\left[1 + \Phi\left(\frac{\mu}{\sigma_u}\right)(F_q - 1)\right] + \frac{\frac{\mu}{\sigma_u^2}\phi\left(\frac{\mu}{\sigma_u}\right)(1 - F_{iq})}{\phi\left\{\Phi^{-1}\left[1 + \Phi\left(\frac{\mu}{\sigma_u}\right)(F_{iq} - 1)\right]\right\}} \tag{48}$$

9 Appendix B

In order to avoid the complications presented by 'wrong skew' and simulation chatter, not to mention the large amount of time that would be needed to estimate the full model via MSL for thousands of repetitions, we restrict our attention to the case where $\sigma_u = 0$, so that the model reduces to a regression with Student's t errors. That is, true data generating process is

$$y_i = 1 + v_i \tag{49}$$

where v_i is drawn from a standard normal distribution. We then estimate two regression models: one assuming a normally distributed error term and estimated via OLS, and the other assuming a Student's t error term and estimated via ML. The LR statistic is then calculated as in (32), where $\ln L_0$ and $\ln L_A$ are the resulting log-likelihoods from the former and latter models, respectively. We use a sample size of 5000 for each of 50,000 repetitions. Table 6 below compares the mean, variance, and percentiles of the $\chi^2_{1:0}$

Table 6 Comparison of the $\chi^2_{1:0}$ distribution and the sampling distribution of the LR statistic under the null hypothesis that the errors are normally distributed

		$\chi^2_{1:0}$ distribution	Sampling distribution
Mean		0.500	0.532
Percentiles	25th	0.000	0.000
	50th	0.000	0.011
	75th	0.455	0.527
	90th	1.642	1.718
	95th	2.706	2.759
	99th	5.412	5.585

distribution to those of the sampling distribution of the LR statistic from our Monte Carlo replications.

Comparing the two columns, the evidence from our Monte Carlo simulations seems to support the idea that Case 5 of Self and Liang (1987) applies in this case, i.e. that the LR statistic follows a $\chi^2_{1:0}$ distribution.

References

Aigner D, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. J Econ 6(1):21–37

Azzalini A (1985) A class of distributions which includes the normal ones. Scand J Stat 12(2):171–178

Bak M, Borkowski P (2009) Marginal Cost of Road Maintenance and Renewal in Poland, CATRIN (Cost Allocation of TRansport INfrastructure cost) Deliverable D6, Annex 2. ITS, University of Leeds, Leeds, UK

Bak M, Borkowski P, Musiatowicz-Podbial G, Link H (2006) Marginal Infrastructure Cost in Poland, Marginal Cost Case Studies for Road and Rail Transport Deliverable D3, Annex 1.2C. ITS, University of Leeds, Leeds, UK

Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. J Econ 38(3):387–399

Bauer PW, Berger AN, Ferrier GD, Humphrey DB (1998) Consistency conditions for regulatory analysis of financial institutions: a comparison of frontier efficiency methods. J Econ Bus 50(2):85–114

Berger AN, Humphrey DB (1991) The dominance of inefficiencies over scale and product mix economies in banking. J Monet Econ 28(1):117–148

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–444

Chen C-F (2007) Applying the stochastic frontier approach to measure hotel managerial efficiency in Taiwan. Tour Manag 28(3):696–702

Chen Y, Liang K-Y (2010) On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems. Biometrika 97(3):603–620

Cook RD (1977) Detection of influential observation in linear regression. Technometrics 19(1):15–18

Cullinane K, Wang T-F, Song D-W, Ji P (2006) The technical efficiency of container ports: comparing data envelopment analysis

and stochastic frontier analysis. Transp Res Part A: Policy Pract 40(4):354–374

De Borger B, Kerstens K (1996) Cost efficiency of Belgian local governments: a comparative analysis of FDH, DEA, and econometric approaches. Reg Sci Urban Econ 26(2):145–170

Drake L, Simper R (2003) The measurement of English and Welsh police force efficiency: a comparison of distance function models. Eur J Oper Res 147(1):165–186

Economou P (2011) On model selection in the case of nested distributions—An application to frailty models. Stat Methodol 8 (2):172–184

Fallah-Fini, S., Triantis, K. and de la Garza, J.M. 2009. Performance measurement of highway maintenance operation using data envelopment analysis: Environmental considerations. In: *Proceedings of IIE Annual Conference. Miami, USA*: Institute of Industrial Engineers.

Fan J (1991) On the optimal rates of convergence for nonparametric deconvolution problems. Ann Stat 19(3):1257–1272

Fan J (1992) Deconvolution with supersmooth distributions. Can J Stat / La Rev Can De Stat 20(2):155–169

Fecher F, Kessler D, Perelman S, Pestieau P (1993) Productive performance of the French insurance industry. J Product Anal 4 (1):77–93

Geweke JF, Keane MP, Runkle DE (1997) Statistical inference in the multinomial multiperiod probit model. J Econ 80(1):125–165

Goldberger AS (1983) Abnormal selection bias. In: Amemiya T and Goodman LA, (eds). *Studies in Econometrics, Time Series, and Multivariate Statistics*. New York, USA: Academic Press, 67–84

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

Greene WH (2008) The econometric approach to efficiency analysis. In: Fried HO, Lovell CAK, Schmidt SS, (eds). *The Measurement of Productive Efficiency and Productivity Growth*. Second ed. New York, USA: Oxford University Press, 92–159

Greene WH (2016) *LIMDEP Version 11.0 Econometric Modeling Guide*. Econometric Software

Griffin JE, Steel MFJ (2007) Bayesian stochastic frontier analysis using WinBUGS. J Product Anal 27(3):163–176

Grushka E (1972) Characterization of exponentially modified Gaussian peaks in chromatography. Anal Chem 44(11):1733–1738

Gupta AK, Nguyen N (2010) Stochastic frontier analysis with fat-tailed error models applied to WHO health data. Int J Innov Manag, Inf & Prod 1(1):43–48

Hajargasht G (2015) Stochastic frontiers with a Rayleigh distribution. J Product Anal 44(2):199–208

Hajargasht G, Griffiths WE (2018) Estimation and testing of stochastic frontier models using variational Bayes. J Product Anal 50(1):1–24

Haraldsson M (2006) Marginal Cost for Road Maintenance and Operation—A Cost Function Approach, Marginal Cost Studies for Road and Rail Transport Deliverable D3, Annex. ITS, University of Leeds, Leeds, UK

Horrace WC, Parmeter CF (2011) Semiparametric deconvolution with unknown error variance. J Product Anal 35(2):129–141

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

Horrace WC, and Wright IA, forthcoming. Stationary points for parametric stochastic frontier models. *Journal of Business & Economic Statistics*

Hurst S, 1995 *The Characteristic Function of the Student t Distribution*. Financial Mathematics Research Report 006-95. Australian National University

Jondrow J, Knox Lovell CA, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. J Econ 19(2):233–238

Jonsson L, Haraldsson M (2008) Marginal Costs of Road Maintenance in Sweden, CATRIN (Cost Allocation of TRansport INfrastructure cost) Deliverable D6, Annex 1. VTI, Stockholm, Sweden

Lange K, Sinsheimer JS (1993) Normal/independent distributions and their applications in robust regression. J Comput Graph Stat 2 (2):175–198

Lange KL, Roderick JAL, Jeremy MGT (1989) Robust statistical modeling using the t distribution. J Am Stat Assoc 84(408):881–896

Latruffe L, Balcombe K, Davidova S, Zawalinska K (2004) Determinants of technical efficiency of crop and livestock farms in Poland. Appl Econ 36(12):1255–1263

Lin B, Wang X (2014) Exploring energy efficiency in China's iron and steel industry: A stochastic frontier approach. Energy Policy 72:87–96

Link H (2006) An econometric analysis of motorway renewal costs in Germany. Transp Res Part A: Policy Pract 40(1):19–34

Link H (2009) Marginal Costs of Road Maintenance in Germany, CATRIN (Cost Allocation of TRansport INfrastructure) Deliverable D6, Annex 3. VTI, Stockholm, Sweden

Link H (2014) A cost function approach for measuring the marginal cost of road maintenance. J Transp Econ Policy (JTEP) 48(1):15–33

Mahalanobis PC (1936) On the generalized distance in statistics. Proc Natl Inst Sci (India) 2:49–55

Meeusen W, van Den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. Int Econ Rev 18(2):435–444

Nguyen, N (2010) *Estimation of Technical Efficiency in Stochastic Frontier Analysis*. PhD thesis, Bowling Green State University.

Ondrich J, Ruggiero J (2001) Efficiency measurement in the stochastic frontier model. Eur J Oper Res 129(2):434–442

Parmeter CF, Kumbhakar SC (2014) Efficiency analysis: A primer on recent advances. Found Trends Econ 7(3-4):191–385

Schreyer C, Schmidt N, Maibach M (2002) Road Econometrics–Case Study Motorways Switzerland. UNITE (UNIfication of accounts and marginal costs for Transport Efficiency) Deliverable 10, Annex A1b. ITS, University of Leeds, Leeds, UK

Sedlacek N, Herry M (2002) Infrastructure Cost Case Studies. UNITE (UNIfication of accounts and marginal costs for Transport Efficiency) Deliverable 10, Annex A1c. ITS, University of Leeds, Leeds, UK

Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82(398):605–610

Simar L, Wilson PW (2010) Inferences from cross-sectional, stochastic frontier models. Econom Rev 29(1):62–98

Stead AD, Wheat P, Greene WH (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: Greene WH, Khalaf L, Makdissi P, Sickles RC, Veall M, and Voia M. (eds). *Productivity and Inequality*. Cham, Switzerland: Springer, 1-19

Stevenson RE (1980) Likelihood functions for generalized stochastic frontier estimation. J Econ 13(1):57–66

Tancredi, A (2002) *Accounting for Heavy Tails in Stochastic Frontier Models*. Working Paper No. 2002.16. Department of Statistical Sciences, University of Padua

Tchumtchoua S, Dey DK (2007) Bayesian estimation of stochastic frontier models with multivariate skew t error terms. Commun Stat - Theory Methods 36(5):907–916

Train KE (2009) *Discrete Choice Methods with Simulation*. Second ed. Cambridge, UK: Cambridge University Press

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

Waldman DM (1982) A stationary point for the stochastic frontier likelihood. J Econ 18(2):275–279

Wang WS, Schmidt P (2009) On the distribution of estimated technical efficiency in stochastic frontier models. J Econ 148(1):36–45

Wheat P (2017) Scale, quality and efficiency in road maintenance: evidence for English local authorities. Transp Policy 59: 46–53