Accounting for Heavy Tails in Stochastic Frontier Models

Andrea Tancredi *

Università di Roma "La Sapienza"

Summary

This paper aims at introducing a new class of stochastic frontier models that can take account for fat tails in the composed error. Quite surprisingly, at least to our knowledge, all the stochastic frontier models proposed in literature cannot handle situations where the empirical distribution of the composed error has heavy tails. These situations are instead very common in applications. In particular, we will propose to model the composed error with the skew-t distribution. This is equivalent to assume a Student-t distribution for the measurement error and a half-t distribution for the inefficiency. In this way, we extend quite naturally, the stochastic frontier model where a normal distribution is assumed for the symmetric error and a half-normal distribution is assumed for the inefficiency term.

*Dipartimento di Studi Geoeconomici Statistici e Storici per l'Analisi Regionale, Università di Roma "La Sapienza", Via del Castro Laurenziano 9, 00161, Roma, Italy; Phone: +39-0649766231; Fax: +39-064957606; E-mail andrea.tancredi@uniroma1.it. This research has been supported by MIUR, under grant scheme PRIN 2000: Extensions of the Gaussian distribution and their applications. This paper has been written while the author was research associate at the Department of Statistics, University of Padova. The author thanks Adelchi Azzalini for many interesting and fruitful discussions.

Some key words: Composed error; Efficiency analysis; Skew-t distribution

1 Introduction

Stochastic frontier models, introduced in Meeusen and van den Broeck (1977) and Aigner *et. al.* (1977), are useful tools to evaluate the efficiency of economic agents, such as firms, individuals or countries. The principle underlying the model is that the observed production of a single unit cannot exceed the unobserved potential production i.e. the frontier, which is the maximum possible production given input quantities. The difference between the frontier and the observed production is a measure of inefficiency. Such difference is modelled by a one-sided random variable. The model is completed adding a symmetric random variable capturing the mesaurement error of the frontier.

In the commonly adopted formulation a normal distribution is assumed for the symmetric error while a half-normal distribution is assumed for the one-sided random variable. This framework leads to model the composed error by a probability distribution known, in the statistical literature, as skew-normal distribution, see Azzalini (1985) and Azzalini and Capitanio (1999). The skew-normal distribution is a family of distributions including the normal, but with an extra parameter to regulate the skewness. The stochastic frontier model introduced by Aigner *et. al* (1977), ALS model in the following, is a skew normal regression models with negative skewness.

Several other distributions for the inefficiency term, have been proposed, in different times, in place of the half-normal distribution. For example, Meeusen and van den Broeck (1977) adopt the exponential distribution, Stevenson (1980) the truncated normal and Greene (1990) the gamma distribution. A unified approach is proposed in van den Broeck *et al* (1994) where the results obtained with different inefficiency distributions are pooled together by Bayesian model averaging. Finally, a semi-parametric Bayesian approach is proposed in Griffin and Steel (2002) where a Dirichelet process with gamma mean is assumed for the inefficiency distribution.

This paper aims at introducing a new class of stochastic frontier models that can take account for fat tails in the composed error. Quite surprisingly, all the proposals existing in literature, at least to our knowledge, cannot handle situations where the empirical distribution of the composed error has heavy tails which, instead, are very common in applications. In particular, we will propose to model the composed error with the univariate skew-t distribution introduced by Branco and Day (2001) and by Azzalini and Capitanio (2002). This is equivalent to assume a Student-t distribution for the measurement error and a half-t distribution for the inefficiency. In this way, we extend quite naturally, the ALS model, which becomes a limit case of our model.

Section 2 describes the new model with particular emphasis on frequentist inference both for testing the presence of the inefficiency term and for estimating individual technical efficiencies. In particular it is shown that, in estimating individual efficiencies, the skew-t model has a completely different approach with respect to the ALS model when we have observations which are suspected of being outliers. In the ALS model, observations with a large postive deviation from the estimated frontier lead to estimates of individual efficiency concentrated in one. Instead, for the skew-t model these observations are not considered informative for estimating individual efficiency.

In section 3 we apply the model to the well known data set of the American elettrical companies. This data set has been carefully analyzed, in a frequentist setting, by Ritter and Simar (1994), who conclude that the data show more evidence for the normal linear model without inefficiency than for stochastic frontier models. We will show that assuming a skew-t model for the composed error provides a more reasonable fit than the normal linear model and that taking account for fat tails increases the evidence for the inefficiency term. Moreover we will study the behaviour of the estimates of individual efficiencies respect to the distance from the estimated frontier both for the ALS model and the skew-t model.

In section 4 we give a brief discussion for subsequent modelling and research.

2 The model

Let us consider the model

$$y_i = h(x_i, \beta) + \epsilon_i - z_i \qquad i = 1, \ldots, n \tag{1}$$

where y_i denotes the log of the output variable for firm i ($i = 1, \ldots, n$) and x_i is a a vector of observations for the explanatory variables for firm i. In stochastic frontier models ϵ_i is a symmetric distribution with zero mean while z_i is a one-sided positive distribution. For istance, in the ALS model ϵ_i is assumed $\mathcal{N}(0, \sigma_\epsilon^2)$ and z_i is assumed half-normal distributed $|\mathcal{N}(0, \sigma_z^2)|$.

Respect to the ALS model we assume that $z_i = |v_i|$ and that the couples ϵ_i, v_i are distributed like a bivariate Student-t distribution with zero means, scale parameters σ_ϵ and σ_z, uncorrelated components and shape parameter ν, independently for $i = 1, \ldots, n$. The density can be written as

$$f(\epsilon_i, v_i) = \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \nu \, \pi \, \sigma_\epsilon \sigma_z} \left\{ 1 + \frac{1}{\nu} \left[\left(\frac{\epsilon_i}{\sigma_\epsilon}\right)^2 + \left(\frac{v_i}{\sigma_z}\right)^2 \right] \right\}^{-\frac{\nu+2}{2}} \qquad i = 1, \ldots, n. \tag{2}$$

The meaning of the variables ϵ_i and z_i remains the same of standard stochastic frontier models. In fact ϵ_i still represents a symmetric disturbance capturing the measurement error of the stochastic frontier and z_i is still a nonnegative random variable modelling the level of inefficiency. The novelty here is that ϵ_i is marginally distributed like a univariate Student-t distribution $\mathcal{T}(0, \sigma_\epsilon, \nu)$ and z_i is marginally distributed like a half Student-t $|\mathcal{T}(0, \sigma_z, \nu)|$.

The sampling distribution of y_i has been derived, in a more general context, by Azzalini and Capitanio (2002). Specifically, if we write

$$y_i = h(x_i, \beta) + \sqrt{\sigma_\epsilon^2 + \sigma_z^2}\left[\frac{\sigma_\epsilon}{\sqrt{\sigma_\epsilon^2 + \sigma_z^2}}U - \frac{\sigma_z}{\sqrt{\sigma_\epsilon^2 + \sigma_z^2}}|U_0|\right], \tag{3}$$

where (U_0, U) is a standard bivariate Student-t distribution with shape parameter ν, then we can apply proposition 9 of Azzalini and Capitanio (2002). Thus we have that

$$\begin{aligned} p(y_i; \beta, \sigma_\epsilon, \sigma_z, \nu) &= 2\, f_t(y_i; h(x_i, \beta), \omega, \nu) \\ &\quad \times T_1\left[\alpha\, \frac{y_i - h(x_i, \beta)}{\omega}\left(\frac{\nu + 1}{\omega^{-2}(y_i - h(x_i, \beta))^2 + \nu}\right)^{1/2}; \nu + 1\right] \end{aligned} \tag{4}$$

where $f_t(y; h(x_i, \beta), \omega, \nu)$ denotes the density function of a Student-t distribution with mean $h(x_i, \beta)$, scale ω and ν degrees of freedom, $\omega = \sqrt{\sigma_\epsilon^2 + \sigma_z^2}$, $T_1(y, \nu + 1)$ denotes the scalar Student-t distribution with $\nu + 1$ degrees of freedom and $\alpha = -\sigma_z/\sigma_\epsilon$. Distributions with density (4) are called skew-t distributions and they generalize the skew-normal distributions (Azzalini, (1985)) which can be obtained when the shape parameter ν goes to infinity.

Note that the same mechanism that generates a Student-t distribution from a normal distribution allow us to generate a skew-t from a skew-normal. In fact Azzalini and Capitanio (2002) show that a skew-t distribution can be obtained as mixture of skew-normal variates with scale parameter $1/\sqrt{\lambda}$ where λ is gamma distributed $\Gamma(\nu/2, \nu/2)$. Thus model (1) can be written also in the following way

$$y_i = h(x_i, \beta) + \frac{1}{\sqrt{\lambda_i}}(\epsilon_i - z_i) \qquad i = 1, \ldots, n \tag{5}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$, $z_i = |\mathcal{N}(0, \sigma_z^2)|$ and $\lambda_i \sim \Gamma(\nu/2, \nu/2)$. Indeed, the difference $\epsilon_i - z_i$ in (5), which is the composed error of the ALS model, follows a skew-normal distribution with location parameter equal to zero, scale parameter equal to $\sqrt{\sigma_\epsilon^2 + \sigma_z^2}$ and shape parameter equal to $-\sigma_z/\sigma_\epsilon$.

2.1 Testing the presence of the inefficiency term

The natural use of model (4) is whitout constraining the parameter α to be negative. In fact, if we let α to vary in $(-\infty, \infty)$ the statistical model (4) is a regression model that can account for fat tails and for both positive and negative skewness of the error distribution. Positive values for α correspond to composed error distributions where a half-t is added to a Student-t, switching drasticallly the meaning of the model respect to stochastic frontier models. Anyway considering the model (4) with $\alpha \in (-\infty, \infty)$ can be helpful for inferential aims also in stochastic frontier models. Specifically, in testing $H_0 : \alpha = 0$ against $H_1 : \alpha < 0$, i.e. in testing the presence of the ineffieciency term, the signed version of the likelihood ratio statistic may be used. That is

$$R = sgn(\hat{\alpha})\{2[\ell(\hat{\theta}) - \ell(\hat{\theta}^*)]\}^{1/2} \tag{6}$$

where $\theta = (\beta, \omega, \alpha)$, $\hat{\alpha}$ and $\hat{\theta}$ denote the maximum likehood estimates of α and θ when $\alpha \in (-\infty, \infty)$ and $\hat{\theta}^*$ denote the maximum likelihood estimate when $\alpha = 0$. The asymptotic distribution of R under the null model is standard normal and having observed R_{obs} the evidence of H_0 against H_1 is given by $P(\mathcal{N}(0,1) < R_{obs})$.

2.2 Estimation of firm-level technical efficiencies

In stochastic frontier models the main interst is not on the parameters themselves, but in the individual technical efficiencies, measured by $r_i = \exp(-z_i)$. Estimates of these quantities are obtained considering the conditional expected values $E(r_i|y_i)$, see for example Coelli *et al.* (1998). In the appendix we prove that for model (1) the conditional density of z_i given y_i is

$$f(z_i \,|\, y_i) = \frac{\left(\frac{\nu}{2}\right)^{\nu/2} \Gamma\left(\frac{\nu+2}{2}\right)}{f(y_i)\Gamma\left(\frac{\nu}{2}\right) \pi \sqrt{\sigma_\epsilon^2 \sigma_z^2} \left[\frac{1}{2}\left\{\nu + \frac{(y_i - h_i)^2}{\sigma_\epsilon^2 + \sigma_z^2} + \frac{\sigma_\epsilon^2 + \sigma_z^2}{\sigma_\epsilon^2\, \sigma_z^2}\left(z_| + (y_i - h_i)\frac{\sigma_z^2}{\sigma_\epsilon^2 + \sigma_z^2}\right)^2\right\}\right]^{\frac{\nu+2}{2}}}. \tag{7}$$

Thus, a point estimation of the individual technical efficiency can be obtained integrating numerically $\int e^{-z_i} f(z_i|y_i) d\,z_i$ where in $f(z_i|y_i)$ we replace the unknown parameters with the maximum likelihood estimates. A measure of uncertainty of these estimates is obtained considering the plug-in estimates of the standard deviation of e^{-z_i} given y_i. Note that, in this ways, we do not take account of parameter uncertainty, but, from a frequentist point of view, this seems the standard practice since Jondrow *et al.* (1982).

Let us observe that adopting the skew-t model in place of the ALS model leads to a completely different behaviour in estimating individual efficiency of firms with large positive deviations from the estimated frontier. In fact, if we indicate the frontier $h(x_i, \beta)$ with h_i we have that (see the appendix) for the ALS model

$$\lim_{y_i - h_i \to \infty} f(z_i|y_i) = \begin{cases} \infty & if \quad z_i = 0 \\ 0 & if \quad z_i > 0 \end{cases} \tag{8}$$

while for the skew-t model

$$\lim_{y_i - h_i \to \infty} f(z_i|y_i) = 0 \qquad\qquad z_i \geq 0. \tag{9}$$

This means that when $y_i - h_i$ goes to infinity the conditional distribution of z_i in the ALS model is concentrated in zero while for the skew-t model is improper uniform on $[0, \infty)$.

Now let us suppose to have an observation y_i which produces an estimated positive residual $y_i - h(x_i, \hat{\beta})$ very far from the bulk of the other residuals. Thus we should consider y_i as an outlier and, maybe, we should remove it from the data before calculating the required inference. Now, if we are fitting the data with the ALS model we have, for this observation, that the plug-in estimate of r_i will be near one and the plug-in estimates of $Var(e^{-z_i}|y_i)$ near zero. In fact, by (8) the distribution of e^{-z_i} given y_i will be very concentrated near one. Thus we should believe that this observation is very informative for estimating individual technical efficiency. The situation is completely different with the skew-t model. In fact when we have an observation which produces a high estimated residual, the plug-in estimate of $Var(e^{-z_i}|y_i)$ will be very high. This is because the limit distribution of $z_i|y_i$ is completely flat. Thus, this observation will not considered very informative for estimating individual technical efficiency. This means that, when our primary goal is to estimate individual efficiencies, by adopting the skew-t model we do not have to worry whether to discard an outlier or not. In fact the model automatically increases the degree of uncertainty of our conclusions when we are in presence of outliers.

3 Example

We consider the data collected by Christensen and Greene (1976) for 123 electric utility companies in the US in 1970. The data are given in the appendix to Greene (1990) and have been used by van den Brock *et al* (1994) and Tsionas (2002). There are three production factors labor, capital and fuel with prices p_L, p_K, and p_F and the cost function which is usually specified is

$$y_i = -\beta_0 - \beta_1 \ln Q_i - \beta_2 \ln^2 Q_i - \beta_3 \ln \frac{p_{K_i}}{p_{F_i}} - \beta_4 \ln \frac{p_{L_i}}{p_{F_i}} + \epsilon_i - v_i \qquad (10)$$

where $y_i = -\ln C_i/p_{F_i}$, Q_i is the output and C_i the cost of the ith firm.

3.1 Normal against skew-t model

For this data set, Ritter and Simar (1994) compare, from a likelihood point of view, a normal regression model with several stochastic frontier models. They conclude that a normal linear model without inefficiency is enough to explain the data. Thus we firstly try to understand if a skew-t model lead to a significantly improved explanation of the data over a standard regression model.

Maximum likelihood estimates (MLE) and approximated 95% confidence interval are reported in Tab 1 for the normal linear model, the ALS model and the skew-t regression model. We observe that the maximum log-likelihood for the normal model is 65.67, while for the skew-t regression model we have obtained 68.75. Thus the test statistic $D = \ell(\hat{\theta}) - \ell(\hat{\theta}^*)$, where $\ell(\hat{\theta})$ and $\ell(\hat{\theta}^*)$ denote the maximized log-likelihood within the skew-t regression model and the normal regression model, is 6.12. Anyway, comparing the two models by the test statistic D needs some caution. In fact the normal regression model occurs on the frontier of the parametric space of the skew-t regression model, specifically when ν tends to infinity and the ratio σ_z/σ_ϵ tends to zero. Therefore, the chi-square approximation to twice the log-likelihood does not apply in this context. To circumvent the problem we have opted for a bootstrap approach. Specifically we have generated 10000 samples from the estimated normal model $y_i = h(x_i, \hat{\beta}) + \epsilon_i$ with $\epsilon_i \sim \mathcal{N}(0, \hat{\sigma})$ where $\hat{\beta}$ and $\hat{\sigma}$ are the MLE. For each sample we have calculated the difference between the maximum log-likelihood under the skew-t model with ν fixed to the MLE $\hat{\nu} = 4.6461$ and the maximum log-likelihood under the normal model. The associated bootstrap p-value for the normal model is 0.014. This indicates low evidence for the normal linear model respect to the skew-t model. Graphical analysis seem to confirm this. In Fig 1 we report, for the normal, ALS, and skew-t model, the histograms and pp-plots for the estimated standardized residuals $(y_i - h(x_i, \hat{\beta}))/\hat{\omega}$, where ω is equal to σ_ϵ for the normal model and to $\sqrt{\sigma_\epsilon^2 + \sigma_z^2}$ both for the ALS model and the skew-t model. In particular, looking the pp-plots for the normal model and the skew-t model it is evident that the latter fits the data better.

3.2 Testing the inefficiency term

The next goal in analazying the data is to understand if we can drop the inefficiency term when we take account of heavy tails, thus we test $H_0 : \alpha = 0$ against $H_1 : \alpha < 0$, in model (4) with $\alpha \in (-\infty, \infty)$ and we use test statistic R. The MLE of α is $\hat{\alpha} = -0.936$ and the maximum log-likelihood associated to the Student-t regression model is 67.56. The test statistic $R = sign(\hat{\alpha})(\ell(\hat{\theta}) - \ell(\hat{\theta}^*))^{1/2}$ is -1.536 and the associated observed p-value is 0.06 indicating low evidence for the absence of the inefficiency term. Note that inference on the inefficiency term is affected by accounting for heavy tails. Indeed, in comparing the normal model against the ALS model, we may test $H_0 : \alpha = 0$ against $H_1 : \alpha < 0$, in model (4) with $\alpha \in (-\infty, \infty)$ and $\nu = \infty$. In this case the test statistic R is equal to -0.94 and the observed p-value is $P(\mathcal{N}(0, 1) < -0.94) = 0.17$. Thus, for the elettric companies data set, when we allow for thick tails the evidence for the presence of the inefficiency term increases.

3.3 Estimating individual efficencies

Finally we present our results for the efficiency of firms within the sample. In table 2 we compare the quantities $r_i = E(e^{-z_i}|y_i)$ for the first five firms in the sample both for the ALS model and the skew-t model. These are the same firms analyzed by van den Broeck *et al.* (1994). We see that,with the skew-t model the estimates for r_i are generally bigger than for the ALS model, with the exception of the second firm. More insight into the beheaviour of the models about efficiency analysis is given in Fig 2. This plots, for all the firms, the plug-in estimates of $r_i = E(e^{-z_i}|y_i)$ and s_i where $s_i^2 = Var(e^{-z_i}|y_i)$ against the estimated residuals $y_i - h(x_i, \hat{\beta})$ both for the ALS model and the skew-t model. We see that for the ALS model r_i is always increasing while s_i first increases and then decreases in according to the fact the e^{-z_i} given y_i should be concentrated in one when

$y_i - h(x_i, \hat{\beta})$ becomes very large. For the skew-t model we have a different behaviour, when the residuals becomes larger both r_i and s_i start to increase in according with our findings that the density of z_i given y_i is completely flat when $y_i - h(x_i, \hat{\beta})$ becomes very large.

4 Discussion

In this paper we have introduced a new stochastic frontier model with several attractive features. First of all, it generalizes the common stochastic frontier model, where a normal distribution is assumed for the error term and a half-normal distribution is assumed for the inefficiency term, allowing for fat tails in the composed error distribution. Adopting this new model we do not have to worry to remove outlier observations before drawing inference. In fact, if our aim is to estimate individual technical efficieny, the model automatically increases the uncertainty of our estimates when we have observations lying above and far from the estimated frontier. We have discussed frequentist inference and further research will be conducted on estimating the model from a Bayesian point of view.

Appendix A: density for the individidual inefficiency in the skew-t model

Let us observe that

$$f(z_i \,|\, \lambda_i, y_i) = \frac{e^{-\frac{1}{2}\frac{\lambda_i(\sigma_\epsilon^2+\sigma_z^2)}{\sigma_\epsilon^2\,\sigma_z^2}\left(z_i - \frac{-\sigma_z^2(y_i-h_i)}{\sigma_\epsilon^2+\sigma_z^2}\right)^2}}{\sqrt{2\pi\frac{\sigma_\epsilon^2\,\sigma_z^2}{\lambda_i(\sigma_\epsilon^2+\sigma_z^2)}}\,\Phi\left(\frac{-\sqrt{\lambda_i}\sigma_z\,(y_i-h_i)}{\sigma_\epsilon\sqrt{\sigma_\epsilon^2+\sigma_z^2}}\right)}$$

and that

$$f(\lambda_i|y_i) = \frac{f(y_i|\lambda_i)f(\lambda_i)}{f(y_i)} = \frac{2\sqrt{\lambda_i}e^{-\frac{1}{2}\frac{\lambda_i(y_i-h_i)^2}{\sigma_z^2+\sigma_\epsilon^2}}\,\Phi\left(\frac{-\sqrt{\lambda_i}\sigma_z\,(y_i-h_i)}{\sigma_\epsilon\sqrt{\sigma_\epsilon^2+\sigma_z^2}}\right)}{f(y_i)\sqrt{2\pi(\sigma_\epsilon^2+\sigma_z^2)}}\frac{\left(\frac{\nu}{2}\right)^{\nu/2}e^{-\lambda_i\frac{\nu}{2}}\lambda_i^{\nu/2-1}}{\Gamma\left(\frac{\nu}{2}\right)}.$$

Then we have

$$\begin{aligned} f(z_i\,|\,y_i) &= \int f(z_i\,|\,\lambda_i, y_i)f(\lambda_i|y_i)d\lambda_i \\ &= \frac{\left(\frac{\nu}{2}\right)^{\nu/2}}{f(y_i)\Gamma\left(\frac{\nu}{2}\right)\pi\sqrt{\sigma_\epsilon^2\sigma_z^2}}\int_0^\infty \lambda_i^{\frac{\nu+2}{2}-1}e^{\frac{-\lambda_i}{2}\left[\nu+\frac{(y_i-h_i)^2}{\sigma_\epsilon^2+\sigma_z^2}+\frac{\sigma_\epsilon^2+\sigma_z^2}{\sigma_\epsilon^2\,\sigma_z^2}\left(z_i+(y_i-h_i)\frac{\sigma_z^2}{\sigma_\epsilon^2+\sigma_z^2}\right)^2\right]}d\lambda_i \\ &= \frac{\left(\frac{\nu}{2}\right)^{\nu/2}\Gamma\left(\frac{\nu+2}{2}\right)}{f(y_i)\Gamma\left(\frac{\nu}{2}\right)\pi\sqrt{\sigma_\epsilon^2\sigma_z^2}\left[\frac{1}{2}\left\{\nu+\frac{(y_i-h_i)^2}{\sigma_\epsilon^2+\sigma_z^2}+\frac{\sigma_\epsilon^2+\sigma_z^2}{\sigma_\epsilon^2\,\sigma_z^2}\left(z_i+(y_i-h_i)\frac{\sigma_z^2}{\sigma_\epsilon^2+\sigma_z^2}\right)^2\right\}\right]^{\frac{\nu+2}{2}}}. \end{aligned}$$

Appendix B: limiting behaviour for the individual inefficiency

1) ALS model

Let us consider the following notation $A = \frac{1}{2}\frac{\sigma_z^2\sigma_\epsilon^2}{\sigma_z^2+\sigma_\epsilon^2}$, $B = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2+\sigma_z^2}$, $C = \frac{\sigma_z}{\sigma_\epsilon}\frac{1}{\sqrt{\sigma_\epsilon^2+\sigma_z^2}}$, and observe that $-\frac{1}{2}C^2 = -AB^2$. We have

$$\lim_{y_i-h_i\to\infty} f(z_i|y_i) = \lim_{y_i-h_i\to\infty} \frac{e^{-A(z_i+B(y_i-h_i))^2}}{\Phi(-C(y_i-h_i))} = \lim_{y_i-h_i\to\infty} \frac{-2e^{-A(z_i+B(y_i-h_i))^2}AB(z_i+B(y_i-h_i))}{-Ce^{-\frac{1}{2}C^2(y_i-h_i)^2}}$$

Thus if $z_i = 0$ we have

$$\lim_{y_i-h_i\to\infty} f(z_i|y_i) = \lim_{y_i-h_i\to\infty} 2\frac{AB^2}{C}(y_i-h_i) = \infty$$

while if $z_i > 0$ we have

$$\lim_{y_i-h_i\to\infty} f(z_i|y_i) = \lim_{y_i-h_i\to\infty} e^{-A(z_i+B(y_i-h_i)^2+\frac{1}{2}C^2(y_i-h_i)^2)} = 0$$

2) skew-t model

$$\lim_{y_i-h_i\to\infty} f(z_i|y_i) = \lim_{y_i-h_i\to\infty} \frac{\left[1+\frac{1}{\nu}\left(\frac{y_i-h_i}{\sqrt{\sigma_\epsilon^2+\sigma_z^2}}\right)^2\right]^{\frac{\nu+1}{2}}}{\left[\nu+\frac{(y_i-h_i)^2}{\sigma_\epsilon^2+\sigma_z^2}+\frac{\sigma_\epsilon^2+\sigma_z^2}{\sigma_\epsilon^2\,\sigma_z^2}\left(z_i+(y_i-h_i)\frac{\sigma_z^2}{\sigma_\epsilon^2+\sigma_z^2}\right)^2\right]^{\frac{\nu+2}{2}}} = 0 \quad \forall z \geq 0$$

References

AIGNER D, LOVELL CAK, SCHMIDT P. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* **6**: 21-37.

AZZALINI A. 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**: 171-178.

AZZALINI A, CAPITANIO A. 1999. Statistical applications of the multivariate skew normal distribution. *Journal of Royal Statistical Society, Series B* **61**: 579-602.

BRANCO MD, DEY DK. 2001 A general class of multivariate skew elliptical distributions. Journal of Multivariate Analysis **79**: 99-113.

AZZALINI A, CAPITANIO A 2002. Distributions generated by permuation of symmetry with emphasis on a multivariate skew t distribution. *Journal of Royal Statistical Society, Series B* **65**:367-389

COELLI T, PRASADA RAO DS, BATTESE GE. 1998. *An Introduction to Efficiency and Productivity Analysis* Kluwer Academic Publisher.

CHRISTENSEN LR, GREENE WH. 1976 Economies of scale in U.S. electric power generation. *Journal of Political Economy* **84**: 655-676.

GREENE WH. 1990. A gamma-distributed stochastic frontier model. *Journal of Econometrics* **46**: 141-163.

GRIFFIN JE, STEEL MFJ. 2002 Semiparametric Bayesian inference for stochastic frontier models Institute of Mathematics and Statistics. Technical Report UKC/IMS/02/19

JONDROW J, LOVELL CAK, MATEROV IS, SCHMIDT P. 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics* **19**: 233-238.

MEEUSEN WJ. VAN DEN BROECK J. 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic review* **8**: 435-444.

RITTER C, SIMAR L. 1994. Another look at the American elettrical utility data. *Technical report B1348, Institut de Statistique, Universitè Catholique de Louvain* Louvain-la-Neuve (Belgium).

STEVENSON RE. 1980. Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics* **13**: 57-66.

VAN DEN BROECK J, KOOP G, OSIEWALSKY J, STEEL M. 1994 Stochastic frontier models. A Bayesian perspective. *Journal of Econometrics* **61**: 273-303.

TSIONAS E. 2002. Stochastic frontier models with random coefficients. *Journal of Applied Econometrics* **17**: 127-147.

	normal linear model		ALS model		skew-t model	
	mle	95% confidence interval	mle	95% confidence interval	mle	95% interval.
β_0	-7.2047	(-7.8512 -6.5582)	-7.4071	(-8.0389 -6.7752)	-7.8208	(-8.4419 -7.1997)
β_1	0.3860	(0.3126 0.4595)	0.4081	(0.3316 0.4846)	0.4549	(0.3871 0.5228)
β_2	0.0316	(0.0264 0.0368)	0.0306	(0.0254 0.0357)	0.0278	(0.0231 0.0325)
β_3	0.2462	(0.1178 0.3746)	0.2439	(0.1186 0.3692)	0.2952	(0.1724 0.4181)
β_4	0.0792	(-0.0390 0.1974)	0.0592	(-0.0606 0.1790)	0.0344	(-0.0725 0.1412)
σ_z	—-	—-	0.1558	(0.0853 0.2844)	0.0900	(0.0302 0.2682)
σ_ϵ	0.1419	(0.1252,0.1608)	0.1069	(0.0703 0.1627)	0.0949	(0.0669 0.1347)
ν	—-	—-	—-	—	4.6461	(1.912 11.2441)
log lik	65.67		66.14		68.75	

Table 1: Estimates, 95% confidence intervals and maximum loglikelihood values

	ALS model	skew-t model
r_1	0.7325	0.7749
r_2	0.9650	0.9408
r_3	0.9145	0.9432
r_4	0.8980	0.9252
r_5	0.9510	0.9524

Table 2: Efficiencies for the first five firms

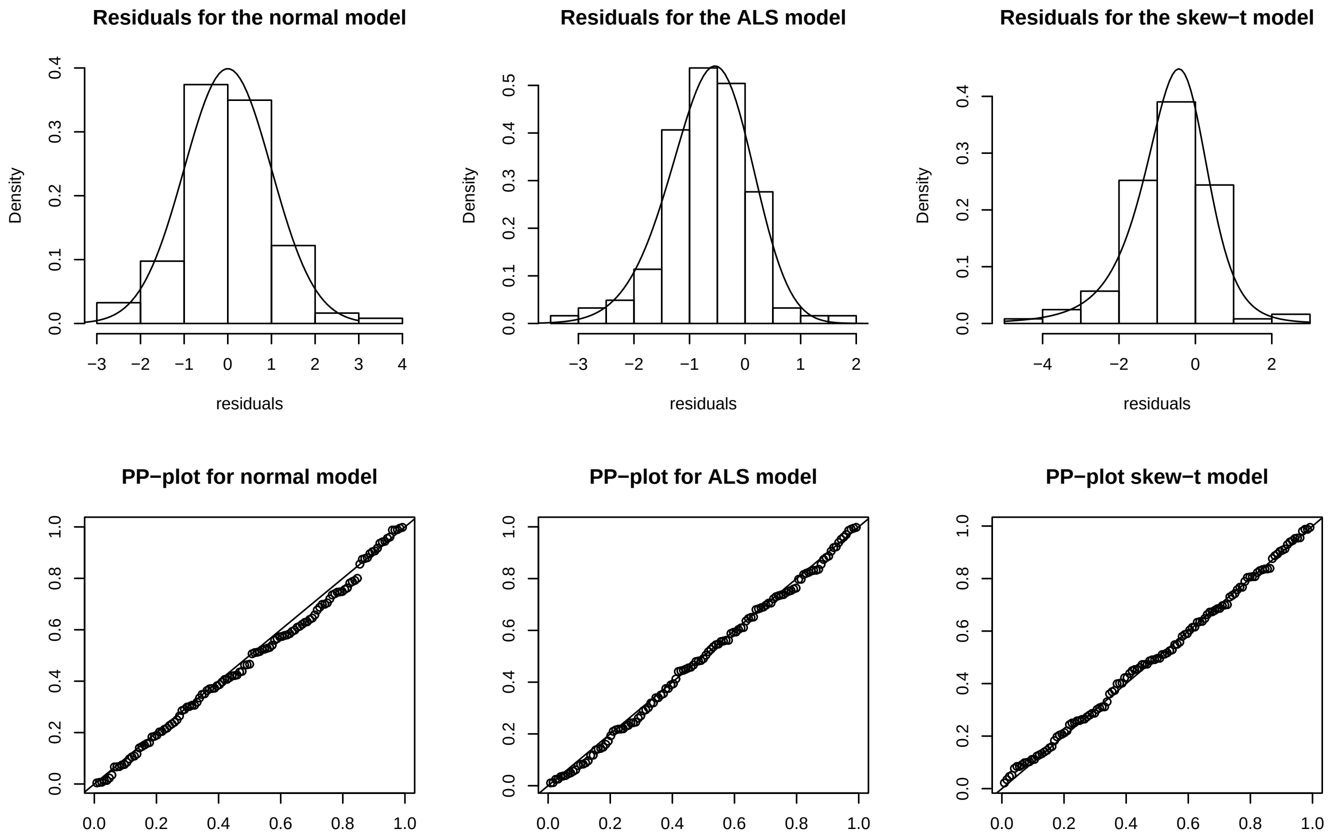

Figure 1: Residuals analysis

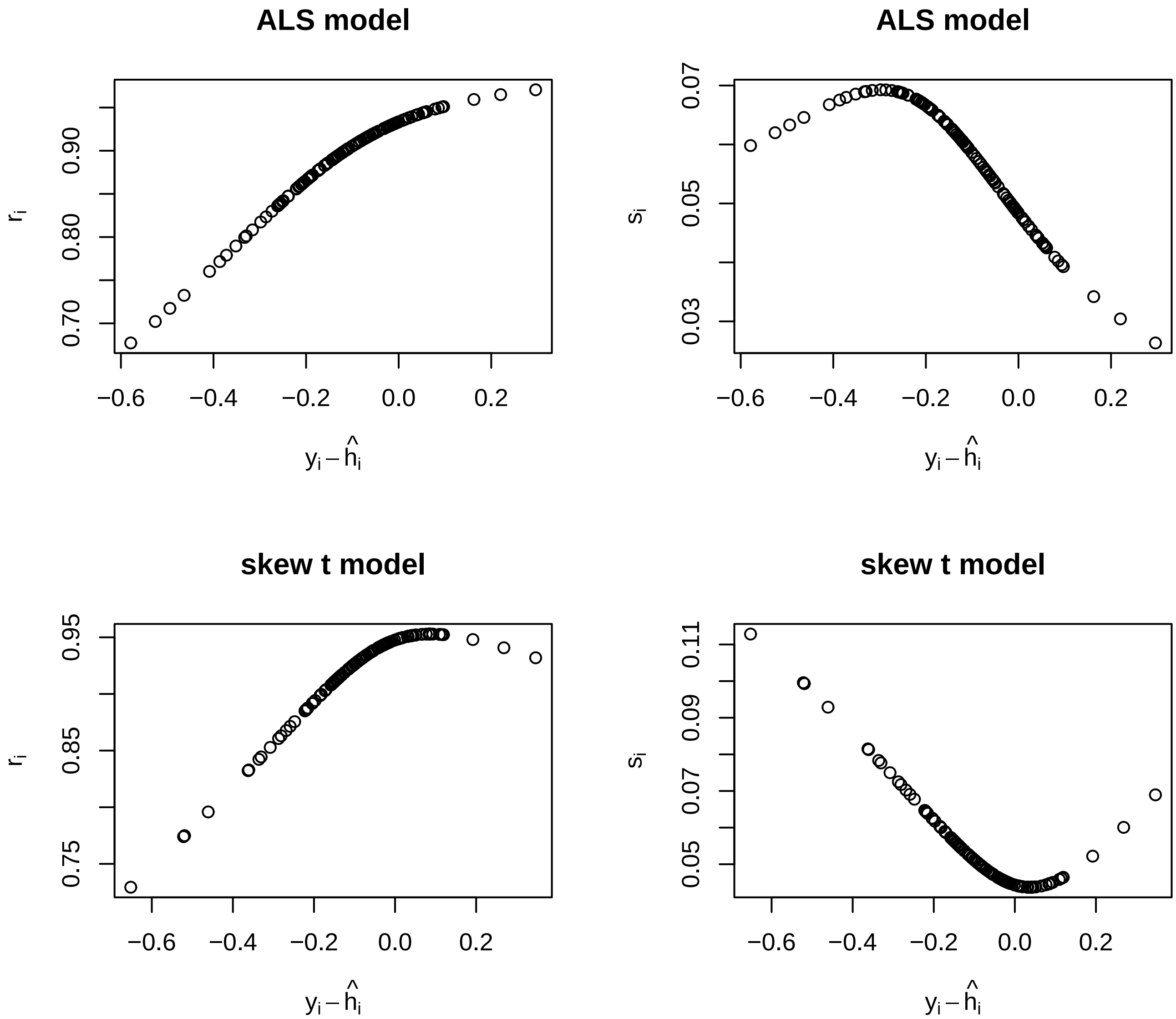

Figure 2: Eefficiencies respect to the estimated residuals