# Evaluating the Importance of Multiple Imputations of Missing Data on Stochastic Frontier Analysis Efficiency Measures

**Saleem Shaik[1] and Oleksiy Tokovenko[2]**

***Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2013 AAEA & CAES Joint Annual Meeting, Washington, DC, August 4 - 6, 2013.***



[1]Saleem Shaik (saleem.shaik@ndsu.edu) Assistant Professor, Department of Agribusiness and Applied Economics, North Dakota State University, Fargo, ND.

[2]Oleksiy Tokovenko (oleksiy.tokovenko@sas.com) Econometrician, SAS Institute, Inc

# Evaluating the Importance of Multiple Imputations of Missing Data on Stochastic Frontier Analysis Efficiency Measures

Saleem Shaik Oleksiy Tokovenko

**Abstract**

The robustness of the multiple imputation of missing data on parameter coefficients and efficiency measures is evaluated using stochastic frontier analysis in the panel Bayesian context. Second, the implications of multiple imputations on stochastic frontier analysis technical efficiency measures under alternative distributional assumptions – half-normal, truncation and exponential is evaluated. Empirical estimates indicate difference in the between-variance and within-variance of parameter coefficients estimated from stochastic frontier analysis and generalized linear models. Within stochastic frontier analysis, the between-variance and within-variance of technical efficiency are different across the three alternative distributional assumptions. Finally, results from this study indicate that even though the between- and within variance of multiple imputed data is close to zero, between- and within-variance of production function parameters, as well as, the technical efficiency measures are different.

**Research in progress. Do not quote without authors' permission.**

## 1 Introduction

Missing data are universally problematic in survey and longitudinal research. The uninformed researcher may analyze these incomplete data inappropriately or not at all. Missing data are problematic for a number of reasons. First, most statistical procedures rely on complete-data methods of analysis (Allison, 2000). Specifically,

computation programs require that all cases contain values for all variables used in the analysis. Thus, most statistical software programs exclude from analysis cases that have missing data on any of the variables (also known as listwise deletion). This can lead to two potentially serious problems: compromised analytic power and nonresponse bias. The analytical power may be significantly reduced if the researcher excludes all cases missing data for one or more variables (Allison, 2000). Discarding cases with missing data can bias a study severely. Non-response data is when respondents who do not answer a particular question, leading to a systematic pattern or bias characterize the missing data (Tabachnick and Fidell, 1999). Nonresponders may decide to omit certain questions for very distinct reasons that researchers may never know. Missing data problems are also prevalent in economic studies using aggregated data (county, state and county).

There are several approaches to handling missing data. These include weighting techniques, single imputation, and multiple imputation (MI). MI aims to create plausible imputations for the missing values, to accurately reflect uncertainty, and to preserve important data relationships and aspects of the data distributions (Schafer, 1997). Most of the studies have imputed missing data by taking into account the pattern of missingness (missing completely at random or missing at random; monotone or non-monotone missing data), type of imputed variable (continuous or discrete) and methods [regression and propensity score (Rubin, 1987), discriminant function and Markov chain Monte Carlo (Schafer, 1997)].

Current survey data analyses have been successfully using MI procedure to address missing data issues as well drawing inferences on parameter coefficients. However, the implications and inference on technical efficiency and productivity measures are seldom evaluated. Even though, aggregate country data are always faced with missing data. The missing data are more prominent with the inputs

used in the production of outputs in the low and middle income group countries. Further, data envelopment analysis and stochastic frontier analysis require non zero inputs and outputs to estimate efficiency and productivity measures. So many observations are dropped from the analysis and evaluated with short time series.

In this paper, the importance of accounting for inefficiency on MI of missing data is evaluated by comparing the stochastic frontier analysis (SFA) to generalized linear models (GLM) statistical procedures. Second, the importance of MI on technical efficiency measures is evaluated using SFA. Specifically, the importance of MI on SFA technical efficiency measures estimated under alternative distributional assumptions – half-normal, truncation and exponential is evaluated.

In Section 2, we present multiple imputation procedure and the three steps involved to impute, analyze, and draw inferences from imputation of missing data. The GLM and the SFA models with three alternative distributions models are also presented. Section 3 will provide details of the data used in MI analysis. Application of the three alternative SFA and technical efficiency measures is presented in Section 4, and some conclusions are drawn in Section 5.

# 2 Multiple Imputation

Modeling the missingness as missing at random (MAR) for a univariate data series can be represented by GLM in matrix notation as:

$$f\left(Y_i|X_i,\beta\right) \tag{1}$$

where $Y_i$ represents data collected on individual units $i$ (or subjects at the individual and countries at the aggregate level) conditional on $X_i, \beta$. Instead of missing completely at random (MCAR), we assume missing at random (MAR). This assumption of MAR will be used in the three-step approach of MI proposed by Rubin (1976).

The MI procedure (Rubin, 1987) replaces each missing value with a set of plausible values that represent the uncertainty about the correct values to impute. The multiple imputed data sets are then analyzed by using standard statistical or econometric procedures for complete data and combining the results from these analyses. No matter which complete-data analysis is used, the process of combining results from different data sets is essentially the same. The MI procedure does not attempt to estimate each missing value through simulated values but rather to represent a random sample of the missing values. In essence, MI procedure requires the analyst specify an imputation model, imputes several data sets, analyzes them separately, and then combines results. MI yields a single set of test statistics, parameter estimates, and standard errors.

There are several advantages to the MI procedures due to the underlying assumptions[1]. First, the MI procedure builds on the benefits of single imputation. Secondly, MI allows use of complete data methods for data analysis and also includes the data collector's knowledge. Third, MI incorporates random error because it requires random variation in the imputation process. Fourth, MI can accommodate any model and any data and does not require specialized software.

[1]First, missing data should be missing at random (MAR). Second, the imputation model must match the model used for analysis (Allison, 2000). Rubin (1987) termed this a "proper" imputation model. Schafer (1999) explained that the imputation model must preserve all important associations among variables in the data set, including interactions if they will be part of the final analysis. Further, the algorithm used to generate imputed values must be "correct", i.e., it must accommodate the necessary variables and their associations.

Fifth, MI simulates proper inferences from data; it also increases efficiency of the estimates because MI minimizes standard errors (Rubin, 1987). Finally, MI allows randomly drawn imputations under more than one model.

There are several disadvantages of the MI method. According to Rubin (1987) the three disadvantages of MI compared with other imputation methods are - more effort to create the multiple imputations, more time to run the analyses, and more computer storage space for the imputation-created data sets. Rubin (1987) also notes that MI was unacceptable because it uses simulation and adds random noise to the data. A final disadvantage of MI is its not producing a unique answer. Because randomness is preserved in the MI process, each data set imputed will yield slightly different estimates and standard errors. Therefore, the reproducibility of exact results may be problematic.

To overcome the disadvantages of MI, Rubin's (1976, 1987 and 1996) three-step approach of MI strategy is a useful method to analyze the uncertainty associated with the correct value to compute for missing values. The first step involves creating plausible values for missing observations that reflect uncertainty about the nonresponse model. These values will be used to impute the missing values multiple times to create number of completed datasets. Second step involves the use of these datasets to be analyzed using GLM and SFA that accounts for inefficiency. Finally in the third step the results are combined to evaluate the uncertainty regarding the MI on the production function parameter coefficients and technical efficiency measures of a production function.

## 2.1 First Step of Multiple Imputation

Suppose the complete data, $Y$ can be decomposed into two components – observed complete component,$Y_{obs}$ and missing component,$Y_{mis}$. Depending upon the missing patterns for continuous data, missing data is generated by randomly drawing from its distribution of the observed complete component,$Y_{obs}$ assuming multivariate normal distribution with mean vector and covariance matrix. Missing data are filled in $m$ times to generate $m$ complete data sets assuming expectation-maximization (EM) algorithm. For parametric models the EM algorithm is a two-step iterative process that finds the maximum likelihood estimates. In the expectation step of the EM procedure, the conditional expectation of the log likelihood given observed complete data and the mean and covariance are computed. The second step of the EM procedure involves computing the log likelihood that maximizes for the mean vector and covariance matrix of the parameters coefficients (see Schafer, 1997 for a detailed description and applications of the EM algorithm). For multivariate data with $G$ groups with distinct missing patterns, the log likelihood being maximized can be expressed as

$$\begin{aligned} \log\, L\left(\theta|Y_{obs}\right) &= \sum_{g=1}^{G} \log\, L\left(\theta|Y_{obs}\right) \\ &= -\tfrac{n_g}{2} \log\left|\textstyle\sum_g\right| \; - \tfrac{1}{2} \sum_{tg} \left(y_{t,g} - \mu_g\right)' \; \textstyle\sum_g^{-1} \; \left(y_{t,g} - \mu_g\right) \end{aligned} \tag{2}$$

where $n_g$ is the number of observations in the group, $y_{t,g}$ is a vector of observed values corresponding to observed variables, $\mu_g$ is the mean vector and $\sum_g$ is the corresponding covariance matrix.

To validate or draw inferences from $m$ multiple imputed complete data sets, $m$ different sets of point estimates, $\hat{Q}_m$ and variance estimates, $\hat{W}_m$ can be computed. The combined point estimate from multiple imputation is the average of

$m$ complete data sets and defined as:

$$\bar{Q} = \frac{1}{m} \sum_{m=1}^{M} \hat{Q}_m \tag{3}$$

Similarly, the combined within-imputation variance estimate, $\bar{W}$ is the average of $m$ complete data sets and defined

$$\bar{W} = \frac{1}{m} \sum_{m=1}^{M} \hat{W}_m \tag{4}$$

and the combined between-imputation variance estimate, $B$ can be computed as

$$B = \frac{1}{m-1} \sum_{m=1}^{M} \left( \hat{Q}_m - \bar{Q} \right)^2 \tag{5}$$

The total variance estimate, $T$ can be computed from the combined within and between imputation variance as

$$T = \bar{W} + \left( 1 + \frac{1}{m} \right) B \tag{6}$$

These within, between and total variance estimates would be used to evaluate the multiple imputation of the data.

## 2.2 Second Step of Multiple Imputation

In the second stage we use two alternative methods to perform MI, namely, generalized linear models (GLM) and stochastic frontier analysis (SFA) statistical techniques.

### 2.2.1 General Linear Model Statistical Procedure

Modeling for output can be represented by general linear model in matrix notation as:

$$Y=X\beta+Z\gamma+\varepsilon \tag{7}$$

where $Y$ represents endogenous (output quantity index) variable with $t$ data points and $i$ cross section observations, $X$, the vector of explanatory (inputs quantity indexes) variables; $\beta$ is the unknown fixed-effects parameter vector, $\varepsilon$ is the unobserved vector of independent and identically distributed Gaussian random errors, $Z$ represents the random matrix, and $\gamma$ represents the associated parameters. Each model shares the exact same $X$ but different composition of the $Z$ matrix.

The composition and dimension of $Z$ depends on the use of time series (TS), panel random effects (PRE) or hierarchical linear model (HLM) statistical procedure. For example, consider a three-way panel random effects model that includes three different factors. The three factors are income groups, region and country, and are treated as independent random variables. However, the three-way panel model does not consider the common characteristics that each country shares within a particular region. In contrast, three-way hierarchical linear model does consider the commonalities that arise because of the nesting or hierarchical structure of country within a region and income groups. In a similar way, two-way panel random effects differ from two-way hierarchical linear model with respect to hierarchical structure of the two factors - country and region. If $Z$ matrix is set to zero, equation 1 boils down to a TS statistical procedure.

**2.2.1.1 Time-series Statistical Procedure** The time-series statistical procedure to examine the importance of inputs used in the production function in-

cluding technology can be represented as:

$$y_t = \alpha + \beta x_t + \varepsilon_t \tag{8}$$

where $y_t$ represents a $1 \times T$ matrix; $x_t$ represents a $K \times T$ matrix of exogenous input quantity and time trend variables with $T$ representing the temporal (time series) dimension; $\alpha$ is the intercept, $\beta$ is the associated parameters of input quantity variables; and $\varepsilon$ represents a $1 \times T$ matrix of pure random error.

### 2.2.2 Stochastic Frontier Analysis Statistical Procedure

In 1977, Aigner Lovell and Schmidt, Meeusen and van den Broeck, and Battese and Corra simultaneously introduced the stochastic frontier model that decomposes the error term, $\varepsilon$ into a symmetrical random error, $v$ and a one-sided error or inefficiency, $u$. The normal-half normal and an exponential distribution was assumed by Aigner, Lovell and Schmidt (1977), while Meeusen and van den Broeck (1977) assumed an exponential distribution of the inefficiency term.

In 1982, Jondrow, Materov, Lovell and Schmidt suggested a method to estimate firm specific inefficiency measures. The stochastic frontier model can be used to represent a Cobb-Douglas production function as

$$y_t = \alpha + \beta x_t + v_t - u_t \equiv \alpha + \beta x_t + \varepsilon_t \tag{9}$$

where $y_t$ represents a $1 \times T$ matrix; $x_t$ represents a $K \times T$ matrix of exogenous input quantity and time trend variables with $T$ representing the temporal (time series) dimension; $\alpha$ is the intercept, $\beta$ is the associated parameters of input quantity variables; and $\varepsilon$ represents a $1 \times T$ matrix of pure random error and decomposed

into $v$ represents the random error and $v \sim N(0, \sigma_v^2)$, $u$ represents the negatively skewed one-sided inefficiency and can be represented with alternative distributions including half normal, exponential, or truncated normal distribution. Details for the distribution can be found in the set of 1977 articles.

**2.2.2.1 Half-Normal Distribution of Stochastic Frontier Analysis** For the normal-half normal distribution assumed by Aigner, Lovell and Schmidt (1977), the joint density of random error $v \sim N(0, \sigma_v^2)$ and $u \sim N(0, \sigma_u^2)$ can be written as

$$f(u, v) = \frac{2}{2\pi\sigma_u\sigma_v} \exp\left\{-\frac{u^2}{2\sigma_u} - \frac{(\varepsilon + u)^2}{2\sigma_v}\right\} \tag{10}$$

For convenient parameterization, substituting $v = \varepsilon + u$ and integrating $u$ out to obtain the marginal density function of $\varepsilon$ can be written as

$$f(\varepsilon) = \int_0^\infty f(u, \varepsilon)\, du = \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\Phi\left(-\frac{\varepsilon\lambda}{\sigma}\right) \tag{11}$$

where $\phi$ is the standard normal density, $\Phi$ is the standard normal cumulative distribution function (CDF), $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$ and $\lambda = \sigma_u/\sigma_v$.

The likelihood function for the production function can be written as:

$$\ln L(\alpha, \beta, \sigma, \lambda) = constant - T \ln \sigma + \sum_{t=1}^{T} \ln \Phi\left(\frac{-\varepsilon_t\lambda}{\sigma}\right) - \frac{1}{2}\left(\frac{\varepsilon_t}{\sigma}\right)^2 \tag{12}$$

where $\varepsilon_t = \ln y_t - \alpha - \beta x_t$ and others are defined above.

**2.2.2.2 Exponential Distribution of Stochastic Frontier Analysis** For the exponential distribution assumed by Meeusen and van den Broeck (1977) and Aigner, Lovell and Schmidt (1977), the joint density of random error $v \sim N(0, \sigma_v^2)$

and $u \sim N(0, \sigma_u^2)$ can be written as

$$f(u,v) = \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \exp\left\{-\frac{u}{\sigma_u} - \frac{v^2}{2\sigma_v}\right\} \tag{13}$$

For convenient parameterization, substituting $v = \varepsilon + u$ and integrating $u$ out to obtain the marginal density function of $\varepsilon$ can be written as

$$f(\varepsilon) = \int_0^\infty f(u,\varepsilon)\, du = \frac{1}{\sigma_u}\Phi\left(-\frac{\varepsilon}{\sigma_v} - \frac{\sigma_v}{\sigma_u}\right) \exp\left(\frac{\varepsilon}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2}\right) \tag{14}$$

and the likelihood function can be written as:

$$\ln L(\alpha,\beta,\sigma,\lambda) = constant - T \ln \sigma + T\left(\frac{\sigma_v^2}{2\sigma_u^2}\right) + \sum_{t=1}^{T}\left(\frac{\varepsilon_t}{\sigma_u}\right) + \sum_{t=1}^{T}\left\{\ln \Phi\left(\frac{\varepsilon}{\sigma_v} - \frac{\sigma_v}{\sigma_u}\right)\right\} \tag{15}$$

**2.2.2.3 Truncated Distribution of Stochastic Frontier Analysis** For the truncated distribution assumed by Aigner, Lovell and Schmidt (1977), the joint density of random error $v \sim N(0, \sigma_v^2)$ and $u \sim N(0, \sigma_u^2)$ can be written as

$$f(u,v) = \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v\Phi(\mu/\sigma_u)} \exp\left\{-\frac{(u-\mu)^2}{2\sigma_u} - \frac{v^2}{2\sigma_v}\right\} \tag{16}$$

For convenient parameterization, substituting $v = \varepsilon + u$ and integrating $u$ out to obtain the marginal density function of $\varepsilon$ can be written as

$$f(\varepsilon) = \int_0^\infty f(u,\varepsilon)\, du = \frac{1}{\sigma_u}\phi\left(\frac{\varepsilon+\mu}{\sigma}\right)\Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\varepsilon\lambda}{\sigma}\right)\left\{\Phi\left(\frac{\mu}{\sigma_u}\right)\right\}^{-1} \tag{17}$$

and the likelihood function can be written as:

$$\begin{aligned} \ln L(\alpha, \beta, \sigma, \lambda) = constant - T \ln \sigma - T \ln \Phi\left(\tfrac{\mu}{\sigma_u}\right) + \\ \sum_{t=1}^{T} \ln \Phi\left(\tfrac{\mu}{\sigma\lambda} + \tfrac{\varepsilon_t \lambda}{\sigma}\right) \quad - \quad \sum_{t=1}^{T} \left(\tfrac{\varepsilon_t + \mu}{\sigma}\right)^2 \end{aligned} \tag{18}$$

### 2.3 Third Step of Multiple Imputation

In the third step of the MI procedure, the parameter coefficients and the technical efficiency measures estimated by stochastic frontier analysis production function in the second step from imputed complete data sets is analyzed to draw statistical inferential. Specifically, the third step involves two sets of analysis to evaluate the importance of MI. First, the production function parameter coefficients and its covariance matrix estimated by GLM and SFA for each imputed data set is used to draw statistical inferential about the parameter coefficients. Second, the mean and standard errors of the production function technical efficiency measures estimated by SFA under three alternative distributions for each of the imputed data set is used to draw statistical inferences.

## 3 Data and Variables used in the Empirical Analysis

This study is based on Food and Agricultural Organization data available online. The study includes 92 countries for the period 1961 to 2007. The set of 92 countries include 10 low income countries, 37 lower middle income countries, 20 upper middle income, 22 high income Organisation for Economic Co-operation and Development countries (OECD) and 3 high income non-OECD countries. For the output and the five inputs, a quantity index with 1961 as the base year was

constructed.

Due to the problems of estimating multiple outputs in primal production functions, an aggregate output variable published by FAO is used in the analysis. The FAO output concept is the output from the agriculture sector net of quantities of various commodities used as feed and seed, which is why feed and seed are not included in the input series. Details on the construction of the aggregate output variable are available on FAO webpage, www.fao.org.

This analysis considers only five input variables following earlier studies estimating a production function. These variables include land, labor, capital, fertilizer and livestock. The land variable includes harvested acres of cereals, fibers, fruits, nuts, oil crops, pulses, roots and tubers, rubber, spices, stimulants, sugar crops, tobacco and vegetables unlike earlier studies that use land under cultivation. The capital variable covers the total number of agricultural tractors, and number of harvesters and threshers used in agriculture. With respect to tractors, no allowance was made to the quality (horsepower) of the tractors. The labor variable refers to the economically active population in agriculture. An economically active population is defined as all persons engaged or seeking employment in an economic activity, whether as employers, own-account workers, salaried employees, or unpaid workers assisting in the operation of a family farm or business. The economically active population in agriculture includes all economically active persons engaged in agriculture, forestry, hunting, or fishing. This variable obviously overstates the labor input used in agricultural production, but the extent of overstatement depends on the level of development of the country. Following other studies on inter-country comparisons of agricultural productivity, this analysis uses the sum of nitrogen, potassium, and phosphate contained in the commercial fertilizers consumed. This variable is expressed in thousands of metric tons.

The livestock input variable used in the study is the sheep-equivalent of five categories of animals. The categories considered are buffaloes, cattle, goats, pigs and sheep. The number of these animals is converted into sheep equivalents using conversion factors of 8.0 for buffalo and cattle and 1.00 for sheep, goats and pigs. Chicken numbers are not included in the livestock figures. Table I presents the summary statistics of the output and inputs variables by income group.

# 4 Empirical Application and Results

To evaluate the importance of missing information, the missing data patterns by income groups for capital and fertilizer variables along with other inputs is presented in table II. The $'X'$ mark indicates non-missing data and a dot $'.'$ indicates missing data. Within each income group, there are four categories based on missing pattern with the exception of high income non-OECD countries. The top four income groups had similar missing patterns with independently missing capital and fertilizer, jointly missing capital-fertilizer, and complete capital and fertilizer data. The missing observations range from 3.5 percent for high income OECD countries to 8.5 percent for high income non-OECD countries. However, the means of the inputs are not only different across the four missing pattern categories but also by income groups.

To evaluate the importance of alternative distributional assumptions – half-normal, truncation and exponential, the technical efficiency measures are estimated without the missing data. Then these three datasets are then used for MI of the missing capital and fertilizer data. Specifically, the importance of MI on stochastic frontier analysis technical efficiency measures estimated under alternative distributional assumptions is evaluated. Second, we also wanted to compare

the within and between variances of the production function parameter coefficients estimated for multiple imputed datasets under the alternative distributional assumptions – half-normal, truncation and exponential.

Next, to evaluate the importance of missing information, the within, between and total variance measures of multiple imputed data for the two input variables – capital and fertilizer by three alternative distributional assumptions are presented in table III. The results for table III suggest very low or nearly insignificant between-variance compared to within-variance of capital and fertilizer variables across the income group countries and three alternative distributional assumptions of SFA models. This would suggest a low or insignificant influence on technical efficiency measures across the three SFA models.

To reflect the importance of accounting for inefficiency on the parameter coefficients of imputed missing data in the second step of multiple imputation, equation (7) is estimated using GLM statistical procedure, and equation (8) is estimated using SFA statistical procedures under three alternative distributional assumptions – half-normal, truncation and exponential.

We specified a Cobb-Douglas functional form for half-normal, truncation and exponential stochastic frontier production function models. The Cobb-Douglas production function for a GLM was specified as:

$$Output_{it} = \beta_0 + \beta_1 Capital_{it} + \beta_2 Land_{it} + \beta_3 Labor_{it} + \beta_4 Fertilizer_{it} + \beta_5 Livestock_{it} + \varepsilon_{it} \tag{19}$$

and for SFA was specified as:

$$Output_{it} = \beta_0 + \beta_1 Capital_{it} + \beta_2 Land_{it} + \beta_3 Labor_{it} + \beta_4 Fertilizer_{it} + \beta_5 Livestock_{it} + v_{it} - u_{it} \tag{20}$$

In the third step of the multiple imputation, the parameter coefficients and the technical efficiency measures estimated by SFA and GLM production function are analyzed to draw statistical inferential. Specifically, we use the parameter estimates and associated covariance matrix computed by the GLM and SFA for each imputed dataset to draw statistical inferential about the parameter coefficients or input elasticity of production function. The results for the exponential, half-normal and truncation SFA models are presented in table IV, V and VI respectively. The parameter coefficients estimated by the GLM statistical procedure are also presented to evaluate the importance of accounting for inefficiency on the parameter coefficients. Note that we do not present the results for the high income (OECD), high income (nonOECD), and lower middle income countries due to presence of close to zero values in the covariance matrix. Due to the presence of zero in the covariance matrix, the between-variance and within-variance cannot be estimated.

It is particularly interesting to note that between-variance and within-variances of the production function's input elasticity or parameter coefficients estimated by SFA and GLM models are very small. In the case of low income countries, the between-variance of the production function input elasticity estimated by exponential and truncated SFA models is higher than the between-variance production function input elasticity estimated by GLM model, as evident by the ratio of SFA/GLM (last 3 column). The difference in the between-variance of the production function input elasticity estimated by SFA and GLMs model are much higher compared to difference in the within-variance of the production function input elasticity estimated by SFA and GLM models, which are less than 100. Further, total variance (ratio of SFA/GLM) and within-variance of the production function input elasticity are below 100 for the half-normal SFA model. In the half-normal

SFA model, we also note that the difference in the between-variance between SFA and GLM models production function input elasticity is much higher compared to the within-variance difference between SFA and GLM models.

However, the opposite is true when one considers the case of upper middle income countries. Specifically, for exponential and truncated models the difference in the between-variance between SFA and GLM models production function input elasticity is much lower compared to difference between within-variance SFA and GLM models. However, in the case of half-normal SFA model, we observe that the difference in the between-variance of production function input elasticity estimated by SFA and GLM models are much higher compared to within-variance, as evident by less than 100 percent variation in between-variance category (table V, column 9-second panel). These findings suggest the importance of correcting for inefficiency when estimating output and technical efficiency, while imputing for missing values.

Statistical inferential about efficiency measures estimated by the SFA statistical procedures are drawn using the mean and standard errors of the technical efficiency estimated from univariate statistics. Inference based on t-tests is also derived for the difference in the technical efficiency measures estimated from completed data and multiple imputed datasets. The results for the exponential, half-normal and truncation SFA models technical efficiency measures are presented in table VII by income groups.

The difference in technical efficiency measures estimated by exponential (table VII) SFA model for actual and imputed values (using multiple imputed datasets values) is significant for the all income groups with the exception of low income group countries. The highest difference in parameter estimates for actual and imputed values in exponential SFA model is observed in the case of upper middle

income groups countries, followed by lower middle income group countries, high income (non-OECD), and high income (OECD) group countries. This could be due to the distributional assumption of the inefficiency and the inability to account for inefficiency measures during MI of the missing data. In the case of half normal (table VII) SFA model, we observe that the technical efficiency measures for actual and imputed values (using multiple imputed dataset values) are slightly lower than those obtained in the exponential SFA model. Table VII also presents technical efficiency measures for actual and imputed values for truncated SFA model. However, technical efficiency estimated by truncated SFA model for completed data is about half of those obtained in exponential and half normal SFA models. Finally, the difference in parameter estimates for actual and imputed values (using imputed data values) is significant in all but one case – lower middle income group countries.

Further, the within, between and total variance measures of technical efficiency measures from multiple imputed datasets are presented in table VIII. Results in table VIII show that between-variance is very low for all three models of SFA and zero for three income groups, namely, upper middle income, high income (OECD), and high income (non-OECD) group countries. This suggests the efficiency measures from multiple imputation is very close to the actual efficiency measures. On the other hand, within-variance is much higher compared to between-variance, particularly in the case of low income group countries. Findings here suggest the importance of missing values and the imputation of such missing values on technical efficiency measures.

# 5 Challenges and Conclusions

The contribution of the research presented in this paper is twofold. First, the importance of accounting for inefficiency on the multiple imputation of missing data is evaluated by comparing the stochastic frontier analysis to the generalized linear models statistical procedures in the second step of the multiple imputation process. In particular, we use and compare the alternative distributional assumptions of stochastic frontier analysis – half-normal, truncation and exponential. Second, the importance of multiple imputations on stochastic frontier analysis technical efficiency measures under alternative distributional assumptions – half-normal, truncation and exponential is evaluated.

Empirical estimates indicate difference in the between-variance and within-variance of production function input elasticity or parameter coefficients estimated from SFA compared to the GLM statistical procedures. The between-variance and within-variance of technical efficiency measures are also different across the three alternative distributional assumptions of SFA statistical procedures. Further, the technical efficiency measures are different with complete data and multiple imputed dataset. Finally, results from this study indicate that even though the between-variance and within-variance of multiple imputed datasets is almost zero, the between-variance and within-variance of production function parameters as well as the technical efficiency measures is statistically different.

Future research could examine the implications of accounting for heteroskedasticity not only in the multiple imputation of missing data but also in the SFA and GLM statistical procedures in the second step of the MI procedure. Further, research needs to be done with varying number of cross-sections and number of years. Compared to aggregate production analysis, individual farm-level data re-

sults may vary with regard to production of agricultural output and the impact on technical efficiency measures.

# References


[1] Aigner, D.J., Lovell, C.A.K., and Schmidt, P. 1977. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics*, 6: 21-37.

[2] Allison, P.D. 2000. *Missing Data*, Thousand Oaks, CA Sage Publication.

[3] Battese, G. and Corra, G. 1977. "Estimation of a production frontier model: With application for the pastoral zone of eastern Australia." *Australian Journal of Agricultural Economics*, 21: 167-179.

[4] Jondrow, J., Materov, I., Lovell, C.A.K., and Schmidt, P. 1982. "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model." *Journal of Econometrics*, 19: 233-238.

[5] Meeusen, W. and van den Broeck, J. 1977. "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error." *International Economic Review*, 18: 435-444.

[6] Rubin, D. B. 1976. "Inference and Missing Data." *Biometrika*, 63: 581-592.

[7] Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley and Sons.

[8] Rubin, D. B. 1996. "Multiple Imputation after 18+ Years." *Journal of the American Statistical Association*, 91: 473-489.

[9] Schafer, J. L. 1997. Analysis of Incomplete Multivariate Data, New York: Chapman and Hall.

[10] Schafer, J. L. 1999. "Multiple Imputation: A Primer." *Statistical Methods in Medical Research*, 8: 3-15.

[11] Tabachnick, B.G., and Fidell, L.S. 2000. *Using multivariate statistics (4th eds.)*, New York: Harper Collins College Publishers.

Table 1: Summary Statistics of Output and Input Variables of World Agriculture Sector, 1961-2007

| Income Group | Variable | Nonmissing | Missing | Mean | Standard |
|---|---|---|---|---|---|
| Low | Output | 470 | 0 | 72.087 | 27.86 |
| | Land | 470 | 0 | 84.091 | 27.13 |
| | Labor | 470 | 0 | 74.085 | 22.83 |
| | Capital | 451 | 19 | 496.99 | 4256.71 |
| | Fertilizer | 448 | 22 | 98.944 | 126.69 |
| | Livestock | 470 | 0 | 86.262 | 19.09 |
| Lower middle | Output | 799 | 0 | 67.897 | 29.33 |
| | Land | 799 | 0 | 90.733 | 37.43 |
| | Labor | 799 | 0 | 82.667 | 22.37 |
| | Capital | 769 | 30 | 84.945 | 713.6 |
| | Fertilizer | 771 | 28 | 69.331 | 83.23 |
| | Livestock | 799 | 0 | 84.89 | 30.24 |
| Upper middle | Output | 940 | 0 | 88.714 | 34.2 |
| | Land | 940 | 0 | 102.37 | 37.18 |
| | Labor | 940 | 0 | 109.04 | 67.39 |
| | Capital | 922 | 18 | 194.69 | 570.52 |
| | Fertilizer | 900 | 40 | 116.29 | 124.64 |
| | Livestock | 940 | 0 | 100.16 | 40.33 |
| High:OECD | Output | 1034 | 0 | 89.082 | 18.34 |
| | Land | 1034 | 0 | 106.7 | 23.02 |
| | Labor | 1034 | 0 | 148.55 | 52.48 |
| | Capital | 1006 | 28 | 79.838 | 44.73 |
| | Fertilizer | 1012 | 22 | 121.63 | 63.26 |
| | Livestock | 1034 | 0 | 106.33 | 27.52 |
| High:nonOECD | Output | 141 | 0 | 85.104 | 21.61 |
| | Land | 141 | 0 | 143.38 | 64.66 |
| | Labor | 141 | 0 | 185.42 | 117.6 |
| | Capital | 129 | 12 | 63.382 | 60.15 |
| | Fertilizer | 138 | 3 | 122.5 | 58.7 |
| | Livestock | 141 | 0 | 100.24 | 40.93 |

Table 2: Missing Data Patterns of the Variables, 1961-2007

| Income Group | Missing Patterns | | | | | | Group Means | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Land | Labor | Capital | Livestock | Fertilizer | Percent | Land | Labor | Capital | Livestock | Fertilizer |
| Low | X | X | X | X | X | 92.55 | 77.51 | 68.39 | 32.13 | 82.62 | 46.32 |
| | X | X | X | X | . | 3.4 | 93.71 | 88.7 | 445.2 | 97.49 | . |
| | X | X | . | X | X | 2.77 | 120.87 | 111.29 | . | 105.95 | 179.51 |
| | X | X | . | X | . | 1.28 | 117.99 | 116.32 | . | 113.71 | . |
| Lower middle | X | X | X | X | X | 94.24 | 83.72 | 78.25 | 26.85 | 77.69 | 39.12 |
| | X | X | X | X | . | 2 | 92.38 | 100.98 | 173.71 | 110.34 | . |
| | X | X | . | X | X | 2.25 | 103.18 | 103.51 | . | 103.62 | 106.85 |
| | X | X | . | X | . | 1.5 | 107.45 | 99.58 | . | 109.28 | . |
| Upper middle | X | X | X | X | X | 94.68 | 94.42 | 99.48 | 71.88 | 93.43 | 72.18 |
| | X | X | X | X | . | 3.4 | 90.54 | 100.11 | 119.01 | 93.95 | . |
| | X | X | . | X | X | 1.06 | 97.38 | 95.87 | . | 98.67 | 109.07 |
| | X | X | . | X | . | 0.85 | 112.88 | 97.79 | . | 98.07 | . |
| High:OECD | X | X | X | X | X | 96.52 | 104.84 | 142.78 | 54.72 | 103.31 | 110.9 |
| | X | X | X | X | . | 0.77 | 95.62 | 82.42 | 126.92 | 100.64 | . |
| | X | X | . | X | X | 1.35 | 96.62 | 83.49 | . | 94.12 | 104.81 |
| | X | X | . | X | . | 1.35 | 96.96 | 80.89 | . | 94.45 | . |
| High:nonOECD | X | X | X | X | X | 91.49 | 134.775 | 161.81 | 45.018 | 91.63 | 110.35 |
| | X | X | . | X | X | 6.38 | 103.14 | 88.59 | . | 101.49 | 138.48 |
| | X | X | . | X | . | 2.13 | 98.479 | 87.74 | . | 100.85 | . |

Table 3: Multiple Imputed Between, Within and Total Variances of Missing Capital and Fertilizer Variables

| Group | Variable | Between | Within | Total |
|---|---|---|---|---|
| | | SFA Exponential Model | | |
| High:OECD | Capital | 0.000038724 | 0.001864 | 0.001907 |
| | Fertilizer | 0.000001349 | 0.000178 | 0.000179 |
| High:nonOECD | Capital | 0.000267 | 0.006535 | 0.006829 |
| | Fertilizer | 0.000013687 | 0.001274 | 0.001289 |
| Low | Capital | 0.000521 | 0.01773 | 0.018303 |
| | Fertilizer | 0.000098115 | 0.004406 | 0.004514 |
| Lower middle | Capital | 0.000058988 | 0.00284 | 0.002904 |
| | Fertilizer | 0.000034226 | 0.00201 | 0.002048 |
| Upper middle | Capital | 0.000010762 | 0.00129 | 0.001302 |
| | Fertilizer | 0.000076046 | 0.001161 | 0.001245 |
| | | SFA Half-Normal Model | | |
| High:OECD | Capital | 0.00003685 | 0.001865 | 0.001906 |
| | Fertilizer | 0.000002488 | 0.000178 | 0.00018 |
| High:nonOECD | Capital | 0.000184 | 0.006587 | 0.006789 |
| | Fertilizer | 0.000023941 | 0.001275 | 0.001302 |
| Low | Capital | 0.000804 | 0.017678 | 0.018563 |
| | Fertilizer | 0.000080056 | 0.004409 | 0.004497 |
| Lower middle | Capital | 0.000095047 | 0.002836 | 0.00294 |
| | Fertilizer | 0.000036653 | 0.002005 | 0.002046 |
| Upper middle | Capital | 0.000009803 | 0.00129 | 0.0013 |
| | Fertilizer | 0.00003598 | 0.001169 | 0.001208 |
| | | SFA Half-Normal Model | | |
| High:OECD | Capital | 0.000057504 | 0.001859 | 0.001922 |
| | Fertilizer | 0.000003493 | 0.000178 | 0.000182 |
| High:nonOECD | Capital | 0.000127 | 0.006571 | 0.00671 |
| | Fertilizer | 0.000010054 | 0.001268 | 0.001279 |
| Low | Capital | 0.00174 | 0.017605 | 0.019519 |
| | Fertilizer | 0.000203 | 0.004354 | 0.004578 |
| Lower middle | Capital | 0.000045849 | 0.002856 | 0.002906 |
| | Fertilizer | 0.000023606 | 0.002005 | 0.002031 |
| Upper middle | Capital | 0.000018979 | 0.001293 | 0.001314 |
| | Fertilizer | 0.000035114 | 0.001157 | 0.001196 |

Table 4: Between, Within and Total Variances of Regression Parameters from SFA Exponential Model

| | Stochastic Frontier Analysis | | | Generalized Linear Models | | | Ratio of SFA/HLM | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | Between | Within | Total | Between | Within | Total | Between | Within | Total |
| | | | | Low Income Group | | | | | |
| Intercept | 0.026083 | 0.023167 | 0.051858 | 0.000533 | 0.032497 | 0.033084 | 4894% | 71% | 157% |
| Land | 0.000057 | 0.000539 | 0.000602 | 0.000016 | 0.000815 | 0.000832 | 369% | 66% | 72% |
| Labor | 0.005604 | 0.002176 | 0.00834 | 0.000129 | 0.003128 | 0.00327 | 4344% | 70% | 255% |
| Capital | 0 | 0.000003 | 0.000003 | 0.000003 | 0.000009 | 0.000012 | 12% | 32% | 27% |
| Livestock | 0.000294 | 0.000694 | 0.001017 | 0.00002 | 0.00126 | 0.001283 | 1436% | 55% | 79% |
| Fertilizer | 0.000003 | 0.000019 | 0.000022 | 0.000006 | 0.000027 | 0.000034 | 45% | 70% | 65% |
| trend | 0.000001 | 0.000001 | 0.000002 | 0 | 0.000001 | 0.000002 | 1699% | 59% | 147% |
| Sigma_v | 0.000306 | 0.000025 | 0.000361 | | | | | | |
| Sigma_u | 0.018199 | 0.22441 | 0.244429 | | | | | | |
| | | | | Upper Middle Income Group | | | | | |
| Intercept | 0.000246 | 0.05086 | 0.051131 | 0.001431 | 0.025759 | 0.027333 | 17% | 197% | 187% |
| Land | 0.000006 | 0.00104 | 0.001047 | 0.00001 | 0.000394 | 0.000405 | 66% | 264% | 259% |
| Labor | 0.000008 | 0.002883 | 0.002891 | 0.000041 | 0.001205 | 0.00125 | 19% | 239% | 231% |
| Capital | 0 | 0.000197 | 0.000198 | 0.000014 | 0.000117 | 0.000133 | 3% | 168% | 149% |
| Livestpck | 0.000006 | 0.00254 | 0.002546 | 0.000036 | 0.001169 | 0.001209 | 15% | 217% | 211% |
| Fertilizer | 0.000006 | 0.000285 | 0.000292 | 0.000092 | 0.000102 | 0.000204 | 7% | 279% | 143% |
| trend | 0 | 0.000001 | 0.000002 | 0 | 0 | 0 | 109% | 400% | 379% |
| Sigma_v | 0 | 0.000369 | 0.000369 | | | | | | |
| Sigma_u | 0.000001 | 0.00062 | 0.000621 | | | | | | |

Table 5: Between, Within and Total Variances of Regression Parameters from SFA Half-Normal Model

| | Stochastic Frontier Analysis | | | Generalized Linear Models | | | Ratio of SFA/HLM | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | Between | Within | Total | Between | Within | Total | Between | Within | Total |
| | | | | Low Income Group | | | | | |
| Intercept | 0.000674 | 0.026052 | 0.026793 | 0.000446 | 0.032539 | 0.03303 | 151% | 80% | 81% |
| Land | 0.000017 | 0.000521 | 0.00054 | 0.000012 | 0.000815 | 0.000827 | 148% | 64% | 65% |
| Labor | 0.000031 | 0.00236 | 0.002394 | 0.000044 | 0.003128 | 0.003177 | 70% | 75% | 75% |
| Capital | 0 | 0.000003 | 0.000003 | 0.000001 | 0.000009 | 0.00001 | 27% | 31% | 31% |
| Livestpck | 0.000024 | 0.000788 | 0.000814 | 0.000017 | 0.001262 | 0.001281 | 141% | 62% | 64% |
| Fertilizer | 0.000001 | 0.000018 | 0.000018 | 0.000002 | 0.000027 | 0.000029 | 30% | 66% | 63% |
| trend | 0 | 0.000001 | 0.000001 | 0 | 0.000001 | 0.000001 | 63% | 63% | 63% |
| Sigma_v | 0.000004 | 0.000028 | 0.000032 | | | | | | |
| Sigma_u | 0.001564 | 0.250754 | 0.252475 | | | | | | |
| | | | | Upper Middle Income Group | | | | | |
| Intercept | 0.000128 | 0.0509 | 0.051041 | 0.000696 | 0.025908 | 0.026674 | 18% | 196% | 191% |
| Land | 0.000002 | 0.00104 | 0.001043 | 0.000004 | 0.000397 | 0.000401 | 59% | 262% | 260% |
| Labor | 0.000012 | 0.002883 | 0.002896 | 0.000038 | 0.001216 | 0.001258 | 30% | 237% | 230% |
| Capital | 0.000001 | 0.000198 | 0.000198 | 0.000023 | 0.000115 | 0.000141 | 3% | 172% | 140% |
| Livestpck | 0.000002 | 0.002543 | 0.002545 | 0.000009 | 0.001176 | 0.001186 | 24% | 216% | 215% |
| Fertilizer | 0.000006 | 0.000284 | 0.00029 | 0.000054 | 0.000103 | 0.000162 | 11% | 276% | 179% |
| trend | 0 | 0.000002 | 0.000002 | 0 | 0 | 0 | 49% | 399% | 375% |
| Sigma_v | 0 | 0.00037 | 0.00037 | | | | | | |
| Sigma_u | 0.000002 | 0.000619 | 0.00062 | | | | | | |

Table 6: Between, Within and Total Variances of Regression Parameters from SFA Truncated Model

| | Stochastic Frontier Analysis | | | Generalized Linear Models | | | Ratio of SFA/HLM | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | Between | Within | Total | Between | Within | Total | Between | Within | Total |
| | | | | Low Income Group | | | | | |
| Intercept | 0.026552 | 0.023543 | 0.052751 | 0.000449 | 0.032699 | 0.033193 | 5914% | 72% | 159% |
| Land | 0.000038 | 0.000541 | 0.000584 | 0.000016 | 0.000816 | 0.000834 | 245% | 66% | 70% |
| Labor | 0.005739 | 0.00224 | 0.008553 | 0.000076 | 0.003145 | 0.003229 | 7575% | 71% | 265% |
| Capital | 0.000001 | 0.000003 | 0.000004 | 0.000003 | 0.000009 | 0.000012 | 25% | 32% | 30% |
| Livestpck | 0.000382 | 0.000716 | 0.001136 | 0.000013 | 0.001264 | 0.001279 | 2922% | 57% | 89% |
| Fertilizer | 0.000002 | 0.000019 | 0.000021 | 0.000009 | 0.000027 | 0.000037 | 22% | 70% | 58% |
| trend | 0.000002 | 0.000001 | 0.000003 | 0 | 0.000001 | 0.000002 | 2162% | 61% | 166% |
| Sigma_v | 0.000332 | 0.000025 | 0.000391 | | | | | | |
| Sigma_u | 0.016233 | 0.192854 | 0.21071 | | | | | | |
| | | | | Upper Middle Income Group | | | | | |
| Intercept | 0.000121 | 0.050913 | 0.051046 | 0.000728 | 0.025836 | 0.026637 | 17% | 197% | 192% |
| Land | 0.000001 | 0.00104 | 0.001042 | 0.000002 | 0.000396 | 0.000398 | 97% | 263% | 262% |
| Labor | 0.000006 | 0.002882 | 0.002889 | 0.000029 | 0.001209 | 0.001241 | 22% | 238% | 233% |
| Capital | 0.000001 | 0.000197 | 0.000198 | 0.000013 | 0.000117 | 0.000131 | 5% | 168% | 151% |
| Livestpck | 0.000002 | 0.002542 | 0.002545 | 0.000012 | 0.001174 | 0.001187 | 19% | 217% | 214% |
| Fertilizer | 0.000003 | 0.000286 | 0.000289 | 0.000048 | 0.000104 | 0.000157 | 5% | 275% | 184% |
| trend | 0 | 0.000001 | 0.000002 | 0 | 0 | 0 | 104% | 399% | 389% |
| Sigma_v | 0 | 0.000369 | 0.000369 | | | | | | |
| Sigma_u | 0.000001 | 0.000619 | 0.00062 | | | | | | |

Table 7: Actual, Multiple Imputed and Difference in the Technical Efficiency Measures Estimated from Three SFA Models

| Income Group | Variable | Exponential | | | Half-Normal | | | Truncated | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Estimate | StdErr | tValue | Estimate | StdErr | tValue | Estimate | StdErr | tValue |
| Low | Actual | 0.9 | 0.00394 | | 0.87877 | 0.00385 | | 0.49189 | 0.00138 | |
| | MI | 0.8843 | 0.88642 | | 0.89897 | 0.89897 | | 0.8849 | 0.88697 | |
| | Difference | -0.018 | 0.04567 | -0.39 | 0.01784 | 0.00132 | 13.47 | 0.39062 | 0.04496 | 8.69 |
| Lower middle | Actual | 0.9096 | 0.00216 | | 0.8884 | 0.00172 | | 0.47907 | 0.00047 | |
| | MI | 0.4984 | 0.50265 | | 0.48958 | 0.49358 | | 0.49841 | 0.50257 | |
| | Difference | -0.412 | 0.04941 | -8.34 | -0.4 | 0.04747 | -8.43 | 0.01855 | 0.04912 | 0.38 |
| Upper middle | Actual | 0.9998 | 2.92E-09 | | 0.99999 | 8.28E-12 | | 0.54401 | 0.00341 | |
| | MI | 0.5419 | 0.54193 | | 0.54192 | 0.54192 | | 0.54192 | 0.54192 | |
| | Difference | -0.457 | 0.00337 | -135.72 | -0.4571 | 0.00336 | -135.97 | -0.0013 | 0.00026 | -4.9 |
| High:nonOECD | Actual | 0.9287 | 0.00437 | | 0.85716 | 0.00805 | | 0.54561 | 0.00685 | |
| | MI | 0.5675 | 0.56751 | | 0.56838 | 0.56839 | | 0.56834 | 0.56835 | |
| | Difference | -0.3666 | 0.00534 | -68.63 | -0.2941 | 0.00816 | -36.03 | 0.01744 | 0.00357 | 4.89 |
| High:OECD | Actual | 0.7438 | 0.00352 | | 0.90766 | 0.00142 | | 0.5456 | 0.00209 | |
| | MI | 0.5394 | 0.53942 | | 0.5395 | 0.5395 | | 0.53943 | 0.53943 | |
| | Difference | -0.2067 | 0.00151 | -136.81 | -0.3702 | 0.0014 | -264.97 | -0.0084 | 0.00045 | -18.81 |

Table 8: Between, Within and Total Variances of SFA Technical Efficiency Measures

| SFA model | Income Group | Between Variance | Within Variance | Total Variance |
|---|---|---|---|---|
| Exponential | Low income | 0.001888 | 0.783662 | 0.78574 |
| | Lower middle | 0.002118 | 0.250327 | 0.252656 |
| | Upper middle | 0 | 0.293691 | 0.293691 |
| | High:nonOECD | 0.0000087 | 0.322055 | 0.322064 |
| | High:OECD | 0 | 0.290971 | 0.290971 |
| Half | Low income | 0.0000012 | 0.808151 | 0.808152 |
| | Lower middle | 0.001965 | 0.241455 | 0.243616 |
| | Upper middle | 0 | 0.293678 | 0.293678 |
| | High:nonOECD | 0.0000074 | 0.323062 | 0.32307 |
| | High:OECD | 0 | 0.291059 | 0.291059 |
| Truncated | Low income | 0.001833 | 0.784702 | 0.786719 |
| | Lower middle | 0.002081 | 0.250287 | 0.252576 |
| | Upper middle | 0 | 0.293675 | 0.293675 |
| | High:nonOECD | 0.0000069 | 0.323013 | 0.323021 |
| | High:OECD | 0 | 0.29098 | 0.29098 |