

# **Robust Stochastic Frontier Analysis: A Student's $t$ -Half Normal Model with Application to Highway Maintenance Costs in England**

**A.D. Stead, P. Wheat and W.H. Greene**

## **Abstract**

The presence of outliers in the data has serious implications for stochastic frontier analysis because they may distort parameter estimates and, crucially, lead to an exaggerated spread of efficiency predictions. One way of increasing the robustness of the model to outliers is to alter the distributional assumptions about the two-sided error component so that it allows for thick tails. Several existing proposals specify thick tailed distributions for both error components in order to arrive at a closed form log-likelihood function, limiting the analyst's choice of efficiency distribution. Stead et al. (2017) demonstrate that simulation methods may be used to pair a logistically distributed noise term with any of the canonical efficiency distributions, and that a far narrower range of efficiency predictions is obtained by doing so. We extend this idea by proposing a Student's  $t$  distribution for the noise term, which generalises the normal distribution by adding a shape parameter governing the degree of kurtosis, therefore having the advantages of greater flexibility in the thickness of the tails and of nesting the normal distribution. We estimate a Student's  $t$ -half normal cost frontier for highways authorities in England and find that the model yields a significantly narrower range of efficiency predictions. We examine testing against the standard stochastic frontier model.

---

A.D. Stead • P. Wheat  
Institute for Transport Studies, University of Leeds, Leeds, LS2 9LJ, UK  
e-mail: [a.d.stead@leeds.ac.uk](mailto:a.d.stead@leeds.ac.uk)

W.H. Greene  
Department of Economics, Stern School of Business, New York University, New York City, New York, 10012, USA

## 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 problematic in two ways: it affects the position of each decision making unit (DMU) relative to the frontier, and it affects 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 can still be very large in the presence of data with outlying observations. The specific motivation for this study is 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 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 also came across this issue in a number of other data sets, and a casual look at the empirical SF literature seems to indicate that the ranges of individual efficiency predictions are sometimes difficult to reconcile with a priori expectations.

In this paper we consider methods of dealing with outliers in the context of SFA. After consideration of possible existing approaches, we propose a new 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, and in the fact that the Student's  $t$  distribution nests the

normal distribution as the degrees of freedom parameter approaches infinity. For any given distribution of  $u$ , our model nests the standard SF model. This enables testing against the standard model, in contrast to previous proposals which utilise non-nested specifications.

The structure of this paper is as follows: Section 2 looks at existing methods available to handle a large number of outliers in frontier analysis, and reviews the relevant literature, Section 3 introduces  $t$ -truncated normal and  $t$ -gamma stochastic frontier models for dealing with heavy-tailed noise and discusses efficiency prediction and hypothesis testing, Section 4 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 5 gives our summary and conclusions. Some technical results appear in an Appendix.

## **2. Literature Review: Approaches to Outliers in SFA**

Our focus is on possible existing approaches to dealing with outliers in the context of SFA. We start by looking at ad-hoc solutions using methods not specifically developed for this purpose but having some potential to mitigate one or more of the symptoms, e.g. by reducing the spread of efficiency scores. These generally have limited usefulness in dealing with outliers and involve a degree of arbitrariness. We then move on to discuss methods more appropriate to dealing with the root cause of these symptoms, the presence of outliers in the data. As in linear modelling generally, these essentially revolve around the detection and removal of outliers or the adoption of alternative distributional assumptions about the noise term.

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

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

Where the subscript  $i$  denotes the observation,  $y$  is the dependent variable,  $x$  is a vector of independent variables, and  $\varepsilon$  is an error term with two components. The noise component,  $v$ , follows a symmetric distribution centred at zero, and the inefficiency component,  $u$ , is drawn from a one-sided distribution, and 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,  $v$  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 SFA as an alternative to 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 caused by inefficiency. Indeed 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  $u$  based on the conditional distribution of  $u_i|\varepsilon_i$ . Dropping the subscript  $i$ , this is given by

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

where  $f_v$  and  $f_u$  are the probability density functions of  $v$  and  $u$  respectively, and  $f_\varepsilon$  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  $E[\exp(-u|\varepsilon)]$  (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 it approaches  $f_u(u)$  only as  $VAR(v) \rightarrow 0$ , and  $E(u)$  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—i.e. thick tailed—noise distribution, the normal distribution usually assumed for  $v$  is inappropriate since it is mesokurtic—i.e. neither thin or thick tailed—for any given parameter values. Intuitively, we would 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.

## **2.1. Alternative Efficiency Distributions and Predictors**

Given an apparently implausible distribution of efficiency predictions, an obvious option is to adopt an alternative assumption about the distribution of inefficiency. For example, adopting an inefficiency distribution with thinner tails—or perhaps a truncation of the tail, as proposed by Lee (1996), Almanidis et al. (2014) and others—should intuitively result in a narrower spread of efficiency predictions by attributing outliers to noise to a greater extent. Clearly, however, if the real problem is outliers in the data, then the effectiveness of this solution will be limited since there will be not enough shrinkage at the tails. Nor does this workaround

address the problem of imprecise parameter estimation, and if anything we may be driven to estimate a model in which both the inefficiency and noise distributions are inappropriate.

An even more arbitrary approach would be to adopt an alternative efficiency predictor post-estimation. The two alternatives proposed by Jondrow et al. (1982) are the mean and mode of the distribution of  $u|\varepsilon$ , with the former the most commonly used in the applied literature. When the distribution of  $u$  is skewed in the expected direction—i.e. positively in the cost case and negatively in the production case—the conditional mode predictor yields predictions of efficiency that are higher for the most efficient DMUs, but practically the same for the most inefficient DMUs. There is therefore a trade-off with the conditional mode predictor, between obtaining what might appear to be more reasonable efficiency predictions for the most efficient firms via the conditional mode predictor, and expanding the range of efficiency predictions to be even more unreasonable. Adopting alternative predictors therefore does not seem to be a satisfactory solution.

Another approach would be to qualify the point predictions of efficiency with prediction intervals: Horrace and Schmidt (1996) and Bera and Sharma (1999) propose using the quantile function for the distribution of  $u|\varepsilon$ . However Wheat et al. (2014) note that this does not necessarily yield minimum a width interval, which they derive for the normal-half normal model. When the estimated scale of the inefficiency distribution is inflated by outliers, the distribution of  $u|\varepsilon$  will have a high variance, and therefore the range of probable values will be wide. However, this really just serves to highlight the problem, and of course the prediction interval will include values even less plausible than the point prediction. A better solution would therefore be to obtain a better estimate of the scale of the distribution of  $u|\varepsilon$  by taking outliers into account.

## 2.2. Heteroskedastic SF Models

Outliers may be thought of as a result of heteroskedasticity in one or both error components, with a small proportion of errors having higher variance. Therefore one way of mitigating the impact of outliers is to allow for heteroskedasticity. A number of heteroskedastic SF models have been introduced, although these have tended to focus on heteroskedasticity in  $u$ : the general approach is to modify the standard SF specification so that the variance or standard deviation of  $u$  is modelled as a function of a vector of covariates, as proposed by Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill et al. (1995).

Clearly, however, we would not want outliers to be picked up as heteroskedasticity in  $u$ , since the estimated variance of  $u$  for outlying observations would still be inflated. The application of these models is usually to capture the relationship between a set of covariates and efficiency, similar to the Battese and Coelli (1995) specification in which they enter the pre-truncation mean of a truncated normal inefficiency distribution; Wang (2002) combines these two approaches. Rather, we want to allow for heteroskedasticity in  $v$ : Hadri (1999) generalised the Caudill et al. (1995) heteroskedastic stochastic frontier model by allowing for heteroskedasticity in both error components. It would be straightforward, therefore, to estimate a model with heteroskedasticity in  $v$ .

In general, however, the difficulty with using heteroskedastic SF models to control for the presence of outliers is that this would require an appropriate indicator variable for identifying outlying observations, which would have to be identified either on some arbitrary ex-post basis, or on the basis of established techniques for the detection of outliers, in which case it may be better to simply exclude the observations.

### **2.3. Outlier Detection and Removal**

In general, one simple approach to outliers is to exclude them from the analysis. This, however, requires that appropriate criteria are used to identify which observations are outlying and therefore candidates for removal. Many competing criteria exist. One approach is to exclude observations that have a disproportionate effect on parameter estimates. In the context of OLS, Cook (1977) proposed a measure of the influence of an individual observation on the parameter estimates, known as Cook's distance. A similar measure sometimes used for the same purpose is the Mahalanobis distance (Mahalanobis, 1936). However, while these measures provide an indication of the influence of an observation, the cut-off beyond which an observation should be removed is arbitrary and involves uncertainty which is not reflected in subsequent modelling.

The biggest drawback of these methods in the current context, however, is that they may not be appropriate in the case of skewness of the composed error. Cook's Distance explicitly assumes normally distributed errors, while in SFA the distribution of the composed error depends upon the distribution of  $v$  and  $u$ : in the normal-half normal case, 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), for example. As such, these distributions are more robust to outliers in one direction than the other, and the derivation of some analogous measure of leverage taking this into account would inevitably be bound up with one's distributional assumptions.

### **2.4. Alternative Noise Distributions**

Another potential method of dealing with outliers is to change the distributional assumptions about the error term such that the model is more robust to the existence of outliers. As discussed



in Section 2.1, we should change the distributions of  $v$  specifically, rather than  $u$ . In principle, any distribution that is symmetric, centred around zero and unimodal is an appropriate candidate for the distribution of  $v$ , and 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 as described in Section 2.2. All of these distributions have heavier tails than the normal distribution, and therefore offer greater robustness to outliers.

In the context of SFA, Tancredi (2002) proposes a model in which  $v$  is drawn from a  $t$  distribution and  $u$  from a half  $t$  distribution—i.e. the left truncation of a Student's  $t$  distribution at zero—resulting in a skew  $t$  distribution for the composed error, and shows that the distribution of  $u|\varepsilon$  becomes flat as  $s\varepsilon \rightarrow \infty$  and therefore the uncertainty associated with efficiency predictions increases. In contrast, as  $s\varepsilon \rightarrow \infty$  in the normal-half normal case the distribution of  $u|\varepsilon$  becomes concentrated around zero, implying full efficiency; the  $t$ -half  $t$  model therefore deals with outliers in a more satisfactory way. The author also notes that, since the  $t$  distribution approaches the normal distribution as the degrees of freedom parameter approaches infinity, the model nests the normal-half normal model. The model is applied to the Christensen and Greene (1976) data set on US electric utilities, and  $E[\exp(-u|\varepsilon)]$  and  $VAR[\exp(-u|\varepsilon)]$  are compared to those from the normal-half normal model. These are shown to behave very differently, with the efficiency predictions from the  $t$ -half  $t$  model having a smaller range and being nonlinear in relation to  $s\varepsilon$ , declining for large values of  $s\varepsilon$  as the

variance of the distribution of  $u|\varepsilon$  increases in line with the author's findings on the limiting distribution. Tchumtchoua and Dey (2007) study the  $t$ -half  $t$  model in a Bayesian setting.

Nguyen (2010) introduces two additional alternative heavy tailed distributions for  $v$ : the Laplace distribution and the Cauchy distribution: the latter is paired with both half Cauchy and truncated Cauchy distributions—i.e. the left truncation at zero of a Cauchy distribution with means at zero and freely estimated, respectively—while the former is paired with an exponential distribution for  $u$ . The Cauchy distribution is potentially problematic since its mean, variance and higher order moments are undefined. This is also the case for the skew Cauchy distribution (Arnold and Beaver, 2000), for which only fractional moments exist. Despite this, the author is able to derive  $\exp[-E(u|\varepsilon)]$  for the Cauchy-truncated Cauchy and Cauchy-half Cauchy models and applications are shown in Gupta and Nguyen (2010).

The Laplace distribution also presents possible issues, given that it has a cusp at the mean and hence a non-differentiable point in the likelihood function. Standard results for consistency and asymptotic normality of maximum likelihood (ML) estimates do not apply. The usefulness of the Laplace-exponential model derived by Nguyen (2010) is also limited by simplifying assumptions the author makes in its derivation, specifically that the scale parameters for  $v$  and  $u$  in the Laplace-exponential model are the same. Nevertheless, Horrace and Parmeter (2015) derive Laplace-truncated Laplace and Laplace-exponential SF models without such restrictions (The left truncation of a Laplace distribution at or above zero results in an exponential distribution). The authors show that when the variance in  $u$  is zero, the estimator of the model reduces to the least absolute deviations (LAD) estimator. That a regression with Laplace errors minimises absolute deviations has previously been noted by Keynes (1911). The distribution of  $u|\varepsilon$ , and hence any efficiency prediction, is constant for positive values of  $\varepsilon$ .

The models above each perform better than the standard SF model in the presence of outliers because of their use of heavy tailed distributions of  $v$ . However, in addition to the specific issues identified above, a drawback of each of the models is that they assume that the distribution of  $u$  is a left truncation of the distribution of  $v$ , since this makes it possible to derive a closed form expression for the log-likelihood – see Proposition 9 of Azzalini and Capitanio (2003), which gives the density function for sum of one random variable and the absolute value of another random variable when both random variables follow the same elliptical distribution. While simplifying derivation and estimation of the model, this limits comparability to the standard SF model. A preferable approach would be to combine a heavy tailed distribution of  $v$  with any of the canonical distributions for  $u$ , e.g. half normal or exponential.

This shortcoming is addressed by Stead et al. (2017), who propose a model in which  $v$  follows a logistic distribution and  $u$  follows, e.g. a half normal or exponential distribution. They estimate a logistic-half normal SF model using maximum simulated likelihood and data on English local authorities' road maintenance costs. The authors 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  $|\varepsilon|$ ; the model is therefore better at handling outliers at either end than the standard SF model.

A remaining issue is the arbitrary assumptions on the degree of kurtosis in  $v$  embedded in most of the above models—and therefore the extent to which there are outliers in the data—since the logistic and Cauchy distributions have excess kurtosis of 1.2 and 3 respectively, regardless of parameter values; likewise, though kurtosis is undefined for the Cauchy distribution, its shape does not depend upon parameter values. If the assumed distribution of  $v$  has heavier (lighter) tails than the true distribution, this will result in an underestimate (overestimate) of

the scale of  $u$  and excessive (insufficient) shrinkage of  $u$  according to  $E[\exp(-u|\varepsilon)]$  at the tails.

Ultimately, the kurtosis of  $v$  is an empirical question, and therefore we should ideally use a distribution for  $v$  for which kurtosis is flexible. Of the alternatives discussed above, the Student's  $t$  distribution, in which kurtosis depends upon the degrees of freedom parameter is ideal in this regard. A further advantage of the  $t$  distribution is that it nests the Cauchy distribution when the degrees of freedom parameter is equal to one and the normal distribution as it approaches infinity. By extension, an SF model with a  $t$  distribution for  $v$  nests 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.

Following the discussion above, we propose a robust SF model in which  $v$  follows a Student's  $t$  distribution and  $u$  follows some one-sided distribution, e.g. half normal or exponential. In many cases, this will result in a log-likelihood function with no closed form solution. Griffin and Steel (2007) briefly discuss how to estimate  $t$ -half normal,  $t$ -exponential, and  $t$ -gamma Bayesian SFA models using the WinBUGS software package—which uses Markov Chain Monte Carlo (MCMC) methods for numerical integration—and apply these to the Christensen and Greene (1976) electricity data set. But the model remains unexplored in the classical framework. We propose the use of simulation methods, as used by Greene (2003) to implement the normal-gamma model, to solve the convolution of the Student's  $t$  distribution and the distribution of  $v$  and arrive at a simulated likelihood function.

At this point it is worth mentioning a recent proposal by Hofler (2014) with a similar motivation and using similar techniques, to pair a generalised normal distribution for  $v$  with a half-normal

distribution for  $u$ . The generalised normal distribution, introduced by Nadarajah (2005), is similar to the Student's  $t$  distribution in that it has a kurtosis parameter and nests the normal distribution. However, unlike the Student's  $t$ , the generalised normal distribution can be either leptokurtic or platykurtic when its shape parameter is lower than or greater than two, respectively; in principle, an SF model with a generalised normal distribution for  $v$  can therefore account not only for heavy tails but also for the converse problem of light tails, making it an attractive choice. When the kurtosis parameter is equal to one, the generalised normal distribution also nests the Laplace distribution.

However, the generalised normal distribution has several properties which hinder maximum likelihood estimation: as the thickness of the tails increases—i.e. the kurtosis parameter goes below two—the distribution acquires a cusp at the mean and as a consequence has an infinite number of continuous derivatives only when the kurtosis parameter is a positive, even integer. Otherwise, the number of derivatives is equal to the largest integer less than or equal to the kurtosis parameter. As a consequence the standard results for consistency and asymptotic normality of ML estimates only apply when the kurtosis parameter is greater than or equal to two. When the kurtosis parameter is greater than two—i.e. the distribution is thin tailed—the distribution has only finite support. We conclude that the primary advantage of using a generalised normal distribution for  $v$  rather than a Student's  $t$  distribution, the ability to account for thin tails in the noise term, is more than outweighed by these issues. It is difficult to imagine what might give rise to light tails in practice. To the best of our knowledge, the generalised normal-half normal model has not yet been implemented.

The following section shows the derivation of an SF model with a Student's  $t$  distribution for  $v$ : we focus on the  $t$ -truncated normal and  $t$ -gamma models—which nest the  $t$ -half normal and  $t$ -exponential models—though extension to other distributions for  $u$  is straightforward.

### 3. The $t$ -Truncated Normal and $t$ -Gamma SF Models

In this section we derive simulated log likelihood functions and efficiency predictors for the  $t$ -truncated normal and  $t$ -gamma SF models, and discuss estimation and hypothesis testing. Results for the  $t$ -half normal and  $t$ -exponential models can be obtained via some simple restrictions. Extensions to other distributions of  $u$  are shown to be straightforward if the quantile function of that distribution has a closed form, while in many other cases—as with the  $t$ -gamma—the simulated log likelihood function becomes slightly more complex.

#### 3.1. Formulation and Estimation

In SFA, the error  $\varepsilon$  is composed of a symmetric noise component  $v$  and an inefficiency component  $u$  which is drawn from a one-sided distribution:

$$\varepsilon = v - su \quad (3)$$

where  $s = 1$  in for a production frontier and  $s = -1$  for a cost frontier. In this study, we assume that  $v$  is drawn from a non-standard  $t$  distribution—which includes a scale parameter  $\sigma_v$ —and that  $v$  is from a truncated-normal or gamma distribution. For now we shall assume the former, so that the probability density functions  $f_v$  and  $f_u$  are given by

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

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

where  $\mu$  and  $\sigma_u$  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  $t$  distribution, and  $\Gamma$  is the gamma function. As noted previously, as  $a \rightarrow \infty$  the  $t$  approaches the normal distribution. Thus, our models nest the normal-truncated normal and normal-gamma models. Similarly, when  $a = 1$  we have a Cauchy distributed noise component. The joint density of  $\varepsilon$  and  $u$  is given by

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

The marginal density of  $\varepsilon$  is given by the convolution

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

This is an integral with no closed form, meaning that it is not possible to give an analytic expression for the log-likelihood function. One solution to this problem is to use simulation to approximate the integral and arrive at a simulated log likelihood function—see Train (2003) for an introduction to maximum simulated likelihood—as proposed by Greene (2003) for the normal-gamma model and Stead et al. (2017) 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

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

That can be estimated by

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

where  $u_q$  is a draw from a truncated-normal distribution. This gives us a simulated probability density function for  $\varepsilon$

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

The simulated log-likelihood function is

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

This may be maximised like any conventional log-likelihood function, provided we have our draws from the truncated normal distribution. The usual method of taking draws from a non-uniform distribution is to note that the cumulative density function of a random variable follows a uniform distribution, and therefore the inverse cumulative density function, i.e. the quantile function, gives the value of a random variable as a function of a uniformly distributed random variable. We can therefore transform draws from the uniform distribution into draws from any given distribution using the quantile function of that distribution. From Geweke et al. (1997) and Greene (2003), we have the quantile function of a truncated normal random variable:

$$u_{iq} = \mu + \sigma_u \Phi^{-1} \left\{ \left[ \Phi\left(\frac{t_r - \mu}{\sigma_u}\right) - \Phi\left(\frac{t_l - \mu}{\sigma_u}\right) \right] F_i + \Phi\left(\frac{t_l - \mu}{\sigma_u}\right) \right\} \quad (12)$$



Where  $t_l$  and  $t_r$  are the left and right truncation points, respectively. Since we know that  $t_l = 0, t_r = \infty$ , this simplifies to

$$u_{iq} = \mu + \sigma_u \Phi^{-1} \left[ 1 + \Phi \left( \frac{\mu}{\sigma_u} \right) (F_{iq} - 1) \right] \quad (13)$$

At this point, it is useful to note that in order to modify the model so that the one-sided error follows some other distribution, we need only change  $u_{iq}$  such that we instead obtain draws from the chosen distribution. The most obvious choices are the exponential and gamma distributions, which are the most widely used one-sided error distributions in SFA after the half normal and truncated normal. For the  $t$ -exponential case, we have the quantile function

$$u_{iq} = -\sigma_u \ln F_{iq} \quad (14)$$

And by substituting these into (12), we have the log-likelihood function for the  $t$ -truncated-normal and  $t$ -exponential models. Other proposed distributions for  $u$ , such as the Weibull (Tsonas, 2007) and Rayleigh (Hajargasht, 2015) distributions also have closed form quantile functions which can likewise substituted into (12). The  $t$ -gamma case is less straightforward in this respect, since there is no analytical expression for the inverse cumulative density function. One way to proceed, however, is to note that the convolution of a  $t$  distributed  $v$  and a gamma distributed  $u$  is with shape parameter  $b$  and scale parameter  $\sigma_u$  is

$$f_\varepsilon(\varepsilon) = \int_0^\infty \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\sqrt{\pi a}\sigma_v} \left[ 1 + \frac{1}{a} \left( \frac{\varepsilon + su}{\sigma_v} \right)^2 \right]^{-\frac{a+1}{2}} \frac{u^{b-1} e^{-\frac{u}{\sigma_u}}}{\Gamma(b)\sigma_u^b} du \quad (15)$$

which is the expectation of  $f_v(\varepsilon + su)(u/\sigma_u)^{b-1}/\Gamma(b)$  where  $u$  is drawn from an exponential distribution:

$$h(u) = E \left[ \frac{f_v(\varepsilon + su)}{\Gamma(b)} \left( \frac{u}{\sigma_u} \right)^{b-1} \middle| u \geq 0 \right], \quad u \sim \text{Exponential}(\sigma_u) \quad (16)$$

Following the reasoning above, therefore, we arrive at the simulated probability density function for  $\varepsilon$ :

$$f_\varepsilon(\varepsilon) = \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\Gamma(b)\sqrt{\pi a}\sigma_v} \frac{1}{Q} \sum_{q=1}^Q (-\ln F_{iq})^{b-1} \left[ 1 + \frac{1}{a} \left( \frac{\varepsilon - s\sigma_u \ln F_{iq}}{\sigma_v} \right)^2 \right]^{-\frac{a+1}{2}} \quad (17)$$

The simulated log-likelihood function is

$$\begin{aligned} \ln SL = & -N \ln Q + N \ln \Gamma\left(\frac{a+1}{2}\right) - N \ln \Gamma\left(\frac{a}{2}\right) - N \ln \Gamma(b) - N \ln \sigma_v \\ & - \frac{N}{2} [\ln \pi + \ln a] + \sum_{i=1}^N \ln \sum_{q=1}^Q (-\ln F_{iq})^{b-1} \left[ 1 + \frac{1}{a} \left( \frac{\varepsilon - \sigma_u \ln F_{iq}}{\sigma_v} \right)^2 \right]^{-\frac{a+1}{2}} \end{aligned} \quad (18)$$

We now have simulated log-likelihood functions for the  $t$ -truncated normal and  $t$ -gamma SFA models, and therefore also for their familiar restrictions, the  $t$ -half normal (when  $\mu = 0$ ) and  $t$ -exponential (when  $b = 1$ ), respectively. First order conditions for maximisation in both models are given in the Appendix. 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, as opposed to random draws: this significantly reduces the number of draws needed to approximate the integral (see Greene (2003) for a fuller discussion).

### 3.2. 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:

$$f(u|\varepsilon) = \frac{f_v(\varepsilon + su)f_u(u)}{f_\varepsilon(\varepsilon)} \quad (19)$$

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

$$\exp[-E(u|\varepsilon)] = \exp\left[-\int_0^\infty u \frac{f_v(\varepsilon + su)f_u(u)}{f_\varepsilon(\varepsilon)} du\right] \quad (20)$$

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

$$E[\exp(-u|\varepsilon)] = \int_0^\infty \exp(-u) \frac{f_v(\varepsilon + su)f_u(u)}{f_\varepsilon(\varepsilon)} du \quad (21)$$

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

$$\exp[-E(u|\varepsilon)] = \exp\left[-\frac{\int_0^\infty u f_v(\varepsilon + su)f_u(u) du}{\int_0^\infty f_v(\varepsilon + su)f_u(u) du}\right] \quad (22)$$

$$E[\exp(-u|\varepsilon)] = \frac{\int_0^\infty \exp(-u) f_v(\varepsilon + su)f_u(u) du}{\int_0^\infty f_v(\varepsilon + su)f_u(u) du} \quad (23)$$

Again, for the models 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  $u f_v(\varepsilon + su)$  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:

$$\exp[-E(u|\varepsilon)] = \exp \left[ - \frac{\sum_{q=1}^Q u_q \left[ 1 + \frac{1}{a} \left( \frac{\varepsilon + su_q}{\sigma_v} \right)^2 \right]^{-\frac{a+1}{2}}}{\sum_{q=1}^Q \left[ 1 + \frac{1}{a} \left( \frac{\varepsilon + su_q}{\sigma_v} \right)^2 \right]^{-\frac{a+1}{2}}} \right] \quad (24)$$

$$E[\exp(-u|\varepsilon)] = \frac{\sum_{q=1}^Q \exp(-u_q) \left[ 1 + \frac{1}{a} \left( \frac{\varepsilon + su_q}{\sigma_v} \right)^2 \right]^{-\frac{a+1}{2}}}{\sum_{q=1}^Q \left[ 1 + \frac{1}{a} \left( \frac{\varepsilon + su_q}{\sigma_v} \right)^2 \right]^{-\frac{a+1}{2}}} \quad (25)$$

where  $u_q$  is given by (13) in the  $t$ -truncated normal and (14) in the  $t$ -exponential models, for example. These expressions become slightly more complex in the  $t$ -gamma case, but are derived in the same way, giving

$$\exp[E(-u|\varepsilon)] = \exp \left[ - \frac{\sum_{q=1}^Q (-\ln F_{iq})^b \left[ 1 + \frac{1}{a} \left( \frac{\varepsilon - s\sigma_u \ln F_{iq}}{\sigma_v} \right)^2 \right]^{-\frac{a+1}{2}}}{\sum_{q=1}^Q (-\ln F_{iq})^{b-1} \left[ 1 + \frac{1}{a} \left( \frac{\varepsilon - s\sigma_u \ln F_{iq}}{\sigma_v} \right)^2 \right]^{-\frac{a+1}{2}}} \right] \quad (26)$$

$$E[\exp(-u|\varepsilon)] = \frac{\sum_{q=1}^Q (-\ln F_{iq})^{b-1} \exp(\sigma_u \ln F_{iq}) \left[ 1 + \frac{1}{a} \left( \frac{\varepsilon - s\sigma_u \ln F_{iq}}{\sigma_v} \right)^2 \right]^{-\frac{a+1}{2}}}{\sum_{q=1}^Q (-\ln F_{iq})^{b-1} \left[ 1 + \frac{1}{a} \left( \frac{\varepsilon - s\sigma_u \ln F_{iq}}{\sigma_v} \right)^2 \right]^{-\frac{a+1}{2}}} \quad (27)$$

### 3.3. Hypothesis Testing

As discussed previously, an attraction of the Student's  $t$  distribution in the current context is that it nests the normal distribution as  $a \rightarrow \infty$ , and therefore a stochastic frontier model in which  $v$  follows a Student's  $t$  distribution nests a model in which  $v$  follows a normal distribution, for any given distribution of  $u$ . 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 thick 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

$$\text{LR} = -2(\ln SL_A - \ln L_0) \quad (28)$$

Where  $\ln SL_A$  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  $\chi^2$  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. In such cases, 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  $\chi_0^2$  and  $\chi_1^2$  distributions, denoted  $\chi_{1,0}^2$ . Economou (2011) applies this result to an analogous problem in survival analysis: that of testing down from a three parameter Burr XII distribution to a two parameter Weibull distribution, which it nests as a ‘frailty’ parameter approaches zero. A further analogue 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  $B \rightarrow \infty$ . Note that under the null hypothesis that  $\sigma_u = 0$ , the model reduces to a regression model with Student's  $t$  errors.

#### **4. Application to Highways Maintenance Costs in England**

In this section, we apply a Student's  $t$ -half normal model to the data set on highway maintenance costs in England used by Stead et al. (2017). 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<sup>1</sup> and consist of costs and cost drivers associated with local authorities' highway maintenance activities.

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). One exception to the marginal cost focus of empirical studies, is a study of 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 data set is updated annually for a new round of analysis, and in this study we use the data set 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

---

<sup>1</sup> See <http://www.nhtnetwork.org/cqc-efficiency-network/home/>

directions, that 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 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. The model is

$$\begin{aligned}
\ln TOTEX = & \beta_0 + \beta_1 \ln URL + \beta_2 \ln RRL + \beta_3 \ln URL^2 + \beta_4 \ln RRL^2 \\
& + \beta_5 \ln URL \ln RRL + \beta_6 \ln TRAFFIC + \beta_7 RDCA + \beta_8 RDCBC \\
& + \beta_9 RDCU + \beta_{10} PROP_{UA} + \beta_{11} PROP_{UB} + \beta_{12} PROP_{UC} \\
& + \beta_{13} PROP_{UU} + \beta_{14} PROP_{RA} + \beta_{15} PROP_{RB} + \beta_{16} PROP_{RC} \\
& + \beta_{17} YEAR + \beta_{18} \ln WAGE + \beta_{19} \ln ROCOSM + \varepsilon
\end{aligned} \tag{29}$$

where *TOTEX* is total expenditure, *PROP<sub>UA</sub>* through to *PROP<sub>RC</sub>* 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 1 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 coded into a modified version of *LIMDEP*, along with formulae for efficiency predictions as discussed in the previous section. The simulations were based upon 100 Halton draws.



**Table 1:** 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
$\beta_0$	16.058	0.092	***	16.035	0.145	***
$\beta_1$ (ln URL)	0.149	0.107		0.127	0.171	
$\beta_2$ (ln RRL)	0.895	0.113	***	0.917	0.179	***
$\beta_3$ (ln URL <sup>2</sup> )	0.236	0.043	***	0.241	0.063	***
$\beta_4$ (ln RRL <sup>2</sup> )	0.082	0.010	***	0.085	0.016	***
$\beta_5$ (ln URL ln RRL)	-0.064	0.028	**	-0.081	0.044	*
$\beta_6$ (ln TRAFFIC)	0.366	0.099	***	0.415	0.154	***
$\beta_7$ (RDCA)	0.432	0.094	***	0.464	0.144	***
$\beta_8$ (RDCBC)	-0.071	0.026	***	-0.071	0.039	*
$\beta_9$ (RDCU)	-0.004	0.003		-0.005	0.005	
$\beta_{10}$ (PROP <sub>UA</sub> )	8.690	1.941	***	7.810	3.241	**
$\beta_{11}$ (PROP <sub>UB</sub> )	1.642	2.279		0.662	3.869	
$\beta_{12}$ (PROP <sub>UC</sub> )	0.273	1.196		0.448	2.054	
$\beta_{13}$ (PROP <sub>UU</sub> )	0.969	0.547	*	1.090	0.835	
$\beta_{14}$ (PROP <sub>RA</sub> )	2.612	1.045	**	2.120	1.571	
$\beta_{15}$ (PROP <sub>RB</sub> )	2.417	1.056	**	2.678	1.544	*
$\beta_{16}$ (PROP <sub>RC</sub> )	1.015	0.641		0.983	0.988	
$\beta_{17}$ (YEAR)	0.038	0.011	***	0.045	0.017	***
$\beta_{18}$ (ln WAGE)	0.779	0.223	***	0.891	0.340	***
$(1 - \beta_{18})$ (ln ROCOSM)	0.221	-	-	0.109	-	-
$\sigma_u$	0.535	0.046	-	0.568	0.074	-
$\sigma_v$	0.233	0.016	-	0.276	0.043	-
$a$	5.198	1.510	-	$\infty$	-	-
Log Likelihood	-186.06	-	-	-189.14	-	-

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

We can see from Table 1 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. Note that while the  $\sigma_u$  parameters are comparable between the two models, the  $\sigma_v$  parameters are not, since the distribution of  $v$  varies. The variance of  $u$  is given in both cases by:

$$\text{VAR}(u) = \frac{\pi - 2}{\pi} \sigma_u^2 \quad (30)$$

While the variance of  $v$  is, in the  $t$ -half normal and normal-half normal cases respectively:

$$\text{VAR}(v) = \sigma_v^2 \frac{a}{a - 2} \quad (31)$$

$$\text{VAR}(v) = \sigma_v^2 \quad (32)$$

Our estimates of these are compared in below.

**Table 2:** Estimated error variances

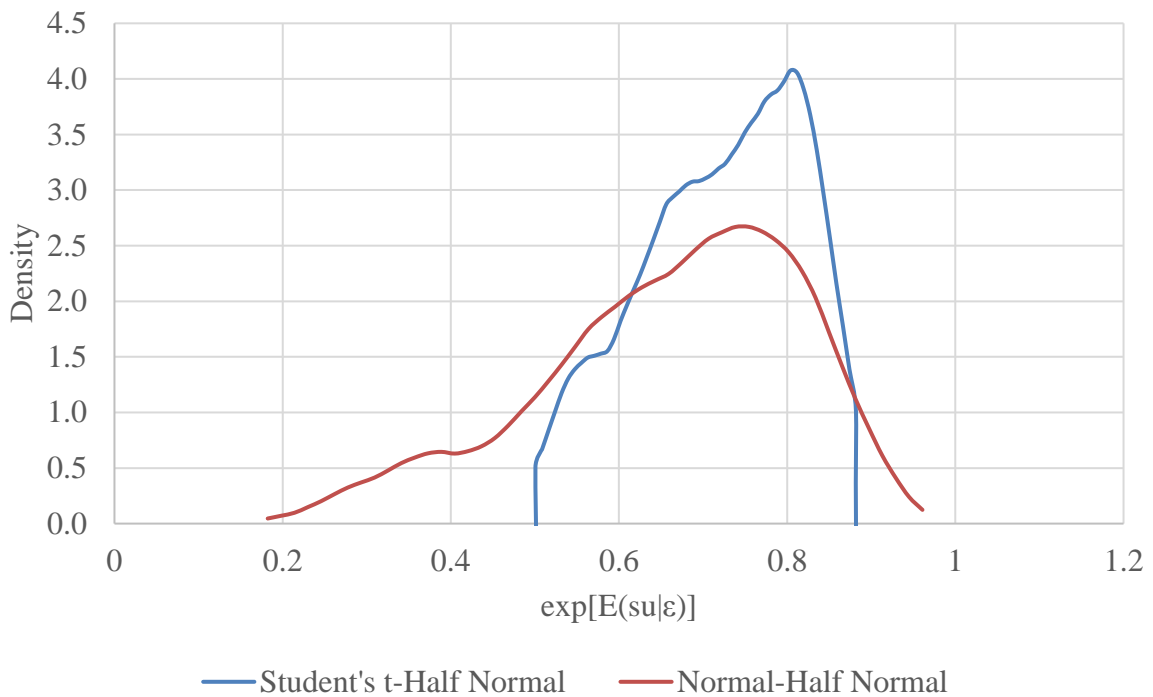
	Student's $t$ -Half Normal	Normal-half normal
$\text{VAR}(u)$	0.104	0.117
$\text{VAR}(v)$	0.088	0.076
$\text{VAR}(\varepsilon)$	0.192	0.194

Table 2 shows that, in line with our expectations, and the findings of Stead et al. (2017) 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, according to the discussion in Section 1, we expect to find a considerably narrower distribution of efficiency predictions from the  $t$ -half normal model owing to both the lower estimated  $\text{VAR}(u)$  and the greater shrinkage towards the mean resulting from the higher estimated  $\text{VAR}(v)$ .

**Table 3:** Summary of efficiency scores

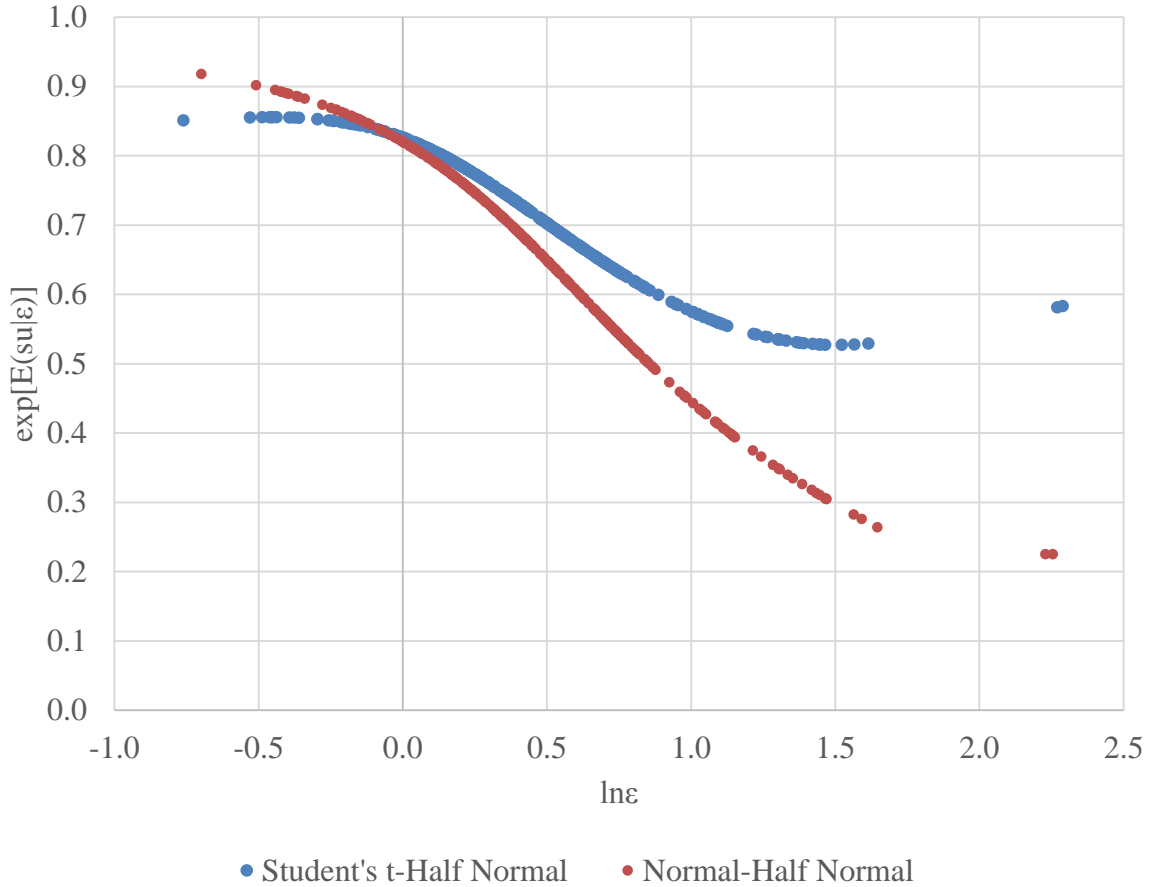
	Student's $t$ -Half Normal	Normal-half normal
Mean	0.721	0.660
Minimum	0.527	0.225
Maximum	0.855	0.918

Table 3 compares the mean, minimum and maximum efficiency scores from the  $t$ -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 Figure 1, which compares the kernel density estimates for the two sets of efficiency scores.



**Figure 1:** Kernel densities of cost 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  $t$ -half normal and normal-half normal models.



**Figure 2:** Scatterplot of cost efficiency scores vs residuals

Given the similarity of the frontier parameter estimates, as shown in Table 1, the residuals from the  $t$ -half normal and normal-half normal models are highly correlated. However the relationships between the residuals and efficiency predictions is shown by Figure 2 to be substantially different between the two models: for values of  $\ln \varepsilon$  between around -0.5 and 1.5, the slope of the function is considerably flatter in the  $t$ -half normal case, so that a change in  $\varepsilon$  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  $t$ -half normal case, with  $\exp[-E(u|\varepsilon)]$  beginning to decrease slightly for the smallest values of  $\varepsilon$  and increase very considerably for the largest values of  $\varepsilon$ . This is in contrast to the Standard SF model, and also the Logistic-half normal model as shown by Stead et al. (2017), in which the relationship is monotonic. The explanation

for this non-monotonicity in the  $t$ -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  $\nu$  to such an extent that there is a reduction in  $\exp[-E(u|\varepsilon)]$ .

As discussed in Section 3.3, we are interested in testing two null hypotheses: first, that  $\nu$  is normally distributed, in which case the  $t$ -half normal nests the normal-half normal model, and second that there is no inefficiency, and the likelihood ratio follows a  $\chi^2_{1,0}$  distribution in both cases. Log-likelihoods are given in Table 1, and are used to calculate the likelihood ratio statistic as shown in ( 28 ) for the first null hypothesis: this gives a likelihood ratio of 4.155 and a corresponding p-value of 0.012. For our second null hypothesis, the  $t$ -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, so in this case we have a likelihood ratio of 4.150 and a corresponding p-value of 0.012. We therefore reject the null hypotheses of normally distributed  $\nu$  and zero inefficiency, indicating that this model performs better than either the standard SF model or the Student's  $t$  regression model in this instance.

## 5. Summary and Conclusions

This paper proposes a new stochastic frontier model as means to account for outliers in the context of SFA. We have reviewed possible methods in the existing literature, including the adopting alternative distributional assumptions for  $\nu$ . Our model develops the approach of Stead et al. (2017), in which maximum simulated likelihood methods are used to estimate a model combining a logistic distribution for  $\nu$  with a half normal distribution for  $u$ , by using a Student's  $t$  distribution for  $\nu$ . The advantages of this distribution is that the kurtosis of  $\nu$  is determined by a degrees of freedom parameter which is freely estimated, and that it nests 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 log-likelihoods and efficiency predictors for the  $t$ -truncated normal and  $t$ -gamma cases, and discuss extension to other distributions for  $u$ , which is straightforward.

We apply a  $t$ -half normal model to estimate a cost frontier using a data set 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. The main differences are a reduced estimate of  $VAR(u)$  and an increased estimate of  $VAR(v)$ , a reduced range of efficiency predictions according to  $\exp[-E(u|\varepsilon)]$ , and a non-monotonic relationship between  $\varepsilon$  and  $\exp[-E(u|\varepsilon)]$ , in contrast to the standard SF model. We discuss testing against the standard SF model, and find that we are able to reject the null hypotheses of normally distributed  $v$  and no inefficiency. This suggests that it is important to account for heavy tails in our data.

## 6. References

- Aigner, D. et al. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*. **6**(1), pp.21-37.
- Almanidis, P. et al. 2014. Stochastic Frontier Models with Bounded Inefficiency. In: Sickles, R.C. and Horrace, W.C. eds. *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications*. New York, NY: Springer New York, pp.47-81.
- Arnold, B.C. and Beaver, R.J. 2000. The skew-Cauchy distribution. *Statistics & Probability Letters*. **49**(3), pp.285-290.
- Azzalini, A. 1985. A Class of Distributions Which Includes the Normal Ones. *Scandinavian Journal of Statistics*. **12**(2), pp.171-178.
- Azzalini, A. and Capitanio, A. 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. **65**(2), pp.367-389.
- Bak, M. and Borkowski, P. 2009. *Marginal Cost of Road Maintenance and Renewal in Poland, CATRIN (Cost Allocation of TRansport INfrastructure cost) Deliverable D6, Annex 2*. Leeds, UK: ITS, University of Leeds.

- Bak, M. et al. 2006. *Marginal Infrastructure Cost in Poland, Marginal Cost Case Studies for Road and Rail Transport Deliverable D3, Annex 1.2C*. Leeds, UK: ITS, University of Leeds.
- Battese, G.E. and Coelli, T.J. 1988. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*. **38**(3), pp.387-399.
- Battese, G.E. and Coelli, T.J. 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*. **20**(2), pp.325-332.
- Bera, A.K. and Sharma, S.C. 1999. Estimating Production Uncertainty in Stochastic Frontier Production Function Models. *Journal of Productivity Analysis*. **12**(3), pp.187-210.
- Caudill, S.B. and Ford, J.M. 1993. Biases in frontier estimation due to heteroscedasticity. *Economics Letters*. **41**(1), pp.17-20.
- Caudill, S.B. et al. 1995. Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity. *Journal of Business & Economic Statistics*. **13**(1), pp.105-111.
- Charnes, A. et al. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research*. **2**(6), pp.429-444.
- Chen, Y. and Liang, K.-Y. 2010. On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems. *Biometrika*. **97**(3), pp.603-620.
- Christensen, L.R. and Greene, W.H. 1976. Economies of Scale in U.S. Electric Power Generation. *Journal of Political Economy*. **84**(4), pp.655-676.
- Cook, R.D. 1977. Detection of Influential Observation in Linear Regression. *Technometrics*. **19**(1), pp.15-18.
- Economou, P. 2011. On model selection in the case of nested distributions — An application to frailty models. *Statistical Methodology*. **8**(2), pp.172-184.
- Fallah-Fini, S. et al. 2009. Performance measurement of highway maintenance operation using data envelopment analysis: Environmental considerations. In: *IIE Annual Conference. Proceedings*: Institute of Industrial Engineers-Publisher, p.693.
- Geweke, J.F. et al. 1997. Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics*. **80**(1), pp.125-165.
- Greene, W.H. 2003. Simulated Likelihood Estimation of the Normal-Gamma Stochastic Frontier Function. *Journal of Productivity Analysis*. **19**(2), pp.179-190.
- Griffin, J.E. and Steel, M.F.J. 2007. Bayesian stochastic frontier analysis using WinBUGS. *Journal of Productivity Analysis*. **27**(3), pp.163-176.
- Grushka, E. 1972. Characterization of exponentially modified Gaussian peaks in chromatography. *Analytical Chemistry*. **44**(11), pp.1733-1738.
- Gupta, A.K. and Nguyen, N. 2010. Stochastic frontier analysis with fat-tailed error models applied to WHO health data. *International Journal of Innovative Management, Information & Production*. **1**(1), pp.43-48.
- Hadri, K. 1999. Estimation of a Doubly Heteroscedastic Stochastic Frontier Cost Function. *Journal of Business & Economic Statistics*. **17**(3), pp.359-363.
- Hajargasht, G. 2015. Stochastic frontiers with a Rayleigh distribution. *Journal of Productivity Analysis*. **44**(2), pp.199-208.
- 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*. Leeds, UK: ITS, University of Leeds.
- Hofler, R.A. 2014. A Generalized Normal-Half Normal Stochastic Frontier Model. In: *North American Productivity Workshop VIII, 4th-7th June 2014, Ottawa, Canada*.

- Horrace, W.C. and Parmeter, C.F. 2015. A Laplace Stochastic Frontier Model. *Econometric Reviews*. pp.1-27.
- Horrace, W.C. and Schmidt, P. 1996. Confidence statements for efficiency estimates from stochastic frontier models. *Journal of Productivity Analysis*. **7**(2), pp.257-282.
- Jondrow, J. et al. 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*. **19**(2), pp.233-238.
- Jonsson, L. and Haraldsson, M. 2008. *Marginal Costs of Road Maintenance in Sweden, CATRIN (Cost Allocation of TRansport INfrastructure cost) Deliverable D6, Annex 1*. Stockholm, Sweden: VTI.
- Keynes, J.M. 1911. The Principal Averages and the Laws of Error which Lead to Them. *Journal of the Royal Statistical Society*. **74**(3), pp.322-331.
- Lange, K. and Sinsheimer, J.S. 1993. Normal/Independent Distributions and Their Applications in Robust Regression. *Journal of Computational and Graphical Statistics*. **2**(2), pp.175-198.
- Lange, K.L. et al. 1989. Robust Statistical Modeling Using the t Distribution. *Journal of the American Statistical Association*. **84**(408), pp.881-896.
- Lee, Y.H. 1996. Tail truncated stochastic frontier models. *Journal of Economic Theory and Econometrics*. **2**, pp.137-152.
- Link, H. 2006. An econometric analysis of motorway renewal costs in Germany. *Transportation Research Part A: Policy and Practice*. **40**(1), pp.19-34.
- Link, H. 2009. *Marginal Costs of Road Maintenance in Germany, CATRIN (Cost Allocation of TRansport INfrastructure) Deliverable D6, Annex 3*. Stockholm, Sweden: VTI.
- Link, H. 2014. A Cost Function Approach for Measuring the Marginal Cost of Road Maintenance. *Journal of Transport Economics and Policy (JTEP)*. **48**(1), pp.15-33.
- Mahalanobis, P.C. 1936. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (India)*. **2**, pp.49-55.
- Meeusen, W. and van Den Broeck, J. 1977. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review*. **18**(2), pp.435-444.
- Nadarajah, S. 2005. A generalized normal distribution. *Journal of Applied Statistics*. **32**(7), pp.685-694.
- Nguyen, N. 2010. *Estimation of Technical Efficiency in Stochastic Frontier Analysis*. PhD thesis, Bowling Green State University.
- Reifschneider, D. and Stevenson, R. 1991. Systematic Departures from the Frontier: A Framework for the Analysis of Firm Inefficiency. *International Economic Review*. **32**(3), pp.715-723.
- Schreyer, C. et al. 2002. *Road Econometrics—Case Study Motorways Switzerland. UNITE (UNification of accounts and marginal costs for Transport Efficiency) Deliverable 10, Annex A1b*. Leeds, UK: ITS, University of Leeds.
- Sedlacek, N. and Herry, M. 2002. *Infrastructure Cost Case Studies. UNITE (UNification of accounts and marginal costs for Transport Efficiency) Deliverable 10, Annex A1c*. Leeds, UK: ITS, University of Leeds.
- Self, S.G. and Liang, K.-Y. 1987. Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical Association*. **82**(398), pp.605-610.
- Stevenson, R.E. 1980. Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics*. **13**(1), pp.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. and Dey, D.K. 2007. Bayesian Estimation of Stochastic Frontier Models with Multivariate Skew t Error Terms. *Communications in Statistics - Theory and Methods*. **36**(5), pp.907-916.
- Train, K.E. 2003. *Discrete Choice Methods with Simulation*. Second Edition (2009) ed. Cambridge University Press.
- Tsionas, E.G. 2007. Efficiency Measurement with the Weibull Stochastic Frontier. *Oxford Bulletin of Economics and Statistics*. **69**(5), pp.693-706.
- Wang, H.-J. 2002. Heteroscedasticity and Non-Monotonic Efficiency Effects of a Stochastic Frontier Model. *Journal of Productivity Analysis*. **18**(3), pp.241-253.
- Wang, W.S. and Schmidt, P. 2009. On the distribution of estimated technical efficiency in stochastic frontier models. *Journal of Econometrics*. **148**(1), pp.36-45.
- Wheat, P. et al. 2014. Understanding prediction intervals for firm specific inefficiency scores from parametric stochastic frontier models. *Journal of Productivity Analysis*. **42**(1), pp.55-65.

## 7. Appendix

First order conditions for maximisation of the  $t$ -truncated normal log likelihood are

$$\frac{\partial \ln SL}{\partial a} = \frac{N}{2} \left[ \psi \left( \frac{a+1}{2} \right) - \psi \left( \frac{a}{2} \right) \right] + \frac{1}{\sum_{q=1}^Q g_i(u_{iq})} \sum_{i=1}^N \sum_{q=1}^Q \frac{\partial g(u_{iq})}{\partial a} = 0 \quad (33)$$

$$\frac{\partial \ln SL}{\partial \beta} = \frac{1}{\sum_{q=1}^Q g_i(u_{iq})} \sum_{i=1}^N \sum_{q=1}^Q \frac{\partial g(u_{iq})}{\partial \varepsilon_i} \frac{\partial \varepsilon_i}{\partial \beta} = 0 \quad (34)$$

$$\frac{\partial \ln SL}{\partial \sigma_v} = -\frac{1}{\sigma_v} + \frac{1}{\sum_{q=1}^Q g_i(u_{iq})} \sum_{i=1}^N \sum_{q=1}^Q \frac{\partial g(u_{iq})}{\partial \sigma_v} = 0 \quad (35)$$

$$\frac{\partial \ln SL}{\partial \sigma_u} = \frac{1}{\sum_{q=1}^Q g_i(u_{iq})} \sum_{i=1}^N \sum_{q=1}^Q \frac{\partial g(u_{iq})}{\partial u_{iq}} \frac{\partial u_{iq}}{\partial \sigma_u} = 0 \quad (36)$$

$$\frac{\partial \ln SL}{\partial \mu} = \frac{1}{\sum_{q=1}^Q g_i(u_{iq})} \sum_{i=1}^N \sum_{q=1}^Q \frac{\partial g(u_{iq})}{\partial u_{iq}} \frac{\partial u_{iq}}{\partial \mu} = 0 \quad (37)$$

In the  $t$ -gamma case, we have the first-order conditions

$$\frac{\partial \ln SL}{\partial a} = \frac{N}{2} \left[ \psi \left( \frac{a+1}{2} \right) - \psi \left( \frac{a}{2} \right) - \ln a \right] + \frac{1}{\sum_{q=1}^Q u_{iq}^{b-1} g_i(u_{iq})} \sum_{i=1}^N \sum_{q=1}^Q u_{iq}^{b-1} \frac{\partial g(u_{iq})}{\partial a} = 0 \quad (38)$$

$$\frac{\partial \ln SL}{\partial \beta} = -N\psi(\beta) + \frac{1}{\sum_{q=1}^Q u_{iq}^{b-1} g_i(u_{iq})} \sum_{i=1}^N \sum_{q=1}^Q u_{iq}^{b-1} \frac{\partial g(u_{iq})}{\partial \varepsilon_i} \frac{\partial \varepsilon_i}{\partial \beta} = 0 \quad (39)$$

$$\frac{\partial \ln SL}{\partial \sigma_u} = N(b-1) + \frac{\sum_{i=1}^N \sum_{q=1}^Q u_{iq}^{b-1} \frac{\partial u_{iq}}{\partial \sigma_u} \left[ (b-1) \frac{g_i(u_{iq})}{u_{iq}} + \frac{\partial g(u_{iq})}{\partial u_{iq}} \right]}{\sum_{q=1}^Q u_{iq}^{b-1} g_i(u_{iq})} = 0 \quad (40)$$

$$\frac{\partial \ln SL}{\partial b} = N \left[ \frac{1}{b} - \psi(b) \right] + \frac{(b-1)}{\sum_{q=1}^Q u_{iq}^{b-1} g_i(u_{iq})} \sum_{i=1}^N \sum_{q=1}^Q u_{iq}^{b-2} g_i(u_{iq}) = 0 \quad (41)$$

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

$$\frac{\partial g(u_{iq})}{\partial a} = -\frac{1}{2} \left[ 1 + \frac{1}{a} \left( \frac{\varepsilon + su_{iq}}{\sigma_v} \right)^2 \right] \left\{ \ln \left[ 1 + \frac{1}{a} \left( \frac{\varepsilon + su_{iq}}{\sigma_v} \right)^2 \right] - \frac{(a+1) \left( \frac{x}{\sigma} \right)^2}{2a^2 \left[ 1 + \frac{1}{a} \left( \frac{x}{\sigma} \right)^2 \right]} \right\} \quad (42)$$

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

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

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

$$\frac{\partial u_{iq}}{\partial \mu} = 1 + \frac{\phi \left( \frac{\mu}{\sigma_u} \right) (F_{iq} - 1)}{\phi \left\{ \Phi^{-1} \left[ 1 + \Phi \left( \frac{\mu}{\sigma_u} \right) (F_{iq} - 1) \right] \right\}} \quad (46)$$

In the  $t$ -truncated normal and  $t$ -gamma cases, respectively, we have:

$$\frac{\partial u_{iq}}{\partial \sigma_u} = \Phi^{-1} \left[ \Phi \left( \frac{\mu}{\sigma_u} \right) F_{iq} + \frac{(1+s)}{2} \Phi \left( \frac{-\mu}{\sigma_u} \right) \right] + \frac{\frac{\mu}{\sigma_u^2} \phi \left( \frac{\mu}{\sigma_u} \right) (1 - F_{iq})}{\phi \left\{ \Phi^{-1} \left[ 1 + \Phi \left( \frac{\mu}{\sigma_u} \right) (F_{iq} - 1) \right] \right\}} \quad (47)$$

$$\frac{\partial u_{iq}}{\partial \sigma_u} = \sigma_u^2 \ln F_{iq} \quad (48)$$