Stochastic Frontier Production Function With Errors-In-Variables

Rajeev Dhawan[*] [†]
Anderson Graduate School of Management
University of California, Los Angeles

&

Peter Jochumzen
Department of Economics
Lund University

July 1999

JEL Codes: C15;C21;D24
Keywords: Errors-In-Variables; Stochastic Frontier; Technical Efficiency; Reliability Ratio

Abstract

This paper develops a procedure for estimating parameters of a cross-sectional stochastic frontier production function when the factors of production suffer from measurement errors. Specifically, we use Fuller's (1987) reliability ratio concept to develop an estimator for the model in Aigner et al (1977). Our Monte-Carlo simulation exercise illustrates the direction and the severity of bias in the estimates of the elasticity parameters and the returns to scale feature of the production function when using the traditional maximum-likelihood estimator (MLE) in presence of measurement errors. In contrast the reliability ratio based estimator consistently estimates these parameters even under extreme degree of measurement errors. Additionally, estimates of firm level technical efficiency are severely biased for traditional MLE in comparison to reliability ratio estimator, rendering inter-firm efficiency comparisons infeasible. The seriousness of measurement errors in a practical setting is demonstrated by using data for a cross-section of publicly traded U.S. corporations.

[*] Corresponding Author: The mailing address is 110 Westwood Plaza, Los Angeles, CA 90095. E-mail: rdhawan@anderson.ucla.edu, Phone: (310) 206-2529 and Fax: (310) 206-9940.

[†]This research originated in Jochumzen's Ph.D. dissertation. The authors wish to thank Alok Bohara, Kurt Brännäs, David Edgerton, Kishore Gawande and Almas Heshmati for their numerous comments and suggestions that substantially improved this paper. We would also like to thank seminar participants at Lund University, University of New Mexico, conference participants at the econometric conference held at Stockholm School of Economics in May 1997, and the 3rd Biennial Georgia Productivity Workshop held at University of Georgia in October 1998 for their helpful comments and suggestions. However, any errors are the responsibility of the authors.

1. Introduction

Since the development of the stochastic frontier production function (SFPF) by Aigner et al (1977) and Meeusen and van den Broeck (1977), evaluating the efficiency of individual firms and industries has become popular with the increasing availability of firm-level data and growing capacity of computers to process them[1]. Econometrically, the most common approach to estimate SFPF is to specify a deterministic, parametric production function common to all economic units and a stochastic component that consists of a two-part error term[2]. The first component of this error term is a symmetric disturbance that represents statistical noise and follows a normal distribution. The second part is a one-sided error that represents technical inefficiency and which typically follows a half-normal or truncated normal distribution[3]. This approach implicitly assumes that the explanatory variables or factors of production are measured without any errors. It is difficult to justify that the data collected on input use represents true measurements of their theoretical counterparts. For example, a major source of measurement error in measuring flow of services from capital is the lack of information on the vintage nature of capital stock. Additionally, the labor usage figures reported by firms can also be mismeasured due to the non-availability of information on the skill levels of the labor force being used. Consequently, the bias introduced by measurement errors can potentially be rather severe and this paper devises a methodology to investigate the severity of this bias.

---

[1] Both economists and policy makers have made use of this trend as the notion of frontier is consistent with the theory of optimization in addition to identifying factors that can explain relative efficiencies of economic units. A partial list of studies that use the SFPF approach for efficiency measurement related issues is: Kumbhakar (1987,1988), Battese and Coelli (1992), Bauer (1992), Kumbhakar and Hjalmarsson (1995) and Dhawan and Gerdes (1997).

[2] Bauer (1990) and Greene (1993) contain detailed surveys of different econometric techniques for estimating SFPF and technical efficiency.

[3] An estimate of the technical inefficiency is then obtained from the mean or mode of the conditional distribution of the one-sided error term given the composed error term (Jondrow et al 1982).

We utilize and extend Fuller's (1987) *reliability ratio* concept to investigate the measurement bias due to errors-in-variables in cross-sectional SFPF framework[4]. Briefly, the concept is as follows. If x is the true (but unobserved) value of a variable, u is the measurement error and $z = x + u$ is the observed measurement, then the reliability ratio may be defined as the ratio of variance of x to z. This means that a variable with no measurement error ($z = x$) has a reliability ratio of one. Thus, lower the value of reliability ratio, higher is the degree of measurement errors in the observed data. Next, given the reliability ratio consistent estimates can be derived for the parameters of the model under consideration. As the reliability ratio may be unknown most of the time in a practical setting, the best alternative is to derive a *range* of estimates given plausible values for the reliability ratio[5]. One can then examine the sensitivity of the estimates to this range in the reliability ratio. For example, while estimating the SFPF model an issue is how sensitive is the estimated firm-specific technical efficiency to the degree of measurement errors in the input data.

The outline of the paper is as follows. In section 2, we set up the cross-sectional SFPF with no measurement errors as developed in Aigner, Lovell and Schmidt (1977). We then define the setup of a SFPF model when inputs (capital and labor) are measured with errors and illustrate how to estimate it. Section 3 presents a Monte-Carlo simulation study that illustrates the superiority of the method developed to deal with measurement errors in section 2. An empirical

[4] The concept of reliability ratio is not new and Fuller (1987) contains a detailed exposition of how to use this concept to derive maximum-likelihood estimate of a multiple regression equation when the reliability ratio is known.

[5] The word plausible is used in the statement as not all degrees of measurement errors or reliability ratios can be supported by the data in a practical setting. Thus, in this paper we also derive the expression for the upper bound for the variance of the measurement error (lower bound for the reliability ratio) that can be supported by a given empirical data set.

example based on U.S. firm-level data is given in section 4. Section 5 concludes with a summary of findings and directions for future research.

2. SFPF in Cross-Section

2.1 Basic set-up

Starting with Aigner et al. (1977), the cross sectional SFPF may be written as:

$$y_i = f(x_i\,;\,b) + e_i - x_i\,. \tag{1}$$

where, for each firm, y_i is output in logs and x_i is the (actual) $k \times 1$ vector of inputs in log terms. e_i is an IID random variable which represents the statistical noise to the production and b is a $k \times 1$ vector of unknown parameters. Aigner et al. assume that e_i ~ N(0, s_e^2) so that the maximum output firm i can produce using x_i is then $f(x_i\,;\,b) + e_i$. They also focused on a linear model, i.e. $f(x_i\,;\,b) = a + x_i'b$ that is consistent with the Cobb-Douglas type production function assumption. Technical inefficiency is then introduced as a positive random variable x_i. The most common assumption made in the literature is that x_i follows a truncated normal with mean zero (the positive half), i.e. x_i ~ iid$N^+(0,\ s_x^2)$ [6,7]. Defining $e_i = e_i - x_i$ as the compound residual, equation (1) can be written as:

$$y_i = a + x_i'b + e_i. \tag{2}$$

The density of e_i is well known (see Weinstein (1964)) and given below:

$$f_e(e_i) = \frac{1}{\sqrt{2ps^2}}\,\text{erfc}\left[\frac{l e_i}{\sqrt{2s^2}}\right]\exp\left[-\frac{e_i^2}{2s^2}\right] \tag{3}$$

where $l = \dfrac{s_x}{s_e}$, $s^2 = s_e^2 + s_x^2$, $s_x^2 = \dfrac{l^2 s^2}{1+l^2}$, $s_e^2 = \dfrac{s^2}{1+l^2}$

---

[6] Other common assumptions are exponential (Meeusen and van den Brock (1977) and also in Aigner, Lovell and Schmidt (1977)), gamma (Beckers and Hammond (1987) and truncated normal with non zero mean (Stevenson (1980)). See Greene (1997a) for a detailed discussion regarding the merits and shortcomings of these different distributional assumptions.

[7] Where the variance of x_i is equal to $(1\text{-}2/\pi)s_x^2$.

and erfc(x) is defined as 1-erf(x) where the error function erf(x) is defined as:

$$\mathrm{erf}(x) = \frac{2}{\sqrt{p}} \int_0^x e^{-t^2}\, dt .$$

l is directly related to the skewness of e while s^2 is directly related to the variance of e[8]. The empirical distribution of e is key to estimation of the SFPF model in equation (2) as the variance of the technical efficiency (s_x^2) is derived from the estimated skewness of e.

With no measurement errors, estimation of parameters (a, b', l and s^2) in equation (2) is straightforward. OLS will give us a consistent but inefficient estimate of b and an inconsistent estimate of a. OLS, however, would not allow us to estimate the variance of the technical efficiency s_x^2 [9]. An alternative is to use maximum likelihood method. The log-likelihood function following Aigner et al. (1977) can be written as :

$$\begin{aligned} l(a, b, l, s^2) &= \sum_{i=1}^{n} \ln f_e(y_i - a - x_i b) \\ &= \sum_{i=1}^{n} \ln\left\{ \mathrm{erfc}\left[\frac{l\,(y_i - a - x_i'\, b)}{\sqrt{2s^2}} \right] \right\} - \sum_{i=1}^{n} \frac{(y_i - a - x_i'\, b)^2}{2s^2} - \frac{n}{2} \ln(2ps^2) \end{aligned} \qquad (4)$$

where the above likelihood can easily be modified for a different specification of the production function $f(x_i\, ;\, b)$ as well as for different distributional assumptions on x_i.

2.2 Technical efficiency

The production function implicit in equation (2) when written in level terms is:

$$Y_i = \mathrm{e}^{a} \left[\prod_{j=1}^{k} X_{ij}^{b_j} \right] \mathrm{e}^{e_i} \mathrm{e}^{-x_i}$$

---

[8] Note that s^2 is not exactly the variance of e which is given by the following expression:

$$\mathrm{Var}(e) = \mathrm{Var}(e) + \mathrm{Var}(x) = s_e^2 + \left(1 - \frac{2}{p}\right) s_x^2$$

[9] But OLS does a good job of testing whether the error term is skewed or not. Another alternative would be to use Corrected OLS which gives consistent but inefficient estimates of the regression parameters (see Greene (1997)).

where the output Y_i and input X_{ij} are in levels. Thus, the technical efficiency of firm i can be defined as j_i = exp($-x_i$)[10] and it may be interpreted as the percentage of maximum possible output achieved when the residual e_i is zero. This involves the technical inefficiency effect x_i , which is an unobservable random variable. Even if the true value of the parameter vector β was known, only the difference, $\mathrm{e}_i = e_i - x_i$, could be observed. Thus, if e_i is the compound residual for firm i , the predictor of firm-specific technical efficiency is the random variable ($j_i \mid e_i$). The expected value of ($j_i \mid e_i$), with e_i replaced by the residual $\hat{e}_i$, is our best prediction of technical efficiency for firm i given the residual $\hat{e}_i = y_i - \hat{a} - x_i{}'\hat{b}$. Jondrow et al. (1982) compute E($x_i \mid \hat{e}_i$) as the predicted technical *in*efficiency. This is based on the approximation that $x_i = -\ln f_i \approx 1 - f_i$. We prefer to avoid this approximation and calculate directly the conditional expected value of f_i as suggested by Battese and Coelli (1988). This after some tedious algebra is[11]:

$$\mathrm{E}[\exp(-x_i) \mid \hat{e}_i] = \exp\left\{\frac{\left(2\hat{e}_i + s_e^2\right)s_x^2}{2s^2}\right\}\left\{1-\Phi\left[\frac{\left(\hat{e}_i + s_e^2\right)l}{s}\right]\right\}\left\{1-\Phi\left[\frac{\hat{e}_i l}{s}\right]\right\}^{-1} \tag{5}$$

The average or mean technical efficiency of firms is straightforward to derive as:

$$\mathrm{E}[\exp(-x_i)] = \mathrm{E}(j_i) = 2\exp(s_x^2/2)[1-\Phi(s_x)] \tag{6}$$

2.3 SFPF with errors-in-variables

Let x_i denote the actual (unobserved) $k \times 1$ vector of inputs in log terms for firm i. We also observe an equal number of variables z_i which are related to x_i in the following manner:

$$z_i = x_i + u_i \tag{7}$$

---

[10] Since $x_i \geq 0$, the technical efficiency will always be between zero and one. A firm is defined as fully efficient or located on the frontier if $x_i = 0$ in which case the technical efficiency measure is equal to one.

[11] Note that this expression converges to exp($\hat{e}_i$) as $s_e^2 \to 0$ which is what we should expect since $\hat{e}_i$ in this case estimates $-x_i$.

where u_i denotes the $k \times 1$ vector of measurement errors (also in logs) and $E(u_i \mid x_i) = 0$. This is the error model approach[12], which implies that the measurement errors are multiplicative in nature[13]. In addition this model is non-calibrated in the sense that z_i is not specified to be related to x_i in a systematic manner [14]. The non-calibrated model is more suitable as we have no reason, a priori, to believe that there is a systematic bias in the measurements of inputs. Given the production function specification in equation (2), and assuming normal distributions for the measurement errors and the unobserved explanatory variables, we get the following model specification:

$$y_i = a + x_i'b + e_i - x_i$$

$$z_i = x_i + u_i, \text{ where } E(u_i \mid x_i) = 0, \qquad (8)$$

$$e_i \sim N(0, s_e^2), \; x_i \sim \text{iid}N^+(0, s_x^2), \; x_i \sim N(m, \Sigma_x) \text{ and } u_i \sim N(0, \Sigma_u).$$

This specification implies that x and u are independent and that $z_i \sim N(m, \Sigma_z)$ with $\Sigma_z = \Sigma_x + \Sigma_u$.[15] Finally, we assume that the true factors (x_i), the measurement errors (u_i), the stochastic frontier error term (e_i) and the technical inefficiency (x_i) are all independent random variables for all i.

---

[12] The other approach to modeling errors is the Berkson model where $x_i = z_i + v_i$, and v_i are the measurement errors. Here, $E(v_i \mid z_i) = 0$ which implies that the expectation conditional on observed z_i is zero as opposed to the error models where the expectation conditional on actual x_i is zero. In addition, in the error model, the observed value is correlated with the measurement error while the actual value is correlated with the measurement error in the Berkson model. Which approach to use depends on what one believes is independent of the measurement errors: the true values of inputs or the observed values. In our view it is natural to assume that the true values are independent of the measurement errors and thus we follow the error model approach.

[13] They are multiplicative in the sense that that $\exp(z_i) = \exp(x_i) \cdot \exp(u_i)$ where $\exp(z_i)$ is the actual observations on z (not in logs) and similarly for x and u.

[14] An error model would be calibrated if it was specified that $z_i = g_0 + g_1 x_i + u_i$.

[15] Pal, Neogi and Ghosh (1998) have analyzed a similar setup with nonstochastic explanatory variables (both observed and unobserved z and x). Although this approach has some advantages, in the sense that the the estimated coefficients do not depend on the distributional assumption made for x and z and the likelihood function is easier to derive, it is an unconventional assumption in the errors-in-variable literature (see for example Fuller (1987) and Greene(1997b)). It is hard to motivate the errors-in-variables assumption that $z_i = x_i + u_i$ assuming that the measurement errors are stochastic without assuming that x and z are stochastic too.

At this point it is clear that the model (8) is unidentifiable as we only observe z_i while x_i and u_i are unobserved. This implies that Σ_u and Σ_x cannot be identified separately which means that ones needs additional information such as instrumental variables. As appropriate instruments for labor and capital are hard to justify, one can then try to identify the "consistent bounds" for parameter b as advocated by Klepper and Leamer (1984). In Klepper and Leamer approach ones identifies all *b's* that imply positive estimates for the variance of $e_i = e_i - x_i$ and positive semi-definite estimates of Σ_x and Σ_u.[16] For the specification in equation (8) we cannot use this approach as OLS does not provide estimates for s_x^2 which is needed to calculate the conditional as well as unconditional technical efficiency estimates. Instead, we will follow (and extend) Fuller's *reliability ratio approach* which, although computationally more expensive, allows us to use maximum-likelihood method to consistently estimate all the parameters of the model as well as the technical efficiencies.

The relevant reliability ratios are defined as following:

$$p_i = \mathrm{Var}(x_i)/\mathrm{Var}(z_i) \quad i = 1,\dots,k \tag{9a}$$

$$p_{ij} = \mathrm{Cov}(x_i, x_j)/\mathrm{Cov}(z_i, z_j) = \mathrm{Cov}(x_i, x_j)/[\mathrm{Cov}(x_i, x_j) + \mathrm{Cov}(u_i, u_j)]$$
$$i = 1,\dots,k \text{ and } j = 1,\dots,k,\ i \neq k \tag{9b}$$

where p_i is the (traditional) reliability ratio associated with variable i, $0 \leq p_i \leq 1$ and p_i is equal to one if there are no measurement errors for variable i. Typically, one specifies the covariance matrix of the measurement errors u to be diagonal (see for example Klepper and Leamer (1984). If this is the case, the reliability ratio p_i's are all we need for identifying the model. However, we would like to allow for covariances between the different measurement error, u_i's, which then

[16] The consistent bounds are found by running $k + 1$ regressions and if all these regressions are in the same orthant, then the set of maximum likelihood estimates will be the convex hull of these estimates.

warrants the introduction of p_{ij}. p_{ij} is the ratio of the true (unobserved) covariance between two variables to the observed covariance and will be called the *covariance reliability ratio.* Note that if one assumes that the measurement errors of different variables are uncorrelated, then all the p_{ij}'s are equal to one[17,18]. Since x_i and u_i are independent, it follows that:

$$\mathrm{Var}(u_i) = (1 - p_i)\mathrm{Var}(z_i) \quad i = 1,\ldots,k \tag{10a}$$

$$\mathrm{Cov}(u_i, u_j) = (1 - p_{ij})\mathrm{Cov}(z_i, z_j) \quad i = 1,\ldots,k \text{ and } j = 1,\ldots,k,\ i \neq k. \tag{10b}$$

It is possible to collect all the reliability ratios into one matrix Π where:

$$\Pi = \begin{bmatrix} p_1 & \cdots & p_{1k} \\ \vdots & \ddots & \vdots \\ p_{k1} & \cdots & p_k \end{bmatrix}$$

With this notation, we may write, $\Sigma_x = \Pi.*\Sigma_z$ and $\Sigma_u = (\mathbf{1} - \Pi).*\Sigma_z$ where the notation " .* " means element by element multiplication and where **1** is a $k \times k$ matrix of ones. Because z_i is observed, it is possible to estimate the variance matrix Σ_z. In order to identify the model we have specified so far we must also know either one of Π, Σ_x or Σ_u[19].

2.4 Maximum likelihood estimation

Our goal is to estimate the parameters a, b, s_e^2, s_x^2, or equivalently, a, b, s^2 and l, in model (8) given different values for reliability ratio Π. However, the joint distribution of the observations on y_i and z_i will involve all the parameters. Given the matrix of reliability ratios P these parameters are given by $w' = (m', \mathrm{vech}(\Sigma_z)', a, b', l, s^2)$. We will use a two-step procedure also known as the method of limited information maximum likelihood (LIML) to maximize the

---

[17] Also, unless the measurement errors are negatively correlated, the p_{ij}'s will be less than one.
[18] If the variables x_i and x_j are uncorrelated, we let p_{ij} be equal to zero, irrespective of the fact whether the measurement errors are correlated or uncorreleated.
[19] Now, even if the reliability ratio is unknown for the particular variables under study one can then investigate the sensitivity of the estimates to the changes in the reliability ratio.

likelihood function[20]. In the first step of this procedure we use the sample moments of z to estimate m and Σ_z. Substituting these estimates into the full likelihood will, under weak assumptions, lead to consistent estimates of $(a,\ b',\ l,\ s^2)$. This second stage likelihood function is much easier to maximize than the full likelihood, but the estimated covariance matrix of $(a,\ b',\ l,\ s^2)$ at this stage will no longer be consistent. However, by using the results of Murphy and Topel (1985), we will show how the estimated covariance matrix of $(a,\ b',\ l,\ s^2)$ can be adjusted to yield consistent results.

Now, if we assume that x, and therefore z, is weakly exogenous with respect to $(a,\ b',\ l,\ s^2)$ we can then write the log-likelihood of the parameters as:

$$\ell(w) = \sum_{i=1}^{n} \ln f(y_i, z_i | w) = \sum_{i=1}^{n} \ln f_1(z_i | w_1) + \sum_{i=1}^{n} \ln f_2(y_i | z_i, w_1, w_2) \tag{11}$$

where $w_1' = (m', \text{vech}(\Sigma_z)')$ and $w_2' = (a,\ b',\ l,\ s^2)$.

In LIML procedure we first maximize the first part of the likelihood expression in (11):

$$\ell_1(w_1) = \sum_{i=1}^{n} \ln f_1(z_i | w_1) \tag{12}$$

over w_1. This will give us an estimate $\hat{w}_1(z)$, which is then used to maximize the remaining

portion of the likelihood:

$$\ell_2(\hat{w}_1, w_2) = \sum_{i=1}^{n} \ln f_2(y_i | z_i, \hat{w}_1(z), w_2) \tag{13}$$

over w_2. This gives us an estimate $\hat{w}_2(y,z)$. The exact expression for the first-step likelihood is:

$$\ell_1 = \text{constant} - \tfrac{n}{2}\ln|\Sigma_z| - \tfrac{n}{2}\text{tr}S_z\Sigma_z^{-1} - \tfrac{n}{2}(\bar{z}-m)'\Sigma_z^{-1}(\bar{z}-m) \tag{14}$$

[20] Estimating these parameters using full information maximum likelihood (FIML) could not be done as the likelihood function was too flat, mostly due to the elements in $\mathbf{\Sigma}_z$, to allow us to find the maximum.

where $S_z = \frac{1}{n}\sum_i (z_i - \bar{z})(z_i - \bar{z})'$. It is well known that $\bar{z}$ and S_z are the ML estimates of m and Σ_z, and that they are independently distributed as $N(m, \frac{1}{n}\Sigma_z)$ and $\left(\frac{n}{n-1}\right)\text{Wishart}\left(\frac{1}{n}\Sigma_z, n-1\right)$. The sample moments thus consistently estimate the population moments.

These results enable us to calculate an estimate of V_1, the asymptotic covariance matrix of $\hat{w}_1' = (\hat{m}', \text{vech}(\hat{\Sigma}_z)') = (\bar{z}', \text{vech}(S_z)')$. Although we are not directly interested in an estimate of V_1 but is needed later to obtain consistent estimate V_2 which is the asymptotic covariance matrix of $\hat{w}_2(y,z)$. The asymptotic covariance matrix V_1 is of dimension $\frac{1}{2}k(k+3)$ and can be written as following:

$$V_1 = \begin{pmatrix} V_{1mm} & V_{1ms} \\ V_{1sm} & V_{1ss} \end{pmatrix}.$$

If we define $S_i = (z_i - \bar{z})(z_i - \bar{z})$ then we can write following Edgerton and Jochumzen (1999) the elements of V_1 as[21]:

$$\hat{V}_{1mm} = \tfrac{1}{n} S_z$$
$$\hat{V}_{1ss} = \tfrac{1}{n}\left\{\left(\tfrac{1}{n}\sum_i \text{vech}(S_i)(\text{vech}(S_i))'\right) - \text{vech}(S_z)(\text{vech}(S_z))'\right\}$$
$$\hat{V}_{1ms} = \hat{V}_{1sm} = \tfrac{1}{n^2}\left\{\sum_i (z_i - \bar{z})(\text{vech}(S_i))'\right\}$$

To calculate the second-step likelihood we need to find the conditional distribution of y given z. Let v be the random variable $x'\boldsymbol{b}$. Then by considering the joint density of y and v we have:

$$f_{y|z}(y|z) = \int_V f_{y,v|z}(y,v|z)dv = \int_V f_{y|v,z}(y|v,z)\, f_{v|z}(v|z)dv \qquad (15)$$

Because we are conditioning on z and $x'\boldsymbol{b}$, we have:

$$\begin{aligned} f_{y|v,z}(y|v,z) &= f_{e|v,z}(y - a - v|v,z) \\ &= f_{e|v}(y - a - v|v) = f_e(y - a - v) \end{aligned} \qquad (16)$$

The last equality follows as e is independent of z and $x'\boldsymbol{b}$. By substituting (16) into (15) we get:

$$f_{y|z}(y|z) = \int_v f_{v|z}(v|z)\, f_e(y - a - v)dv \qquad (17)$$

[21] Note that if z_i is normal, the estimate of V_{1sm} will be zero.

where the integral is a single one over all possible values of $x'b$. Since x and z are normal, $v \mid z$ will also be normal. Straightforward application of the results for conditional density of a multivariate normal gives us the expected value and variance of $x'b \mid z$ [22]

$$\mathrm{E}(v \mid z) = m'b + (z - m)'\Sigma_z^{-1}\Pi.*\Sigma_z b \tag{18a}$$

and

$$\mathrm{Var}(v \mid z) = b\,'(\Pi.*\Sigma_z - \Pi.*\Pi.*\Sigma_z)b = b\,'(\Pi.*\Sigma_z .*\Pi^c)b \tag{18b}$$

where $\Pi^c = 1 - \Pi$.

Thus, the density $f_{v/z}$ can be written as:

$$f_{v|z}(v|z) = \frac{1}{\sqrt{2pb'\left(\Pi.*\Sigma_z.*\Pi^c\right)b}}\exp\left[-\frac{\left[x'b - m'b - (z-m)'\Sigma_z^{-1}\Pi.*\Sigma_z b\right]^2}{2b'\left(\Pi.*\Sigma_z.*\Pi^c\right)b}\right] \tag{19}$$

Therefore, the conditional density of y_i given z_i is:

$$f(y_i|z_i) = \int_v \frac{1}{2p\sqrt{s^2 b'\left(\Pi.*\Sigma_z.*\Pi^c\right)b}}\,\mathrm{erfc}\left[\frac{l(y_i - a - v)}{\sqrt{2s^2}}\right]\times \exp\left[-\frac{(y_i - a - v)^2}{2s^2} - \frac{\left[v - mb - (z_i - m)'\Sigma_z^{-1}\Pi.*\Sigma_z b\right]^2}{2b'\left(\Pi.*\Sigma_z.*\Pi^c\right)b}\right]dv \tag{20}$$

and the second step likelihood function ℓ_2 is given by

$$\ell_2(\hat{w}_1, w_2) = \sum_{i=1}^{n}\ln f_2\left(y_i \middle| z_i, \hat{w}_1(z), w_2\right) =$$

$$\sum_{i=1}^{n}\int_v \frac{1}{2p\sqrt{s^2 b'\left(\Pi.*\hat{\Sigma}_z.*\Pi^c\right)b}}\,\mathrm{erfc}\left[\frac{l(y_i - a - v)}{\sqrt{2s^2}}\right]\times \exp\left[-\frac{(y_i - a - v)^2}{2s^2} - \frac{\left[v - \hat{m}b - (z_i - \hat{m})'\hat{\Sigma}_z^{-1}\Pi.*\hat{\Sigma}_z b\right]^2}{2b'\left(\Pi.*\hat{\Sigma}_z.*\Pi^c\right)b}\right]dv \tag{21}$$

---

[22] Since $\mathrm{E}(x_i'b) = mb$ and $\mathrm{Cov}(x_i'b, z_i) = \Sigma_x b$ we have

$$\begin{pmatrix} x_i b \\ z_i \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} m'b \\ m \end{pmatrix}, \begin{pmatrix} b'\Sigma_x b & b'\Sigma_x \\ \Sigma_x b & \Sigma_z \end{pmatrix}\right]$$

Then using the formula for the conditional distribution of $x_i'b$ given z_i we get $x_i'b \mid z_i \sim \mathrm{N}[m'b + (z_i - m)'\Sigma_z^{-1}\Sigma_x b,\ b'(\Sigma_x - \Sigma_x\Sigma_z^{-1}\Sigma_x)b]$. This using $\Sigma_x = \Pi.*\Sigma_z$ could then be written as $\mathrm{N}[m'b + (z - m)'\Sigma_z^{-1}\Pi.*\Sigma_z b,\ b\,'[\Pi.*\Sigma_z.*(\mathbf{1} - \Pi)]b]$.

The LIML estimates are found by substituting the first-step estimates of m and Σ_z ($\bar{z}$ and S_z) into (21) which is then maximized. It is possible to show that this integral will not simplify unless every reliability ratio in Π is equal to one in which case the density above will converge to the density of e in section 2. However, it is possible to evaluate the likelihood function numerically for any matrix Π[23]. Note that the likelihood function will simplify further if all the reliability ratios are equal to one common value say $\hat{p}$. Conditional on this reliability ratio $\hat{p}$, we have:

$$E(x_i'b \mid z_i) = b'm + \hat{p}\, b'(z_i - m) \tag{22a}$$

$$Var(x_i'b \mid z_i) = \hat{p}(1-\hat{p})b'S_z b \tag{22b}$$

and a simpler form for the conditional density of y given z is:

$$f(y_i|z_i) = \int_v \frac{1}{2p\sqrt{s^2\hat{p}(1-\hat{p})b'\Sigma_z b}}\,\mathrm{erfc}\left[\frac{l(y_i - a - v)}{\sqrt{2s^2}}\right]\times \exp\left[-\frac{(y_i - a - v)^2}{2s^2} - \frac{[v - m'b - \hat{p}(z_i - m)'b]^2}{2\hat{p}(1-\hat{p})b'\Sigma_z b}\right]dv \tag{23}$$

The adjusted asymptotic covariance matrix of the second-step estimates w_2 (V_2^*) has to be calculated as following:

$$V_2^* = V_2 + V_2(CV_1C - RV_1C - CV_1R)V_2 \tag{24}$$

where V_2 is the unadjusted second-step covariance matrix. Also, C and R, following Murphy and Topel (1985) who establish consistency of LIML under the usual regularity conditions, are defined as[24]:

$$C = E\left[\left(\frac{\partial \ell_2}{\partial w_2}\right)\left(\frac{\partial \ell_2}{\partial w_1'}\right)\right],\; R = E\left[\left(\frac{\partial \ell_2}{\partial w_2}\right)\left(\frac{\partial \ell_1}{\partial w_1'}\right)\right].$$

---

[23] We need to evaluate n integrals each time we calculate the value of the likelihood function which is not a problem for a fast computer.

[24] We used numerical derivatives to calculate C and R in our Monte-Carlo analysis.

2.5 Bounds on reliability ratio

Not all Π-matrices are possible as some Π-matrices will give rise to negative estimates of the variance of e_i. The reasoning is as following. Combining equations (2) and (7) one gets:

$$y_i = a + z_i'b + e_i - u_i'b - x_i \tag{25}$$

where the error term is now composed of three parts. Since $u_i b$ is normal, the skewness of the data will determine the relative share of the variance between $e_i - u_i b$ on one hand and x_i on the other in equation (25). Given this division, the data will determine variance of $e_i - u_i b$ ($s^2_{e-ub} = s_e^2 + b'\Sigma_u b$), but not how s^2_{e-ub} is shared between the two terms e_i and $u_i b$. With no measurement errors, all the s^2_{e-ub} variance can be attributed to the variance of the residual e. As the variance of u increases, more of the s^2_{e-ub} variance will be due to the variance in $u_i b$. In the extreme case <u>all</u> the variance of $e_i - u_i b$ is due to measurement errors. In this case $s_e^2 = 0$ and $s^2_{e-ub} = b'\Sigma_u b$. Attempting to increase the variance above this limit will result in negative estimate of s_e^2. This in turn determines the lower limit for the reliability ratios. In practice, the estimated variance of e decreases when we decrease the reliability ratios and the lower limit of the reliability ratios will be found where the estimated variance of e goes to zero.

If we define b_Π as the estimated value of b given the reliability ratio matrix Π, the restriction that the estimate of $s_e^2 \geq 0$ is equivalent to $b_\Pi\,'\Sigma_u b_\Pi = b_\Pi\,'(1-\Pi.*\Sigma_z)b_\Pi \leq s^2_{e-ub}$ which is the restriction for determining feasible values of Π. Additional bounds on the reliability ratios may be found if some other simplifying assumptions are made. This will be discussed in detail in when doing the simulation exercise in section 3.

2.6 Technical efficiency with errors-in-variables

Our aim is now to estimate the technical efficiency of the firm and the mean for the sample for different values of the reliability ratio Π. The expression for the mean technical efficiency, E[exp($-x_i$)], is the same as in equation (6) even with measurement errors since the distribution of x_i is unaffected. However, the expression for the firm-specific technical efficiency requires a slight modification. With measurement errors, the compound residual will be given by $e_{i*} = e_i - u_i'b - x_i$ (see equation (25)) instead of $e_i = e_i - x_i$ as in the case of no measurement errors. Since $u_i'b$ is normal, the expression we derived for E[exp($-x_i$) | $\hat{e}_i$] in equation (5) will be valid if we replace e_i by $e_i - u_i'b$ and redefine s^2 and l as following:

$$s_*^2 = s_e^2 + b'\Sigma_u b + s_x^2 \text{ and } l_* = \frac{s_x}{\sqrt{s_e^2 + b'\Sigma_u b}} \tag{26}$$

Consequently, the expression for conditional technical efficiency under errors will then be:

$$\text{E}[\exp(-x_i)|\hat{e}_{i*}] = \exp\left\{\frac{\left(2\hat{e}_{i*} + s_e^2 + b'\Sigma_u b\right)s_x^2}{2s_*^2}\right\}\left\{1-\Phi\left[\frac{\left(\hat{e}_{i*} + s_e^2 + b'\Sigma_u b\right)l_*}{s_*}\right]\right\}\left\{1-\Phi\left[\frac{\hat{e}_{i*}\,l_*}{s_*}\right]\right\}^{-1} \tag{27}$$

3. Simulation Study

3.1 Simulation set-up

This section compares the new estimator for the cross-sectional SFPF developed in the previous section (henceforth called the EIV estimator) with the traditional ML estimator on simulated data. The aim is to investigate the bias introduced by measurement errors in estimating the production function parameters and the resulting technical efficiency estimates. The model that we choose to simulate is a Cobb-Douglas production function with two inputs, capital and labor. The choice of only two inputs was motivated by the desire to be as similar to our empirical example presented in section 4 where the data allows for identification of only two broadly defined category of inputs: total capital and total labor. In addition, the basic parameters for simulation are chosen so as to closely mimic the actual data analyzed in section 4.

The starting point of the simulation is the following model specification:

$$\ln(Y_i) = a + b_K \ln(K_i) + b_L \ln(L_i) + e_i - x_i \quad (28a)$$

$$\text{where } \ln(\underline{K}_i) = \ln(K_i) + \ln(U_{Ki}) \text{ and } \ln(\underline{L}_i) = \ln(L_i) + \ln(U_{Li}) \quad (28b, 28c)$$

where K_i and L_i are actual but unobserved amount of capital and labor of firm i and $\underline{K}_i$ and $\underline{L}_i$ are the measured counterparts. U, e and x are as defined in section 2 and let s_K^2 denote the variance of $\ln(\underline{K}_i)$, s_L^2 the variance of $\ln(\underline{L}_i)$ and s_{KL} the covariance between $\ln(\underline{K}_i)$ and $\ln(\underline{L}_i)$.

Now, a slight modification of the model in (28) by writing it in per-capita terms is preferred. This is achieved by subtracting $\ln(L_i)$ from both sides of (28a) and subtracting (28c) from (28b). There are three advantages of doing this. First, it is easier to find the maximum of the likelihood function when regressing $\ln(Y/L)$ on $\ln(K/L)$ and $\ln(L)$ instead of regressing $\ln(Y)$ on $\ln(K)$ and $\ln(L)$. Second, the parameter of $\ln(L)$ will directly estimate the degree of departure from the constant returns to scale. Third, the per-capita specification allows us to find bounds on the *feasible* reliability ratios as we discuss later in the next sub-section. Thus, after the transformation the model in equations (28a,b,c) can be written as:

$$\ln(Y_i / L_i) = a + b_K \ln(K_i / L_i) + (b_L + b_K - 1)\ln(L_i) + e_i - x_i. \quad (29a)$$

$$\ln(\underline{K}_i / \underline{L}_i) = \ln(K_i / L_i) + \ln(U_{Ki} / U_{Li}), \ \ln(\underline{L}_i) = \ln(L_i) + \ln(U_{Li}) \quad (29b, 29c)$$

or equivalently as:

$$y_i = a + x_i g + e_i - x_i. \quad (30a)$$

$$z_i = x_i + u_i. \quad (30b)$$

where $y_i = \ln(Y_i / L_i)$, $x_i = [\ln(K_i / L_i), \ln(L_i)]$, $z_i = [\ln(\underline{K}_i / \underline{L}_i), \ln(\underline{L}_i)]$, $u_i = [\ln(U_{Ki} / U_{Li}), \ln(U_{Li})]$ and $g = [b_K, (b_L + b_K - 1)]$[25]. We simulate x_1, x_2, u_1 and u_2 from normal distributions such that x_i

[25] Note that x_i and u_i are independent since $\ln(K_i / L_i)$ and $\ln(U_{Ki} / U_{Li})$ are independent.

~ N(0, $2\widehat{p}$), u_i ~ N(0, $2(1-\widehat{p})$) with $\widehat{p}$ being the common reliability ratio. Then by adding x to u, we get z with the desired properties. Once x and u are simulated we then simulate e_i from a N(0, s_e^2) and x_i from a truncated $N^+(0,\ s_x^2)$ with Var(e_i) = 0.2 and Var(x_i) = 0.8 which implies that $s^2 = 1$ and $l = 2$. Finally, we create y by selecting $a = 1.7$, $g_1 = 0.6$ and $g_2 = 0.1$. This implies increasing returns to scale with coefficients $b_K = 0.6$ and $b_L = 0.5$. This is the model structure that we will use for our simulation study.

3.2 Restrictions on the reliability ratios

In contrast to the generalized bounds discussed in section 2.5, we will derive simplified bounds when simulating independent series for x_1 and x_2. The independence assumption implies that the covariance between *actual* $\ln(K_i / L_i)$ and $\ln(L_i)$ is zero. Obviously, we do not know what this covariance is in reality but we can estimate the covariance between *observed* $\ln(\underline{K}_i / \underline{L}_i)$ and $\ln(\underline{L}_i)$. In the actual data that we have examined in section 4, this covariance is almost zero[26]. Since $\mathrm{Cov}(z_1, z_2) = \mathrm{Cov}(x_1, x_2) + \mathrm{Cov}(u_1, u_2)$, and unless there is a reason to believe that the covariance between the measurement errors of $\ln(\underline{K}_i / \underline{L}_i)$ and $\ln(\underline{L}_i)$ is far from zero, it seems reasonable to assume that $\mathrm{Cov}(x_1, x_2)$ is close to zero as well. The implications of setting these covariances to zero are as following:

1. $s_{KL} = s_L^2$. [27]
2. $p_{KL} = p_L$ where again p_{KL} is the "covariance reliability ratio" $\mathrm{cov}[\ln(K_i), \ln(L_i)] / \mathrm{cov}[\ln(\underline{K}_i), \ln(\underline{L}_i)]$[28]. This simplification is very useful since it decreases the number of unknown parameters from three to two when we do the simulations.

---

[26] The actual correlation between $\ln(\underline{K}_i / \underline{L}_i)$ and $\ln(\underline{L}_i)$ in the empirical data of section 4 was -0.03.

[27] Since $\mathrm{Cov}(z_1, z_2) = \mathrm{Cov}[\ \ln(\underline{K}_i/\underline{L}_i), \ln(\underline{L}_i)\ ] = \mathrm{Cov}[\ \ln(\underline{K}_i), \ln(\underline{L}_i)\ ] - \mathrm{Var}[\ \ln(\underline{L}_i)\ ] = s_{KL} - s_L^2 = 0$, it follows.

[28] Since $\mathrm{Cov}(x_1, x_2) = \mathrm{Cov}[\ \ln(K_i/L_i), \ln(L_i)\ ] = \mathrm{Cov}[\ \ln(K_i), \ln(L_i)\ ] - \mathrm{Var}[\ \ln(L_i)\ ] = p_{KL}\cdot\mathrm{Cov}[\ \ln(\underline{K}_i), \ln(\underline{L}_i)\ ] - p_L\cdot\mathrm{Var}[\ \ln(\underline{L}_i)\ ]=0$ and $\mathrm{Cov}[\ \ln(\underline{K}_i), \ln(\underline{L}_i)\ ] = \mathrm{Var}[\ \ln(\underline{L}_i)\ ]$ from (1), this follows.

3. $\text{Var}(z_1) = \text{Var}[\ \ln(\underline{K}_i/\underline{L}_i)\ ] = s_K^2 - s_L^2$, $\text{Var}(z_2) = s_L^2$.

4. $\text{Var}(x_1) = \text{Var}[\ \ln(K_i/L_i)\ ] = p_K s_K^2 - p_L s_L^2$, $\text{Var}(x_2) = p_L s_L^2$.

5. $\text{Var}(u_1) = \text{Var}[\ \ln(U_{Ki}/U_{Li})\ ] = (1-p_K) s_K^2 - (1-p_K) s_L^2$, $\text{Var}(u_2) = (1-p_K) s_L^2$.

6. The reliability ratio of the variable $\ln(\underline{K}_i/\underline{L}_i)$ expressed in terms of the reliability ratios of Capital and Labor is $\text{Var}[\ln(K_i/L_i)\ ] / \text{Var}[\ \ln(\underline{K}_i/\underline{L}_i)\ ] = (p_K s_K^2 - p_L s_L^2)/(s_K^2 - s_L^2)$ by (4) and (5). Note that if the reliability ratio of Capital and Labor are equal (to say $\hat{p}$), then the reliability ratio of $\ln(\underline{K}_i/\underline{L}_i)$ is $\hat{p}$ itself.

By noting that the reliability ratio of $\ln(\underline{K}_i/\underline{L}_i)$ must itself be between zero and one we can find feasible bounds on the reliability ratios for capital and labor. These bounds are[29]:

$$p_L \frac{s_L^2}{s_K^2} \leq p_K \leq p_L \frac{s_L^2}{s_K^2} + \frac{s_K^2 - s_L^2}{s_K^2} \tag{31}$$

This expression evaluates to $0.845 p_L \leq p_K \leq 0.845 p_L + 0.155$ using $s_K^2 = 8.15$ and $s_L^2 = 6.17$ as observed in the empirical data analyzed in section 4. These are powerful restrictions that together with the condition that the estimate of s_e^2 be positive will limit the set of possible reliability ratios that we can consider during the simulations. The table below shows the feasible values for p_K given the values of p_L..

p_L	1.00	0.98	0.96	0.94	0.92	0.90	0.88	0.86	0.84	0.82	0.80	0.78	0.76	0.74
Min. p_K	0.84	0.83	0.81	0.79	0.78	0.76	0.74	0.73	0.71	0.69	0.66	0.66	0.64	0.62
Max. p_K	1.00	0.97	0.97	0.95	0.93	0.92	0.90	0.88	0.86	0.85	0.83	0.81	0.80	0.78

For each simulation round, we start by setting p_L and p_K in such a way that they are within the bounds defined in equation (31). In practical terms this implies setting $p_L = p_K$ as the bounds

[29] The bounds are calculated by equating the expression $(p_K s_K^2 - p_L s_L^2)/(s_K^2 - s_L^2)$ equal to 0 and 1, respectively.

expression do allow p_L to be equal to p_K[30]. Thus, what matters during the simulations is whether p_L and p_K are large (close to one) or small.

3.3 Simulation results: parameter estimates

Each simulation round consisted of 500 observations to estimate the parameters and this was repeated 100 times. Table 1 presents the averages and the standard deviations of the estimated parameters for the MLE method and the EIV estimator under different levels of reliability ratios. In the table estimates of s_x^2 and s_e^2 are derived from the estimates of s^2 and l using expressions defined in equation (3). The most striking result of the simulation study is the severe downward bias in the MLE estimate of g_1 and g_2 as the common reliability ratio falls. This implies that (where $g_1 = b_K$ and $g_2 = b_L + b_K - 1$) we *underestimate the elasticity of capital while we overestimate that of labor* when there are measurement errors. For example, in the simulated data the elasticity of capital was 0.6 while that of the labor was 0.5. With 80% reliability in the data, the capital elasticity is underestimated by 20%, and for 70% reliability ratio the estimates are completely reversed: 0.4175 for capital and 0.6499 for labor. Thus, the biases are quite severe and clearly show the need for a procedure that consistently estimates the elasticity parameters under even reasonable degree of measurement errors.

Table 1 results also imply that MLE tends to underestimate the return to scale parameter g_2. Therefore, if one wants to test for increasing returns, the MLE does a poor job whereas the EIV estimator will pick out true increasing returns even for a 70% reliability ratio. Table 1 also shows that the MLE based l estimate is biased downward and s^2 is biased upwards. The

[30] If we assume that the covariance between x_1 and x_2 is less than 0.05 in absolute terms, then $-0.00727 + 1.0058p_L \leq p_K \leq 0.00727 + 1.0058p_L$ is the equation that defines the bounds. Then if $p_L = 1$, p_{KL} must be between 0.9985 and 1.013 and if $p_L = 0.9$, p_{KL} must be between 0.8979 and 0.9124. Thus, setting $p_{KL} = p_L$ seems to be the most reasonable choice.

combined effect of these two on the estimate for the variance of technical efficiency (s_x^2) is that it seems to be estimated consistently whereas the variance of the residual (s_e^2) is biased upwards. This is not surprising as the measurement errors being normally distributed, will be captured in the s_e^2 term, thus biasing it upwards, leaving the s_x^2 estimate to be unaffected. Thus, there will be an upward bias in estimated s^2 and a downward bias in estimated l under MLE. From table 1 it is clear that both s^2 and l are consistently estimated by the EIV estimator. To sum, even with extreme measurement errors, the EIV estimator succeeds in estimating the elasticity parameters, returns to scale and the relevant variance estimates consistently.

3.4 Simulation results: technical efficiency

Table 2, column 2 presents the mean technical efficiency estimates for different reliability ratios where the true value using equation (6) and $s_x^2 = 0.8$ is 0.5536. Column 3 and 4 respectively have estimates of the expected value of the technical efficiency based on MLE and EIV estimates from table 1. Here the MLE estimator does as well as the EIV estimator when it comes to estimating the mean value of the technical efficiency. This happens because both MLE and EIV produce estimates of s_x^2 that are identical to the third decimal, and that is the only parameter that determines the average technical efficiency. Hence, if you are only interested in mean technical efficiency of the sample, you may just as well use the traditional MLE estimator, even if the data suffers from measurement errors , as long as these are normally distributed.

Next, we analyze the technical efficiency of firm i once we have estimated the residual for that firm. For each simulation round we compare the true technical efficiencies to the estimated technical efficiencies calculated using the MLE and EIV techniques. This comparison was done by calculating the average absolute deviation between the true technical efficiency and the MLE and EIV estimates of it. The means and standard deviations of the absolute deviations from the 100 simulation rounds are presented in table 3. It is clear from the table that EIV is much more successful in estimating the firm-specific technical efficiency. The average absolute deviation between the true value and the MLE estimate rises as the severity of the measurement errors increases, unlike for the EIV estimator where the absolute deviation stays about the same. In the last column of table 3 we report another test that proves the superiority of EIV over MLE based firm-specific efficiency estimates. In this test the performance criterion is for what percentage of the 500 observations the EIV technique results in an estimate closer to the true value in comparison to the MLE. Based on this test EIV estimator based technical efficiency estimate outperforms MLE very convincingly.

To summarize, the MLE estimator is seriously biased when it comes to estimating the elasticity of labor and capital under measurement errors. MLE is also a poor choice if you want to estimate the technical efficiency for a particular firm. However, both MLE and EIV estimator estimate the mean technical efficiency level very well. Thus, in the presence of measurement errors in the input data the EIV estimator developed in section 2 is the preferred choice.

4. Empirical Example

4.1 Data

In this section we examine the impact of measurement errors on SFPF estimates of a production structure in actual data. We draw a cross-section of firms from the COMPUSTAT industrial data files maintained by Standard and Poor. These files consist of all the publicly traded firms on the U.S. stock exchanges for the period 1970-1989. The files provide information on balance sheet components, cash flow and income statements and other relevant financial information. The frequency of reporting is annual. We chose the year 1988 for our analysis as it provided the most number of firms with relevant information[31]. The number of employees, L_i , it employs measures Labor use by a firm. Standard practice is to define labor in terms of hours worked but this information is not available in COMPUSTAT. As we don't know the proportion of skilled versus unskilled workers as well as their quality level, this imparts a source of measurement error to our labor use variable. To calculate the output of a firm or the value added, Y_i, the cost of goods was subtracted from the sales figure[32]. To complete the value added calculations, total inventories were added to the above measure. The measure of capital K_i is the book value of total assets of a firm[33]**.** Thus, we have full information to estimate the production structure, and accompanying level of technical efficiency for 484 firms.

4.2 Production structure

The model that we estimate is identical to the one considered in the simulation study (see (30)). Tables 4a, 4b and 4c provide the necessary summary statistics for the data variables. In particular note that the covariance between z_1 and z_2 is almost zero. As before, $p_K = \text{Var}(\ln(K)) / \text{Var}(\ln(\underline{K}))$ and $p_L = \text{Var}(\ln(L)) / \text{Var}(\ln(\underline{L}))$ are the reliability ratios of capital and labor respectively, while p_{KL} is the "covariance reliability" ratio equal to $\text{Cov}(\ln(K), \ln(L)) / \text{Cov}(\ln(\underline{K}), \ln(\underline{L}))$. As explained in the simulation section, it is reasonable to set $p_{KL} = p_L$ if

---

[31] We could have chosen the year 1989 which is the terminal year of the database. Because of non-reporting of relevant information by quite a number of firms, the highest number of firms with usable information was present in 1988. Another reason for choosing 1988 was the fact that this year was characterized by a stable economic environment, especially the inflation situation and financial market volatility.

[32] Because the reporting procedure for the cost of goods component contains labor expenses, a component of the value added by a firm, the labor expense component was added to the above calculation. Since not every firm reports this item as an expense separate from cost of goods, this correction dropped the number of firms that could ultimately be used in the analysis.

[33] Using total assets as a proxy for productive, physical capital requires qualifications. First, this measure of assets includes the current investment component of a firm. Second, this measure includes cash and other short term liquid investments which may not be appropriate measures of physical capital. A justification for using this measure is the theoretical models and empirical evidence that extend the notion of production structure by incorporating the effects of liquidity and borrowing constraints [for e.g. see Gertler and Hubbard (1988), Dhawan (1997) etc.].

cov(z_1, z_2) is close to zero which is the case here. Also, we will only consider cases where $p_K = p_L = \widehat{p}$ [34]. Based on the summary statistics in table 4c, we can derive consistent estimates of the expected value and the variance of z. Given a particular reliability ratio $\widehat{p}$, we can then find consistent estimates of the expected value and variance of x as well as of the variance of u and these are:

$$\hat{m}_x = \begin{pmatrix} 4.93 \\ 1.15 \end{pmatrix} \quad \hat{\Sigma}_x = \widehat{p}\begin{pmatrix} 1.31 & -0.03 \\ -0.03 & 6.90 \end{pmatrix} \quad \hat{\Sigma}_u = (1-\widehat{p})\begin{pmatrix} 1.31 & -0.03 \\ -0.03 & 6.90 \end{pmatrix}$$

Thus, given this information and the discussion regarding reliability ratio bounds in section 3.2, the lowest possible value for the reliability ratio is 0.86. Any value lower than that is not feasible given that data characteristics.

4.3 Parameter estimates

Table 5 presents the estimates of the parameters in equations (30a) and (30b) using three techniques: OLS, the traditional MLE and the EIV estimator developed in this paper. The first row presents the estimates when simple OLS technique is used which can be characterized as estimating an "average" production function. As is well known, with no measurement errors, OLS will provide us with consistent but inefficient estimates of g, an inconsistent estimate of a and no estimates for s_e^2 and s_x^2. With measurement errors even the OLS estimate for the parameter g is inconsistent. In the second row the MLE based estimates are presented. Rows 3 to 9 display the estimates using the EIV technique based on likelihood function from equation (23). Each row provides a set of estimates for a particular common reliability ratio. These results should be interpreted as following: *If* the reliability ratio of labor and capital is 0.94 (say

[34] Given a value of p_L , p_K may deviate according to table (A1) in the appendix but we found that the estimated coefficients were not affected by setting it apart from p_L .

for example), then the consistently estimated coefficients are in this row. Based on these estimates for a, b, s^2 and l, we can then derive estimates for the elasticity of capital and labor (b_K and b_L) as well as the variances of e and x (s_e^2 and s_x^2) presented in table 6.

A number of interesting but not surprising results, given our simulation experience, are apparent from Tables 5 and 6. First, MLE underestimates the elasticity of capital. According to MLE, the return to capital is 0.6261 while it is as much as 0.7280 using the EIV technique and for the reliability ratio is 0.86. We also find that MLE estimates return to scale very well which then implies that it is over estimating the elasticity of labor. Second, as the reliability ratio decreases the estimated l increases while the estimated s_e^2 goes to zero. This happens because as the reliability ratio decreases, the variance of ub increases. Since it is the same data set, this will happen at the expense of a decline in the variance of e (s_e^2) and as it goes to zero l which is equal to s_x/s_e will increase[35]. Third, we find that MLE estimates, σ_ξ^2, the variance of x_i very well. This has important implications for the estimates of the technical efficiencies as discussed later in the next sub-section. MLE also overestimates the variance in e_i, which is natural, since it assumes no measurement errors.

4.4 Estimates of Technical Efficiencies

We begin first by considering the mean or average technical efficiency under varying degree of measurement errors presented in table 7. It is interesting to note that the average level of firm efficiency is almost independent of the assumption on measurement errors. The EIV estimates are also close to the MLE estimate of the average technical efficiency. This happens

---

[35] As a matter of fact, $\hat{p}$ = 0.86 is a lower bound for the reliability ratios. There simply is not enough variation in the data to support more measurement errors than this. With $\hat{p}$ = 0.86, the only disturbance to the model, except for the technical inefficiencies, are measurement errors as e vanishes in this case.

because the only parameter that determines the distribution of the technical efficiencies, s_x^2, is almost identical for MLE as well as for EIV technique regardless of the degree of measurement errors. At first, this may suggest that measurement errors are not an issue when it comes to technical efficiencies. However, as we know from the results of the simulation section, the EIV estimator outperforms the MLE estimator for firm-specific efficiency quite well.

Given that we have 486 firms, it is not possible to present the estimates of technical efficiency for each firm. To get an idea of the bias caused by measurement problems, we present technical efficiency estimates of the first ten firms in table 9. From this table, we note that one cannot predict the direction of the bias as the changes seems to be random. To explore this more, and to get an idea about how severe the problem could be, we ranked all the firms in the sample by their MLE based technical efficiency estimates. Then, as the reliability ratio was decreased, it was found that the relative ranking of the firms changed. For the reliability ratio 0.98 the maximum rank change was 23 on the upper side and 19 on the lower side. In addition, 50 percent of changes in ranks were between ± 2. For the lowest feasible reliability ratio of 0.86, 50% of the rank changes were within ± 15. For this particular reliability ratio the maximum rank change was 132 on the upper side and 131 on the lower side! In percentage terms the maximal change in firm level efficiency was 22% on the up side and 14% on the down side. This is an important outcome since the technical efficiency estimate tells us what percentage of "frontier output" the firm is delivering. This precludes the researcher, using MLE method under measurement errors, from establishing a comparative efficiency rankings of the firms in the sample as evident from the EIV estimate[36].

---

[36] In fact, we tested whether the changes in rankings was predictable (non-random) or not by running an AR(1) regression on a given firm's efficiency estimates for different reliability ratio assumption. It was observed that 90% of the auto-regressive coefficients were above 0.95 , with at least 50% of them being at or above 1, making the rank

5. Summary and Conclusions

This paper investigates the impact measurement errors in inputs have on estimates of production function parameters and firm-specific technical efficiency estimates in a cross-sectional SFPF setting. We first develop the methodology for estimating the standard cross sectional SFPF with measurement errors by using Fuller's reliability ratio concept. Next, our numerical simulation results show that the estimates (elasticity parameters) of the deterministic frontier, the distribution of the stochastic part of the frontier and the distribution of the technical inefficiency are very sensitive to the degree of measurement error. Our simulation results indicate that MLE will bias the elasticity coefficient estimates, and consequently the returns to scale feature. These biases are quite severe and clearly demonstrate the need for a method that consistently estimates the production function parameters for even small degree of measurement errors. The simulation exercise also shows that while MLE overestimates the variance of the composite error term, it underestimates the skewness parameter with the result that the variance of the technical efficiency parameter is consistently estimated. Although the mean level of technical efficiency or average sample efficiency is unaffected by the presence of measurement errors, the firm-specific estimate of technical efficiency will be seriously biased as it depends upon the estimated skewness parameter. Additionally, we also develop theoretical bounds regarding the possible values for the reliability ratios given the data summary statistics. These bounds are extremely useful for a researcher in a practical setting when he/she is analyzing the sensitivity of parameter estimates to the varying degree of belief regarding measurement errors.

---

changes to be very much a random outcome. A proper unit root test on these coefficients, although desirable could not be conducted as only 8 observations exist for each firm, which is not enough to test for the presence of unit root.

Next, a practical applicability of the reliability ratio estimator developed in this paper was demonstrated by applying it to actual firm level data from the U.S. industrial sector. For this data set issues regarding returns to scale feature, elasticity coefficients and firm-specific technical efficiency were explored in detail. Here we demonstrated how the relative ranking of the firms, by their technical efficiency estimates, changed when the degree of measurement errors was increased. Most importantly this change in ranking appeared to be random and not related to the change in the degree of measurement error. In addition the percent change in the firm-specific technical efficiency levels from its MLE estimate was also quite severe when the degree of measurement error increased. This exercise has serious implications for economic researchers who are engaged in inter-firm or inter-industry comparisons as ignoring measurement errors and relying solely on simple MLE estimates will most likely lead to erroneous efficiency comparisons.

The analysis in this paper has been undertaken for cross-sectional SFPF model with Cobb-Douglas production structure that in many respects is very simplistic. Consequently, practical issues such as analyzing technical change over time, and evolution of a firm's efficiency levels that requires a more general production structure (for e.g. Translog) in a panel setting are a subject matter of future research.

References

Aigner, D.J., C.A.K. Lovell, and P. Schmidt, 1977, Formulation and estimation of stochastic frontier production function models, Journal of Econometrics 6, 21-37.

Battese, G.E.and T.J. Coelli, 1988, Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data, Journal of Econometrics 38, 387-399.

Battese, G.E.and T.J. Coelli, 1992, Frontier production functions, technical efficiency and panel data: with application to paddy farmers in india, Journal of Productivity Analysis 3, 153-169.

Bauer, P.W., 1990, Recent developments in the econometric estimation of frontiers, Journal of Econometrics 46, 39-56.

Bauer, P.W., 1990, Decomposing TFP growth in the presence of cost in efficiency, non-constant returns to scale, and technological progress, Journal of Productivity Analysis 1, 287-299.

Beckers, D. and C. Hammond , 1987, A tractable likelihood function for the normal-gamma stochastic frontier model, Economics Letters 24, 33-38.

Dhawan, R., 1997, Asymmetric information and debt financing: The empirical importance of size and balance sheet factors, International Journal of the Economics of Business 4, 189-202.

Dhawan, R. and G. Gerdes, 1997, Estimating technological change using a stochastic frontier production function framework: evidence from U.S.. firm-level data, Journal of Productivity Analysis 8, 431-446.

Edgerton, D. and P. Jochumzen, 1999, Estimation in binary choice models with measurement errors, mimeo, Lund University.

Fuller, W.A., 1987, Measurement Error Models, Wiley, New York.

Gertler, M.L. and R.G. Hubbard, 1988, Financial factors and business fluctuations, in Financial Market Volatitlity, Federal Reserve Bank of Kansas City.

Greene, W.H., 1993, The econometric approach to efficiency analysis, in H. Fried, C.A.K. Lovell and S. Schmidt (eds), The Measurement of Productive Efficiency: Techniques and Applications, Oxford University Press, Oxford.

Greene, W.H., 1997a, Frontier production functions, in M. Hashem Pesaran and P. Schmidt (eds), Handbook of Applied Econometrics, Vol II: Microeconomics, Balckwell Publishers, Massachusetts.

Greene, W.H., 1997b, Econometric Analysis, Prentice Hall, New Jersey.

Jondrow, J., C.A.K. Lovell, I.S. Materov, and P. Schmidt, 1982, On the estimation of technical inefficiency in the stochastic frontier production function model, Journal of Econometrics 19, 233-238.

Klepper, S. and E. E. Leamer, 1984, Consistent sets of estimates for regressions with errors in all variables, Econometrica 52, 163-83.

Kumbhakar, S.C., 1987, Production frontiers and panel data: An application to U.S. class 1 railroads, Journal of Business and Economic Statistics 5, 249-255.

Kumbhakar, S.C., 1988, On the estimation of technical and allocative inefficiency using stochastic frontier functions: The case of U.S. class 1 railroads, International Economic Review 29, 727-743.

Kumbhakar, S.C. and L. Hjalmarsson, 1995, Labor-use efficiency in Swedish social insurance offices, Journal of Applied Econometrics 10, 33-47.

Meeusen, W.J. and van der Broeck, 1977, Efficiency estimation from cobb-douglas production functions with composed error, International Economic Review 18, 435-444.

Murphy K. and R. Topel, 1985, Estimation and Inference in Two Step Econometric Models, Journal of Business and Economic Statistics 3, 370-379.

Pal, M., Neogi, C. and B. Ghosh , 1998, Estimation of frontier production function with errors-in-variables: An illustration from Indian industry, in Sr. Chakravarty, D. Coondoo and R. Mukherjee (eds.) Quantitative Economics: Theory and Practice, Allied Publishers Limited, New Delhi, India.

Stevenson, R., (1980), Likelihood functions for generalized stochastic frontier estimation, Journal of Econometrics 13, 58-66.

Weinstein, M., 1964, The sum of values from a normal and a truncated normal distribution, Technometrics 6, 104-105.

Table 1. Traditional Maximum Likelihood Estimates of SFPF for Simulated Data

	α	γ_1	γ_2	σ^2	λ	σ_ξ^2	σ_ε^2
True Value*	1.700	0.6000	0.1000	1.000	2.000	0.8000	0.2000
Traditional Maximum Likelihood Estimates							
$\hat{p}$=1.00	1.711	0.5998	0.1030	1.013	2.132	0.8242	0.1883
	(0.07)	(0.02)	(0.02)	(0.13)	(0.39)	(0.15)	(0.04)
$\hat{p}$=0.90	1.691	0.5376	0.09289	1.055	1.757	0.7903	0.2644
	(0.08)	(0.02)	(0.02)	(0.14)	(0.36)	(0.18)	(0.05)
$\hat{p}$=0.80	1.685	0.4778	0.08143	1.100	1.606	0.7848	0.3154
	(0.09)	(0.02)	(0.03)	(0.14)	(0.35)	(0.19)	(0.06)
$\hat{p}$=0.70	1.676	0.4175	0.06741	1.121	1.494	0.7641	0.3568
	(0.12)	(0.02)	(0.03)	(0.16)	(0.39)	(0.22)	(0.07)
EIV Method Maximum Likelihood Estimates							
$\hat{p}$=1.00	1.711	0.5998	0.1030	1.013	2.132	0.8242	0.1883
	(0.07)	(0.02)	(0.02)	(0.13)	(0.39)	(0.15)	(0.04)
$\hat{p}$=0.90	1.691	0.5973	0.1032	0.9890	2.058	0.7903	0.1987
	(0.08)	(0.03)	(0.02)	(0.14)	(0.49)	(0.18)	(0.05)
$\hat{p}$=0.80	1.685	0.5972	0.1018	0.9833	2.095	0.7848	0.1985
	(0.09)	(0.03)	(0.03)	(0.14)	(0.62)	(0.19)	(0.06)
$\hat{p}$=0.70	1.676	0.5964	0.0963	0.9681	2.116	0.7641	0.2040
	(0.12)	(0.03)	(0.04)	(0.17)	(0.85)	(0.22)	(0.08)

* The data was simulated from the model $y_i = \alpha + x_i\gamma + \varepsilon_i - \xi_i$ with $z_i = x_i + u_i$. $x_i \sim N(0, 2\hat{p})$, $u_i \sim N(0, 2(1-\hat{p}))$ where $\hat{p}$ is the common reliability ratio of log of labor, log of capital (and thus of log capital by labor). $\hat{p}$ is varied in the table and $e \sim N(0, 0.2)$ and $x \sim N^+(0, 0.8)$. The standard errors are reported in parentheses.

Table 2. Mean Technical Efficiency And Reliability Ratio

Reliability ratio	Actual Value	MLE Estimate	EIV Estimate
1.00	0.5536	0.5496	0.5496
0.90	0.5536	0.5553	0.5553
0.80	0.5536	0.5562	0.5562
0.70	0.5536	0.5598	0.5598

Table 3. Comparing Firm-Specific Technical Efficiency Estimates

Reliability ratio	Average absolute deviation between EIV and true value	Average absolute deviation between MLE and true value	Percentage won by EIV
1.00*	N/A	N/A	N/A
0.90	0.0542	0.0755	94.47%
	(1.8×10^{-3})	(1.9×10^{-3})	(1.0%)
0.80	0.0463	0.125	99.27%
	(1.6×10^{-3})	(3.1×10^{-3})	(0.4%)
0.70	0.0406	0.161	99.72%
	(1.4×10^{-3})	(2.5×10^{-2})	(0.2%)

* For a reliability ratio of 1, MLE and EIV will produce exactly the same estimates and the formulas for expected value of the conditional technical efficiencies will coincide. N/A implies not applicable here. The standard errors are reported in parentheses.

Table 4a. Transformed and Non-Transformed Data Variable Means

	ln(Y)	ln(K)	ln(L)	Y = ln(Y/L)	Z_1 = ln(K/L)	Z_2 = ln(L)
Mean:	5.55	6.08	1.15	4.40	4.93	1.15

Table 4b. Untransformed Data Variance and Covariance Matrix

	ln(Y)	ln(K)	ln(L)
Ln(Y)	8.23	7.17	6.86
Ln(K)	7.17	8.14	6.86
Ln(L)	6.86	6.86	6.90

Table 4c. Transformed Data Variance and Covariance Matrix

	Y= ln(Y/L)	Z_1= ln(K/L)	Z_2= ln(L)
Y	0.78	0.76	0.26
z_1	0.76	1.31	-0.03
z_2	0.26	-0.03	6.90

Table 5. SFPF Parameter Estimates: OLS, MLE And EIV*

	α	γ_1	γ_2	σ^2	λ
OLS	1.5036	0.5781	0.04132	N/A	N/A
	(0.33)	(0.07)	(0.03)		
MLE	1.9195	0.6261	0.00071	0.7146	2.7072
	(0.10)	(0.02)	(0.01)	(0.07)	(0.37)
EIV	1.8565	0.6389	0.00073	0.7041	2.8904
$\hat{p}$=0.98	(0.10)	(0.02)	(0.01)	(0.07)	(0.46)
EIV	1.7910	0.6522	0.00074	0.6931	3.1270
$\hat{p}$=0.96	(0.10)	(0.02)	(0.01)	(0.07)	(0.54)
EIV	1.7226	0.6661	0.00076	0.6817	3.4486
$\hat{p}$=0.94	(0.11)	(0.02)	(0.01)	(0.06)	(0.68)
EIV	1.6512	0.6806	0.00078	0.6698	3.9188
$\hat{p}$=0.92	(0.11)	(0.02)	(0.01)	(0.06)	(0.99)
EIV	1.5767	0.6957	0.00079	0.6573	4.6979
$\hat{p}$=0.90	(0.11)	(0.02)	(0.01)	(0.06)	(1.66)
EIV	1.499	0.7115	0.00081	0.6443	6.376
$\hat{p}$=0.88	(0.12)	(0.03)	(0.01)	(0.07)	(4.19)
EIV	1.417	0.7280	0.00083	0.6307	18.51
$\hat{p}$=0.86	(0.14)	(0.03)	(0.01)	(0.07)	(169)

* The standard errors are in parentheses and N/A means not applicable.

Table 6. Basic Production Structure Estimates

	β_K	β_L	σ_ξ^2	σ_ε^2
OLS	0.5781	0.4351	N/A	N/A
MLE	0.6261	0.3746	0.6288	0.0858
EIV $\hat{p}$=0.98	0.6389	0.3618	0.6288	0.0753
EIV $\hat{p}$=0.96	0.6522	0.3485	0.6288	0.0643
EIV $\hat{p}$=0.94	0.6661	0.3347	0.6288	0.0529
EIV $\hat{p}$=0.92	0.6806	0.3203	0.6289	0.0409
EIV $\hat{p}$=0.90	0.6957	0.3051	0.6288	0.0285
EIV $\hat{p}$=0.88	0.7115	0.2893	0.6288	0.0155
EIV $\hat{p}$=0.86	0.7280	0.2728	0.6289	0.00184

Table 7. Average Technical Efficiency For the Sample

	MLE ($\hat{p}$=1)	$\hat{p}$=0.98	$\hat{p}$=0.96	$\hat{p}$=0.94	$\hat{p}$=0.92	$\hat{p}$=0.90	$\hat{p}$=0.88	$\hat{p}$=0.86
Mean	0.6008	0.6008	0.6008	0.6008	0.6008	0.6012	0.6068	0.6180

Table 8. Predicted Firm Efficiency of the First 10 Firms

	MLE($\hat{p}$=1)	$\hat{p}$=0.98	$\hat{p}$=0.96	$\hat{p}$=0.94	$\hat{p}$=0.92	$\hat{p}$=0.90	$\hat{p}$=0.88	$\hat{p}$=0.86
Firm 1	0.753	0.759	0.765	0.771	0.777	0.783	0.788	0.793
Firm 2	0.487	0.487	0.487	0.487	0.488	0.489	0.489	0.492
Firm 3	0.885	0.886	0.888	0.890	0.891	0.891	0.891	0.891
Firm 4	0.933	0.933	0.932	0.931	0.930	0.928	0.926	0.923
Firm 5	0.810	0.814	0.818	0.822	0.826	0.829	0.832	0.834
Firm 6	0.827	0.829	0.832	0.834	0.836	0.838	0.839	0.840
Firm 7	0.281	0.275	0.271	0.266	0.262	0.258	0.250	0.250
Firm 8	0.615	0.623	0.630	0.638	0.647	0.655	0.664	0.673
Firm 9	0.714	0..712	0.710	0.707	0.704	0.701	0.697	0.694
Firm 10	0.596	0.595	0.594	0.593	0.592	0.592	0.591	0.591