The Folded Normal Stochastic Frontier

Gholamreza Hajargasht

Department of Economics

University of Melbourne, Australia

Abstract

We introduce a stochastic frontier model with a *folded normal* inefficiency distribution. This model may be a reasonable alternative to the popular truncated normal model. We study the model and show how it can be estimated using a maximum likelihood approach. We also apply the model to two data sets and compare the results to those from the truncated normal model.

.

Keywords: Technical Efficiency, Truncated Normal, Skew Normal

JEL Classifications: C11, D24, C23, C12

Corresponding Author

Gholamreza Hajargasht
Department of Economics, University of Melbourne
Vic 3010, Australia
Email: har@unimelb.edu.au

1- *Introduction*

We consider a stochastic frontier model with a *folded normal* distribution. This model may be considered as an alternative to the popular truncated normal model. In section 2 we review the folded normal distribution and briefly study its properties. Section 3 introduces a new skew normal distribution formed as convolution of a folded normal and a normal distribution. In Section 4 we introduce the folded normal stochastic frontier model and derive its likelihood function and provide the estimator for the inefficiency effects. Section 5 applies the model to two real data sets.

2- *The Folded Normal Distribution*

The folded normal distribution is a one-sided distribution derived from the normal distribution. If variable Y is normally distributed with mean μ and variance σ^2, the random variable $X = |Y|$ has a distribution known as the folded normal i.e.

$$Y \sim N(\mu,\sigma^2) \Rightarrow X = |Y| \sim FN(\mu,\sigma^2)$$

It is called folded because probability mass to the left of $x = 0$ is "folded" over by taking the absolute value [see Tsagris et al. (2014) and references cited therein for more information on this distribution]. PDF and CDF of this distribution are given by

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(X-\mu)^2}{2\sigma^2}\right\} + \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(X+\mu)^2}{2\sigma^2}\right\}$$

$$F(X) = \Phi\left(\frac{X-\mu}{\sigma}\right) + \Phi\left(\frac{X+\mu}{\sigma}\right) - 1$$

The first and second order moments of this distribution can be obtained as

$$E(X) = \sigma\sqrt{\frac{2}{\pi}}\exp\left\{-\frac{\mu}{2\sigma^2}\right\} + \mu\left[1-2\Phi\left(-\frac{\mu}{\sigma}\right)\right]$$

$$E(X^2) = \mu^2 + \sigma^2$$

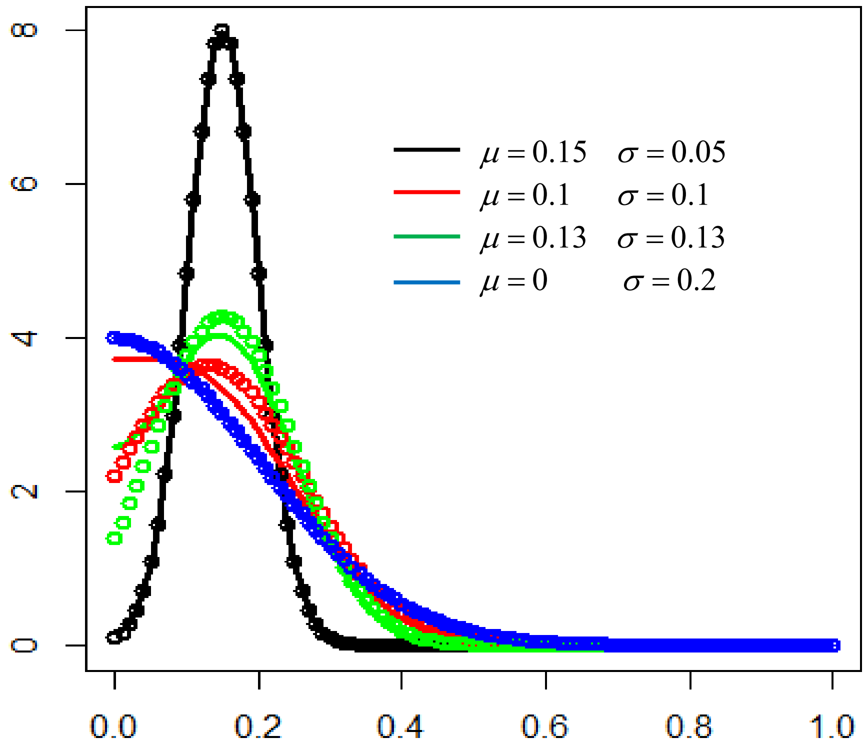
Mode can be obtained by solving the following equation in terms of x

$$x = -\frac{\sigma^2}{2\mu}\log\left(\frac{\mu - x}{\mu + x}\right)$$

It can be seen that when $\mu < \sigma$, the $Mode = 0$. When $\mu \geq \sigma$, the $Mode > 0$, and when μ becomes greater than 3σ, the maximum approaches μ since, in this case, the folded normal approaches the normal distribution. Note also that this distribution is equal to half-normal when $\mu = 0$.

Figure -1 compares Folded normal density function with a truncated normal density with the same values for μ and σ parameters. As it can be seen, where $\mu = 0$ they coincide. If $\sigma \to 0$ they approach each other while the biggest difference is where $\mu = \sigma$. Of course by choosing different values for parameters of each model one may obtain similar shapes. Another difference is that the folded normal distribution is unchanged if μ is replaced with $-\mu$.

Figure-1: Folded and Truncated Normal for some values of μ and σ



Solid lines represents $FN(m,\sigma^2)$ and Points represents $TN(m,\sigma^2)$

3- A New Skew Normal Distribution

The conventional skew normal distribution [see e.g. Azzalini (2013)] is obtained as convolution of a truncated normal and a normal distribution. Similarly, one can obtain another class of skew normal distributions through convolution of the folded normal and the normal distributions. Let $v_i \sim N(0,\sigma_v^2)$ and $u \sim FN(\mu,\sigma_u^2)$ and define $\varepsilon = -u + v$. To derive density function for ε we proceed by writing

$$f(\varepsilon)=\int_0^\infty f(\varepsilon+u)f(u_i)du=\frac{1}{(2\pi)^{(T+1)/2}\sigma_v^T\sigma_u}\left\{\begin{array}{l}\int_0^\infty \exp\left\{\frac{-1}{2\sigma_v^2}(u+\varepsilon)^2-\frac{(u-\mu)^2}{2\sigma_u^2}\right\}du\\ +\int_0^\infty \exp\left\{\frac{-1}{2\sigma_v^2}(u+\varepsilon)^2-\frac{(u+\mu)^2}{2\sigma_u^2}\right\}du\end{array}\right\}$$

It can be shown that

$$f(\varepsilon)=\frac{1}{\sigma}\left\{\phi\left(\frac{\mu-\varepsilon}{\sigma}\right)\Phi\left(-\frac{\mu}{\lambda\sigma}-\frac{\varepsilon\lambda}{\sigma}\right)+\phi\left(\frac{\mu+\varepsilon}{\sigma}\right)\Phi\left(\frac{\mu}{\lambda\sigma}-\frac{\varepsilon\lambda}{\sigma}\right)\right\}$$

where $\sigma=\sqrt{\sigma_u^2+\sigma_v^2}$ and $\lambda=\sigma_u/\sigma_v$. This distribution may be regarded as a new skew normal distribution. We now study its properties. Its CDF can be obtained as

$$F(x;\mu;\sigma,\lambda)=\frac{1}{\sigma}\left\{\phi\left(\frac{x+\mu}{\sigma}\right)\Phi\left(\lambda\frac{x-\mu}{\sigma}\right)+\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x+\mu}{\sigma}\right)\right\}$$

To obtain the moments we derive its moment generating function. Using some tedious integration it can be shown that

$$E(e^{t\varepsilon})=\exp\left(-\frac{t^2\sigma^2}{2}+\mu t\right)\Phi\left(-\frac{\mu(1+\lambda^2)+\sigma^2\lambda^2 t}{\lambda\sigma\sqrt{1+\lambda^2}}\right)+\exp\left(-\frac{t^2\sigma^2}{2}-\mu t\right)\Phi\left(\frac{\mu(1+\lambda^2)-\sigma^2\lambda^2 t}{\lambda\sigma\sqrt{1+\lambda^2}}\right)$$

From this we can derive its moments. The first 4 central moments are

$$m_1=\mu\left\{1-2\Phi\left(\frac{\mu\sqrt{1+\lambda^2}}{\lambda\sigma}\right)\right\}+\frac{2\sigma\lambda}{\sqrt{1+\lambda^2}}\phi\left(\frac{\mu\sqrt{1+\lambda^2}}{\lambda\sigma}\right)$$

$$m_2=\mu^2+\sigma^2$$

$$m_3=(\mu^2+2\sigma^2)m_1-\mu\sigma^2\left(1-2\Phi\left(\frac{\mu\sqrt{1+\lambda^2}}{\sigma}\right)\right)$$

$$m_4=\mu^4+6\mu^2\sigma^2+3\sigma^4$$

The even order moments have a similar form to the even order moments of the folded normal distribution with $\sigma=\sqrt{\sigma_u^2+\sigma_v^2}$.

4- *Stochastic Frontier*

Consider the panel stochastic frontier model

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} \mp u_i + v_{it}$$

where $i = 1,...,N$ indexes the firms and $t = 1,...,T$ indexes time, $\mathbf{x}_{it}$ is a row vector of regressors (e.g., logs of inputs or logs of input prices), y_{it} represents the logarithm of output or cost, $\mathbf{x}_{it}\boldsymbol{\beta}$ is the log of the frontier production or cost function, u_i is a non-negative random error which accounts for the inefficiency of firm *i*, and v_{it} represents noise. Various distributions have been proposed for u_i in the literature. In this paper we consider a folded normal distribution for u_i. To estimate the model via maximum likelihood, we begin by denoting $\varepsilon_{it} = v_{it} \mp u_i = y_{it} - \mathbf{x}_{it}\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1},...,\varepsilon_{iT})$. Then, to obtain the likelihood function we first find

$$f(\boldsymbol{\varepsilon}_i) = \int_0^\infty \prod_{t=1}^{T} f(\varepsilon_{it} \pm u_i) f(u_i) du_i = \frac{1}{(2\pi)^{(T+1)/2}\sigma_v^T\sigma_u} \left\{ \begin{array}{l} \int_0^\infty \exp\left\{ \frac{-1}{2\sigma_v^2}\sum_{t=1}^{T}(u_i + \varepsilon_{it})^2 - \frac{(u_i - \mu)^2}{2\sigma_u^2} \right\} du_i \\ + \int_0^\infty \exp\left\{ \frac{-1}{2\sigma_v^2}\sum_{t=1}^{T}(u_i + \varepsilon_{it})^2 - \frac{(u_i + \mu)^2}{2\sigma_u^2} \right\} du_i \end{array} \right\}$$

It can be shown that

$$f(\boldsymbol{\varepsilon}_i) = \frac{\sigma}{(2\pi)^{T/2}\sigma_u\sigma_v^T} \exp\left(-\sum_{t=1}^{T} \frac{\varepsilon_{it}^2}{2\sigma_v^2} + \frac{\mu^2}{2\sigma_u^2} \right) \left\{ \exp\left(\frac{\mu_{i1}^2}{2\sigma^2} \right) \Phi\left(\frac{\mu_{i,1}}{\sigma} \right) + \exp\left(\frac{\mu_{i2}^2}{2\sigma^2} \right) \Phi\left(\frac{\mu_{i,2}}{\sigma} \right) \right\}$$

where $\mu_{i,1} = \dfrac{\sigma_v^2\mu \mp \sigma_u^2 \sum_{i=1}^{T} \varepsilon_{it}}{(T\sigma_u^2 + \sigma_v^2)}$ $\mu_{i,2} = \dfrac{-\sigma_v^2\mu \mp \sigma_u^2 \sum_{i=1}^{T} \varepsilon_{it}}{(T\sigma_u^2 + \sigma_v^2)}$ and $\sigma^2 = \dfrac{\sigma_u^2\sigma_v^2}{\sigma_v^2 + T\sigma_u^2}$. $\phi(.)$ and $\Phi(.)$ are the *pdf* and *cdf* of the normal distribution. Using this formula log-likelihood function can be obtained as

$$LogL = -\frac{NT}{2}\ln 2\pi - \frac{N(T-1)}{2}\ln\sigma_v^2 - \frac{T}{2}\ln\left(T\sigma_u^2 + \sigma_v^2\right) - \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\varepsilon_{it}^2}{2\sigma_v^2}$$
$$- N\frac{\mu^2}{\sigma_u^2} + \sum_{i=1}^{N}\ln\left\{ \exp\left(\frac{\mu_{i,1}^2}{2\sigma^2} \right)\Phi\left(\frac{\mu_{i,1}}{\sigma} \right) + \exp\left(\frac{\mu_{i,2}^2}{2\sigma^2} \right)\Phi\left(\frac{\mu_{i,2}}{\sigma} \right) \right\}$$

This log-likelihood function is not very different from log-likelihood function of a stochastic frontier model with the truncated normal distribution (see e.g. Kumbhakar and Lovell 2000 or Greene 2008) and it can be maximized using standard numerical methods in a similar fashion. $E(u_i \mid \boldsymbol{\varepsilon}_i)$ is often used for estimation of individual inefficiency effects which are often of primary interest in efficiency studies. To obtain $E(u_i \mid \boldsymbol{\varepsilon}_i)$ note that

$$E(u_i \mid \boldsymbol{\varepsilon}_i) = \frac{\int_0^{\infty} u_i f(\boldsymbol{\varepsilon}_i, u_i) du_i}{f(\boldsymbol{\varepsilon}_i)}$$

Some relatively tedious integration and algebra gives

$$E(u_i \mid \boldsymbol{\varepsilon}_i) = \frac{\exp\left(\frac{\mu_{i,1}^2}{2\sigma^2}\right)\left\{\sigma\phi\left(\frac{\mu_{i,1}}{\sigma}\right) + \mu_{i,1}\Phi\left(\frac{\mu_{i,1}}{\sigma}\right)\right\} + \exp\left(\frac{\mu_{i,2}^2}{2\sigma^2}\right)\left\{\sigma\phi\left(\frac{\mu_{i,2}}{\sigma}\right) + \mu_{i,2}\Phi\left(\frac{\mu_{i,2}}{\sigma}\right)\right\}}{\exp\left(\frac{\mu_{i,1}^2}{2\sigma^2}\right)\Phi\left(\frac{\mu_{i,1}}{\sigma}\right) + \exp\left(\frac{\mu_{i,2}^2}{2\sigma^2}\right)\Phi\left(\frac{\mu_{i,2}}{\sigma}\right)}$$

5- *An Example*

Here we estimate a stochastic frontier model with the folded normal distribution using two real data sets. We first use the famous rice data set, collected by the International Rice Research Institute (IRRI) consists of a panel of 43 Philippine rice farms observed over 8 years from 1990 to 1997 (see Coelli, Rao, O'Donnell and Battese 2005 for further information). We only consider the last 4 years because of the time invariance assumption. The model we consider for this data set is

$$\ln y_{it} = \beta_0 + \beta_1 \ln(land_{it}) + \beta_2 \ln(labor_{it}) + \beta_3 \ln(fert_{it}) + \beta_4 \ln(others_{it}) - u_i + v_{it}$$

where y_{it}, $land_{it}$, $labor_{it}$, $fert_{it}$ and $others_{it}$ represent output, land, hired labour, amount of fertilizer and other inputs respectively.

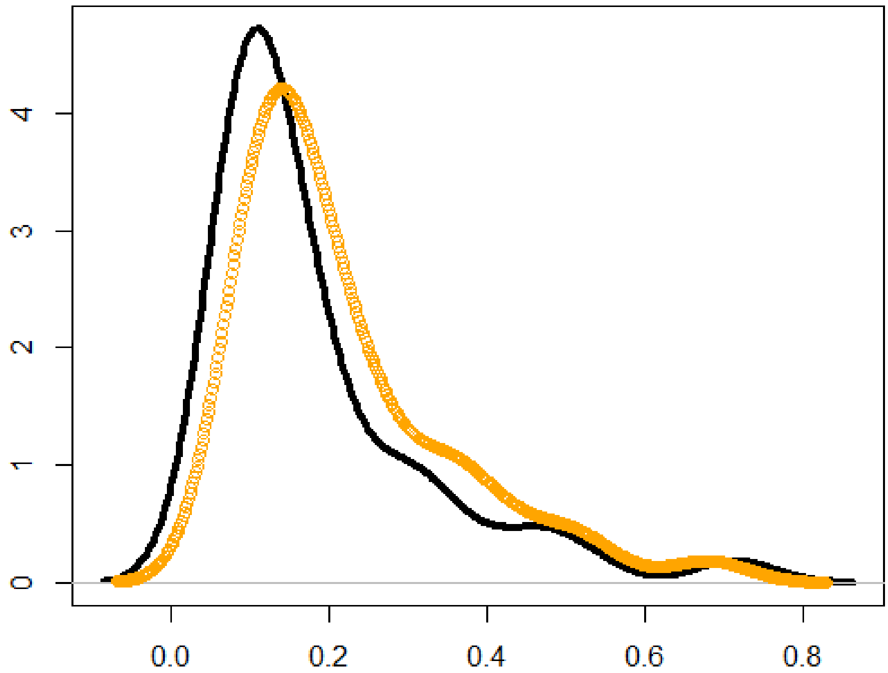
Table-1 present ML estimates for the parameters of both folded and truncated normal models. The estimates of the parameters corresponding to *land*, *labour*, *fert* and o*thers* represent elasticities. These estimates are sensible both in terms of their signs and magnitudes. Note that the estimates for parameters of technology are not substantially different across the models and since our focus is on efficiencies and comparing the models and not analysing the characteristics of these particular examples' production technology, we do not discuss them further. The estimated inefficiency distribution parameters i.e. μ and σ_u are different across the two models. In case of the folded normal, μ is almost zero indicating a half-normal distribution but the estimated μ from truncated normal is -1.028 that is very different from zero (i.e. half-normal) but with large standard errors and not significantly different from zero. In case of truncated normal, we had difficulties in reaching the convergence where the value of μ kept decreasing (up to -7 or even more) without noticeable changes in the Log-likelihood value.

Table-1 ML estimates for Parameters of Various Models – Rice Data

	TN		FN	
	Par	**SE**	**Par**	**SE**
C	-0.937	0.386	-0.954	0.354
Land	0.384	0.099	0.386	0.097
Labor	0.294	0.097	0.301	0.097
Fert	0.196	0.055	0.203	0.061
Other	0.057	0.033	0.053	0.034
σ_v	0.358	0.260	0.308	0.019
μ	-1.028	1.335	0.000	0.880
σ_u	0.967	0.218	0.280	0.060
LogL	-57.778		-58.193	

Figure-2 provides kernel density estimate of the estimated inefficiencies. As it can be seen the estimates from the folded and truncated normal models are somewhat different. They have a similar shape but one looks like a slightly translated version of the other i.e. estimated values from folded normal model are almost always slightly larger (between .01 to 0.5).

Figure-2 Kernel Density Estimate of Distribution of Inefficiencies



Orange Represents FN and Black Represents TN

The second data set consists of six years of observations on 247 Spanish dairy farms. The output, y_{it} is milk production. The four inputs, x_1, x_2, x_3 and x_4 are *feed*, *land*, *labor* and *cows*. We consider the following translog production frontier model (see Alvarez, Arias and Greene 2004 for further information).

$$y_{it} = \beta_0 + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \beta_3 x_{3,it} + \beta_4 x_{4,it} + \beta_{11} x_{1,it}^2 + \beta_{22} x_{2,it}^2 + \beta_{33} x_{3,it}^2 + \beta_{44} x_{4,it}^2 + \beta_{12} x_{1,it} x_{2,it}$$
$$+ \beta_{13} x_{1,it} x_{3,it} + \beta_{14} x_{1,it} x_{4,it} + \beta_{23} x_{2,it} x_{3,it} + \beta_{24} x_{2,it} x_{3,it} + \beta_{34} x_{3,it} x_{4,it} + \gamma_1 t + \gamma_{11} t^2 - u_i + v_{it}$$

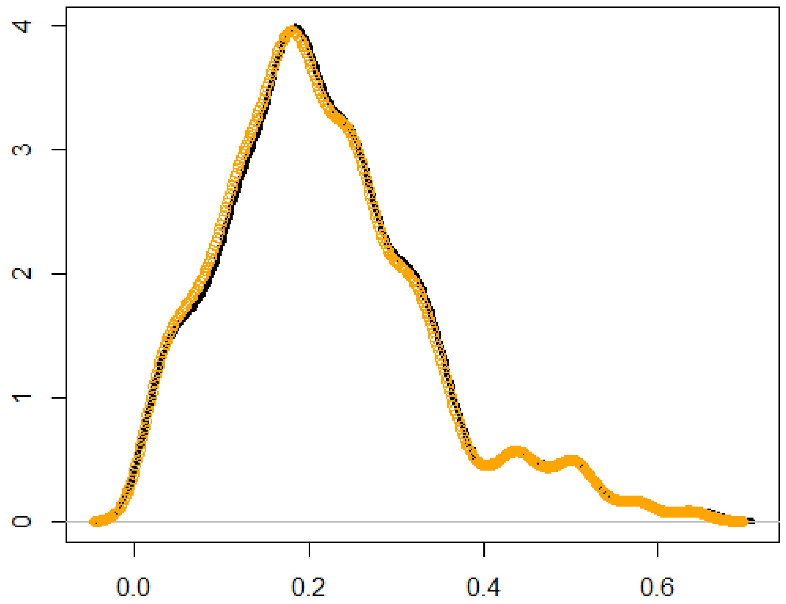
Table-2 presents ML estimates for the parameters of both folded and truncated normal models for dairy data. Similar conclusions can be drawn from Table-2 which contains the results for the dairy data. The estimates for the technology parameters are almost across the same across the models.

Table-2 ML Estimates for Folded and Truncated Normal – Dairy Data

	TN		**FN**	
	Par	**SE**	**Par**	**SE**
C	11.718	0.020	11.721	0.023
β1	0.663	0.022	0.663	0.021
β2	0.040	0.014	0.039	0.014
β3	0.051	0.018	0.052	0.017
β4	0.353	0.013	0.353	0.012
β11	0.294	0.109	0.287	0.103
β22	-0.072	0.052	-0.068	0.052
β33	-0.128	0.113	-0.129	0.110
β44	0.110	0.034	0.108	0.033
β12	-0.053	0.053	-0.052	0.053
β13	0.083	0.063	0.084	0.053
β14	-0.136	0.055	-0.131	0.053
β23	0.000	0.041	-0.003	0.037
β24	0.016	0.027	0.016	0.027
β34	-0.012	0.036	-0.012	0.032
γ_1	0.032	0.006	0.032	0.006
γ_{11}	-0.003	0.001	-0.003	0.001
σ_v	0.077	0.002	0.077	0.002
μ	0.188	0.033	0.214	0.025
σ_u	0.148	0.012	0.131	0.009
LogL	1356.17		1354.33	

Figure-3 provides kernel density estimate of the estimated inefficiencies for dairy data. As it can be seen the estimate from folded normal and truncated normal models are indistinguishable.

Figure-3 Kernel Density Estimate of Distribution of Inefficiencies for Dairy Data



Orange represents FN and Black represents TN

Conclusion

In this paper we introduced a stochastic frontier model with a folded normal distribution for inefficiency distribution. We derived the likelihood function, efficiency estimator and applied the model to two data sets. The model had a reasonable empirical performance although did not outperform the popular truncated normal model. This model however might be more tractable when estimating some of the complex models where modelling of inefficiencies is of interest.

References

Azzallini, A. (2014). *The Skew-Normal and Related Families*, Cambridge University Press,

Aigner, D., C. Lovell, C. and P. Schmidt (1977). Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics*, 12: 21-37.

Coelli, T., Rao, P., O'Donnell, C. and Battese, G. (2005) *An Introduction to Efficiency and Productivity Analysis*. 2nd edition. Boston: Kluwer Academic Publishers.

Elandt, R. (1961) The Folded Normal Distribution: Two Methods of Estimating Parameters from Moments. *Technometrics,* 3, 551–562.

Greene, W., Econometric Approach to Efficiency Analysis, Ch 2 in the *Measurement of Productive Efficiency and Productivity Growth* edited by Fried, Lovell and Schmidt, Oxford University Press, pp. 92- 250,

Johnson, N. (1962) The Folded Normal Distribution: Accuracy of Estimation by Maximum Likelihood. *Technometrics,* 4, 249–256.

Johnson N., S. Kotz and N. Balakrishnan. (1994) *Continuous Univariate Distributions*; John Wiley & Sons, Inc.: New York, NY, USA.

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

Leone, F., L. Nelson and R. Nottingham (1961) The Folded Normal Distribution. *Technometrics*, 3, 543–550.