THE MAXWELL SCHOOL

CENTER FOR POLICY RESEARCH

WORKING PAPER SERIES

A Laplace Stochastic Frontier Model

William C. Horrace and Christopher F. Parmeter

ISSN: 1525-3066

426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020
(315) 443-3114 / email: ctrpol@syr.edu

Paper No. 166
April 2014

http://www.maxwell.syr.edu/CPR_Working_Papers.aspx

CENTER FOR POLICY RESEARCH –Spring 2014

Leonard M. Lopoo, Director
Associate Professor of Public Administration and International Affairs (PAIA)

Associate Directors

Margaret Austin
Associate Director

Budget and Administration
John Yinger
Trustee Professor of Economics and PAIA
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Badi H. Baltagi ..... Economics
Robert Bifulco ..... PAIA
Leonard Burman ..... PAIA
Thomas Dennison ..... PAIA
Alfonso Flores-Lagunes ..... Economics
Sarah Hamersma ..... PAIA
William C. Horrace ..... Economics
Yilin Hou ..... PAIA
Hugo Jales ..... Economics
Duke Kao ..... Economics
Jeffrey Kubik ..... Economics
Yoonseok Lee ..... Economics
Amy Lutz ..... Sociology
Yingyi Ma ..... Sociology
Jerry Miner ..... Economics
Cynthia Morrow ..... PAIA
Jan Ondrich ..... Economics
John Palmer ..... PAIA
David Popp ..... PAIA
Stuart Rosenthal ..... Economics
Rebecca Schewe ..... Sociology
Amy Ellen Schwartz ..... PAIA/Economics
Perry Singleton ..... Economics
Michael Wasylenko ..... Economics
Peter Wilcoxen ..... PAIA

GRADUATE ASSOCIATES

Dana Balter ..... PAIA
Joseph Boskovski ..... PAIA
Christian Buerger ..... PAIA
Brianna Cameron ..... Lerner Center
Emily Cardon ..... PAIA
Sarah Conrad ..... Lerner Center
Carlos Diaz ..... Economics
Pallab Ghosh ..... Economics
Lincoln Groves ..... PAIA
Chun-Chieh Hu ..... Economics
Jung Eun Kim ..... PAIA
Yan Liu ..... Sociology
Michelle Lofton ..... PAIA
Roberto Martinez ..... Lerner Center
Qing Miao ..... PAIA
Nuno Abreu Faro E Mota ..... Economics
Judson Murchie ..... PAIA
Sun Jung Oh ..... Social Science
Katie Oja ..... Lerner Center
Laura Rodriquez-Ortiz ..... PAIA
Jordan Stanley ..... Economics
Kelly Stevens ..... PAIA
Tian Tang ..... PAIA
Liu Tian ..... Economics
Rebecca Wang ..... Sociology
Ian Wright ..... Economics
Pengju Zhang ..... PAIA

STAFF

Kelly Bogart ..... Administrative Specialist
Karen Cimilluca ..... Office Coordinator
Kathleen Nasto ..... Administrative Assistant
Candi Patterson ..... Computer Consultant
Mary Santy ..... Administrative Assistant
Katrina Wingle ..... Administrative Assistant

Abstract

We propose a Laplace stochastic frontier model as an alternative to the traditional model with normal errors. An interesting feature of the Laplace model is that the distribution of inefficiency conditional on the composed error is constant for positive values of the composed error, but varies for negative values. Therefore, it may be ideally suited for analyzing industries with many forms on or close to the efficient frontier. A simulation study suggests that the model performs well relative to the normal-exponential model when the two-sided error is misspecified. A brief application to US Airlines is provided.

JEL No. C12, C16, C44, D24

Keywords: Stochastic frontier, efficient estimation

The authors are grateful to Peng Liu for excellent research assistance. The paper has benefited from excellent suggestions at various conferences. All errors are ours alone.

William C. Horrace - Corresponding Author. Email: whorrace@maxwell.syr.edu, Tel: 315-443-9061. Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244-1020

Christopher F. Parmeter - Email: cparmeter@bus.miami.edu, Tel: 305-284-4397. University of Miami, Department of Economics, 517-E Jenkins Building, Miami, FL, 33124.

1 Introduction

Given a sample of firm-level data, parametric stochastic frontier models specify production output (or cost) as the sum of a linear (in parameters) response function and an additively composed error, consisting of a two-sided error, representing noise, and a one-sided error, representing inefficiency. See, for example, Aigner, Lovell and Schmidt (1977), Battese and Coelli (1988, 1992) and Kumbhakar and Lovell (2001). It is often assumed that the two-sided error is normally distributed and the one-sided error is either truncated normal or exponential, leading to the familiar "normal-truncated normal" and "normal-exponential" stochastic frontier models. In either case the distribution of inefficiency conditional on the composed error (for each firm) is truncated normal, and the traditional predictor of firm-level inefficiency is the mean of this distribution evaluated at the regression residual (in place of the composed error) for each firm. See Jondrow, Lovell, Materov and Schmidt (1982) for the cross-sectional case and Battese and Coelli (1988) for the panel data case. The normal-truncated normal and normal-exponential models have been widely applied, and the conditional mean of inefficiency (evaluated at the residual) for each firm is often reported as the standard inefficiency predictor.[1]

This paper dispenses with the normality assumption of the two-sided error in favor of a Laplace error. There are several reasons why the change may be justified. First, we prove that in the absence of inefficiency the Laplace model reduces to the Least Absolute Deviations (LAD) estimator. It is well-known that the LAD estimator is less sensitive to outliers than OLS, so if outliers are an issue, the Laplace specification may be the preferred choice for empiricists, particularly when inefficiency in the population is close to zero.[2] Consequently, for populations with little to no inefficiency, the Maximum Likelihood Estimator (MLE) of the production function under Laplace errors will be close to the conditional median function. Second, the true data generation process may possess Laplace errors. Third, comparing alternative specifications of the stochastic frontier model sheds

[1] In our discussion of the conditional inefficiency distribution we refer to the conditioning arguments as both the "regression residual" and the "composed error". These are synonymous in the sense that in stochastic frontier analysis it is always assumed (for the purpose of *ex post* inference on inefficiency) that the estimate of the production function equals the true production function. See Horrace and Schmidt (1996) for a discussion of this concept.

[2] Outliers have been considered in the data envelopment analysis (DEA) literature. See Wilson (1993) for an example and Cazals, Florens and Simar (2002). While we do not explicitly address outliers here, ours is the first paper to propose LAD as a solution to outliers in the stochastic frontier literature.

light on robustness features of the approach.

It is interesting to note that nearly all studies that have investigated alternative distributional assumptions for the stochastic frontier model have focused on changes to the distribution of inefficiency. For example, Li (1996) studies a uniform distribution, Carree (2002) proposes a binomial distribution, Tsionas (2007) a Weibull distribution, Almanidis, Qian and Sickles (2014) a doubly truncated normal distribution and Kumbhakar, Parmeter and Tsionas (2013) consider a mixture distribution. However, all these models assume the two-sided error to be normally distributed. While it seems self-evident that changing the distribution of inefficiency may shed light on different patterns of estimated technical efficiency, the fact that *conditional* efficiency is all that can be estimated in the stochastic frontier model suggests that the two-sided error distribution may be equally important.[3]

In addition to these arguments we prove that the Laplace error model possesses the unique feature that the distribution of inefficiency conditional on the composed error is constant for positive realizations of the composed error and varies for negative realizations (regardless of the shape of the unconditional distribution of inefficiency). Therefore, all firms with positive regression residuals receive the same technical inefficiency predicted value. Since large values of the regression residual are associated with small values of inefficiency, the Laplace error model allows for the possibility that multiple firms may be tied for the highest efficiency score in the sample. (Ties occur with probability zero in stochastic frontier models with normally distributed errors.) In this regard, the proposed model may be ideally suited for analyzing highly competitive or mature industries where there may be many firms on or close to the efficient frontier. This is related to the "mostly stars, few dogs" discussion of Almanidis, Qian and Sickles (2014). We discuss this concept more in the sequel.

The constant conditional distribution for positive errors has implications for inference on inefficiency in the Laplace model. Often an empirical goal is to perform inference on these conditional distributions across firms. This is the essence of the marginal prediction intervals of Horrace and Schmidt (1996) and Kim and Schmidt (2008), the simultaneous intervals of Horrace and Schmidt

[3]Nguyen (2010) is the only work we are aware of that considers studying alternative distributions for the two-sided error term (Cauchy and Laplace).

(2000), the efficiency ranking methods of Horrace (2005) and Flores-Lagunes, Horrace and Schmidt (2007), the expected conditional ranks of Horrace, Richards and Wright (2014), the bagging procedure of Simar and Wilson (2009), and the confidence intervals of Wheat, Greene and Street (2013). All these papers use a normal stochastic frontier model specification, resulting in conditional distributions across firms from the family of truncated normals which are identical with probability zero. In contrast the Laplace model can produce identical conditional distributions across some firms, thereby simplifying any subsequent inference. That is, if firms with positive realizations of the regression residual have the same conditional distribution in the sample, then these firms are stochastically equivalent, and any testing procedure with a null hypothesis that the distributions are the same will never reject the hypothesis. We only need conduct inference on those firms with regression residuals less than zero. While we do not tackle inference in the Laplace model here, the model itself opens up an entire new line of inquiry for understanding inefficiency uncertainty when errors are Laplace.

Since the canonical stochastic frontier specification is a normal-truncated normal model (i.e., the unconditional distribution of inefficiency is truncated normal), a natural specification to consider in the case of Laplace errors is the Laplace-truncated Laplace model. Both the truncated normal and truncated Laplace distributions are functions of a location parameter (before truncation) defined on the real numbers. In both cases truncation produces a very rich class of distributional shapes, based on whether or not the location parameter is positive or negative. However, the truncated Laplace possesses the interesting feature that when the location parameter is negative, the resulting distribution is exponential and is no longer a function of the location parameter, so the Laplace-truncated Laplace model nests the Laplace-exponential model.[4] This has implications for maximum likelihood estimation of the Laplace-truncated Laplace model: numerical estimation of the location parameter can be restricted to the non-negative real numbers, potentially making numerical searches simpler (faster). Indeed, while we provide the distributional theory for a Laplace-Truncated Laplace model, we focus the discussion of estimation on the Laplace-exponential case.

[4]This is in contrast to the truncated normal distribution, which is only independent of the location parameter in the half normal case. Nguyen (2010) was the first to consider the Laplace-exponential version of this model.

The constant conditional mean/median feature of the proposed model also has implications for the empirical practice of identifying "super-efficient firms" in empirical research. Super-efficient firms are typically determined based on estimated production frontiers that display residuals with positive skewness (Green and Mayes, 1991) or are simply taken as the firms with the highest values of estimated technical efficiency (see Timmer, 1971). The proposed stochastic frontier model provides a theoretical justification for identification of super-efficient or equally efficient firms that does not hinge on the skewness of the residuals or subjective selection of the highest ranked firms. In any given sample from the Laplace model, one would focus attention on only those firms with negative residuals. We show how this can be done in our brief application to US airlines data. When a Cobb-Douglas cost function is estimated for the pooled cross-section, the least-squares residuals possess the wrong skew, and the normal-exponential estimate of the inefficiency distribution variance is zero. The Laplace-truncated Laplace model produces a non-zero variance estimate, and constant conditional means for about 30% of the observations. The distribution of the conditional mean is discussed with and without these "super-efficient" observations trimmed from the data.

The assumption of a Laplace distribution for the two-sided error term also has implications for deconvolution of the composed error in the cross-sectional case. Horrace and Parmeter (2011) develop deconvolution techniques for the stochastic frontier model for cross-sectional data that allow for estimation of the *unconditional* distribution of inefficiency, given that the distribution of the two-sided error is normally distributed. They find that consistent estimation of the inefficiency distribution has a slow, $\ln n$, convergence rate. It was shown by Fan (1992) that if one of the error components is normally distributed, then $\ln n$ is the best possible convergence rate for consistent estimation of the distribution of the other component. However, if the distribution is Laplace, then estimation of the distribution will have faster polynomial convergence rates. Insofar as the stochastic frontier model is a deconvolution exercise, a Laplace assumption for the two-sided error component may have advantages over the typical normality assumption.[5] These advantages are (indirectly) explored in this paper using simulation techniques. We show that in the fully-parametric model the estimation losses associated with a normality assumption on the two-sided component,

[5]See Meister (2004) for a technical description of the loss associated with assuming normality in deconvolution settings when the true density is Laplace.

when it is actually Laplace, are greater than those associated with a Laplace assumption, when it is actually normally distributed.

Unlike the normal model, identification in the Laplace model does not appear to hinge on negative skewness of the LAD or OLS residuals. Both the normal and Laplace specifications imply that the skew of the composed error is negative, and Waldman (1982) shows in the case of the normal-half normal model that when the skew of the OLS residuals is positive (i.e., their skew has the wrong sign), the maximum likelihood estimate of the variance of inefficiency is zero, there is no inefficiency in the sample, and MLE reduces to OLS.[6] There are two ways in which our proposed model does not suffer from the wrong skew. First, pathologically, given that we dispense with normality, wrong skewness of the OLS residuals has no implications for the Laplace model because LAD (not OLS) is the limiting estimator (as the variance of the inefficiency error component goes to zero). Second, even though a Waldman (1982) type result holds (namely that LAD is a stationary point for our likelihood function), the Hessian is naturally indeterminate due to the non-differentiable point in the likelihood function, so that the stationary point is not stable (nor is it unique in the sense of minimizing the absolute deviations). This is borne out in simulations where no apparent connection between the skewness of the LAD residuals and the stationary point exists and where instability of the results was sometimes encountered.

In what follows we present several propositions related to the Laplace stochastic frontier model and derive the conditional mean and median of inefficiency, based on a truncated Laplace inefficiency distribution. This is followed by a detailed set of simulations which examine the model under misspecification of the error term and under the 'wrong skew' condition; comparisons are made to the normal stochastic frontier model. An empirical exercise contrasts insights on inefficiency when compared to the normal-exponential setup. Conclusions offer avenues for further research.

[6]When this occurs in practice there are several solutions prescribed. Greene (1995) discusses several remedies, while Simar and Wilson (2009) discuss a 'bagging' approach to inference.

2 The Laplace Stochastic Frontier Model

We consider the parametric stochastic frontier model for a cross-section of firms:

$$y_i = x_i\beta + v_i - u_i, \quad i = 1, \ldots, n \tag{1}$$

where y_i is productive output for firm i, x_i is a k vector of production inputs, and β is an unknown parameter vector. The $v_i \in \mathbb{R}$ are random variables representing shocks to the frontier. Let v_i have *iid* symmetric Laplace distribution with scale parameter γ. That is, the probability density of v is,

$$f_v(v) = \frac{1}{2\gamma} e^{-|v|/\gamma}. \tag{2}$$

The distribution of v is absolutely continuous with respect to the Lebesgue measure and possesses a single non-differentiable point at $v = 0$. The $u_i \in \mathbb{R}_+$ are *iid* random variables representing productive inefficiency with absolutely continuous density function $f_u(u)$. The error components are v and u, and the composed error is defined as $\varepsilon = v - u$. Then we have the following result.

Lemma 1 *The conditional distribution $f_u(u|\varepsilon)$ is constant in ε for non-negative values of ε.*

Proof. Since the error components are independent, $f_{uv}(u, v) = f_u(u)f_v(v)$, and

$$f_{u\varepsilon}(u, \varepsilon) = f_u(u)f_v(\varepsilon + u) = f_u(u)\frac{1}{2\gamma}e^{-\frac{|\varepsilon+u|}{\gamma}}.$$

When $\varepsilon \geq 0$,

$$f_{u\varepsilon}(u, \varepsilon) = f_u(u)\frac{1}{2\gamma}e^{-\frac{\varepsilon+u}{\gamma}},$$

so that,

$$f_\varepsilon(\varepsilon) = \int_0^\infty f_{u\varepsilon}(u, \varepsilon)du = \frac{1}{2\gamma}e^{-\frac{\varepsilon}{\gamma}} \int_0^\infty \left(f_u(u)e^{-\frac{u}{\gamma}}du = \frac{1}{2\gamma}e^{-\frac{\varepsilon}{\gamma}}A(1/\gamma) \right. \text{ for } \varepsilon \geq 0,$$

where $A(1/\gamma)$ is the Laplace transform of the density of inefficiency. Then,

$$f_u(u|\varepsilon) = \frac{f_{u\varepsilon}(u, \varepsilon)}{f_\varepsilon(\varepsilon)} = f_u(u)\frac{1}{2\gamma}e^{-\frac{\varepsilon+u}{\gamma}} \left(\frac{1}{2\gamma}e^{-\frac{\varepsilon}{\gamma}}A(1/\gamma) \right)^{-1} = f_u(u)e^{-\frac{u}{\gamma}}A(1/\gamma)^{-1} \text{ for } \varepsilon \geq 0.$$

■

When $\varepsilon < 0$, the conditional distribution varies in ε. The constant conditional distribution is related to (but not precisely due to) the memorylessness property of the exponential distribution, in which conditioning arguments in exponential processes are uninformative. In the context of the present model, this property effectively partitions the observations into those that are strongly informative ($\varepsilon < 0$) and those that are weakly informative ($\varepsilon \geq 0$) . When $\varepsilon < 0$, inefficiency (u) dominates the noise term (v) in the composed error, and $f_u(u|\varepsilon)$ is well-informed (varies across ε). When $\varepsilon \geq 0$, noise dominates, and $f_u(u|\varepsilon)$ is less informed, in which case, the distribution is constant. Furthermore, $\varepsilon \geq 0$ implies $v \geq 0$, so draws of v behave as if they are exponential (half-Laplace), so the near-memoryless feature of the conditioning argument is revealed. This has implications for estimating technical efficiency as we shall see.

Maximum likelihood estimation (MLE) of the frontier model proceeds from the distribution $f_\varepsilon(\varepsilon)$, which is absolutely continuous because it is the convolution of absolutely continuous distributions. A common question in the stochastic frontier literature is what happens to the likelihood function and MLE as the variance of u vanishes or $V(u) \to 0$? Then we have the following result.

Lemma 2 *If the characteristic function of u converges pointwise to 1 as $V(u) \to 0$, then the distribution of u converges uniformly to that of degenerate random variable at 0, and the distribution of ε converges uniformly to a Laplace distribution with scale parameter γ, so the MLE of the stochastic frontier model is the LAD estimator.*

The proof follows immediately from the characteristic function continuity theorem and the fact that the characteristic function of ε is the product of the characteristic functions of v and u. That is, if $\varphi_u(t)$ is the characteristic function of u, and $\lim_{V(u)\to 0} \varphi_u = 1$, then $\lim_{V(u)\to 0} \varphi_\varepsilon$ is the characteristic function of a zero-mean Laplace random variable. Therefore, the likelihood converges uniformly to the Laplace likelihood and in the limit the MLE is the LAD estimator. This result can be reformulated for the case where the distribution of u is degenerate at any finite point in $\mathbb{R}$, in which case the limiting MLE is the LAD estimator displaced by a constant. The result holds more generally for any continuous f_v, so that $f_\varepsilon \to f_v$ uniformly, and in the limit MLE is based on f_v (perhaps displaced by a constant). In practice all the difficulty is in understanding the limiting

behavior of the distribution of u and determining whether or not the limiting MLE is a stable stationary point in the parameter space of the likelihood.[7]

Obviously, the aforementioned lemmas hold for any continuous inefficiency distribution: truncated normal, double truncated normal, exponential, etc.. We now introduce a truncated Laplace specification for inefficiency. Let u_i have an *iid* truncated (at zero) Laplace distribution with location parameter $\mu \in \mathbb{R}$ and scale parameter $\theta > 0$. That is:

$$f_u(u) = \frac{c(\mu)}{2\theta} e^{-|u-\mu|/\theta}, \quad u \geq 0, \tag{3}$$

with

$$c(\mu)^{-1} = \begin{cases} 1 - 0.5e^{-\mu/\theta}, & \mu \geq 0 \\ 0.5e^{\mu/\theta}, & \mu < 0 \end{cases}.$$

The distribution of u is absolutely continuous and possesses a single non-differentiable point at $u = \mu$. Allowing for heterogeneity across i in the inefficiency distributions of u_i is an interesting area of investigation, but it is not considered here.

Notice that when $\mu \leq 0$ the truncated Laplace distribution reduces to an exponential distribution:

$$f_u(u, \mu \leq 0) = \theta^{-1} e^{-u/\theta}, \quad \text{for } u \geq 0, \tag{4}$$

so that the Laplace-truncated Laplace model nests the Laplace-exponential model. In particular, when $\mu \leq 0$, the distribution is no longer a function of μ. This has implications for maximum likelihood estimation of the Laplace-truncated Laplace model, as we shall see. The result that the truncated Laplace distribution reduces to the exponential distribution when the mean of the Laplace (before truncation) is non-positive can be generalized as follows:

Lemma 3 Any *density function that is defined as a multiple of an exponential density to the right of zero will possess an exponential density when truncated to the left of zero.*

Proof. Since the density function before truncation must satisfy non-negativity and have cumulations less than or equal to unity if suffices to restrict it to the class of density functions $f_{u^*}(z) = ae^{-bz}$

[7] This is the essence of the Waldman (1982) results but with v normal and u half normal.

for constants $a > 0$, $b > 0$, and $z \geq 0$ for some random variable $u^* \in \mathbb{R}$. Then it is easy to show that after truncation to the left

$$f_u(y) = \frac{f_{u^*}(y)}{\int_0^\infty f_{u^*}(z)dz} = be^{-by},$$

for $u \geq 0$, which is the exponential density. ■

Therefore, the Laplace distribution with non-positive mean μ and scale parameter θ is a member of the class of distributions that satisfy the requirements of the lemma. In particular to the right of zero, the Laplace distribution satisfies the lemma with $a = 0.5/\theta$ and $b = 1/\theta$. Hence, after truncation we have an exponential distribution $\theta^{-1}e^{-u/\theta}$ when $\mu \leq 0$.

Moments for the truncated Laplace distribution of u when $\mu \leq 0$ are standard exponential results: $E(u, \mu \leq 0) = \theta$, $E(u^2, \mu \leq 0) = 2\theta^2$, $E(u^3, \mu \leq 0) = 6\theta^3$. Moment results when $\mu > 0$ are non-standard:

$$\begin{aligned}
E(u, \mu > 0) =& c(\mu)\left[\mu + 0.5e^{-\mu/\theta}\right] > 0, \\
E(u^2, \mu > 0) =& c(\mu)\left[\mu^2 + \theta^2\left(2 - e^{-\mu/\theta}\right)\right] > 0, \\
E(u^3, \mu > 0) =& c(\mu)\left[\mu^3 + 6\mu\theta + 3\theta^3 e^{-\mu/\theta}\right] > 0,
\end{aligned}$$

so the inefficiency distribution exhibits positive skewness ($E(u^3) > 0$), not unlike the traditional truncated normal distribution. For completeness the characteristic function of the truncated Laplace distribution is given in the following lemma:

Lemma 4 *If random variable u has the truncated Laplace distribution given in equation 3, then its characteristic function is for $\iota = \sqrt{-1}$:*

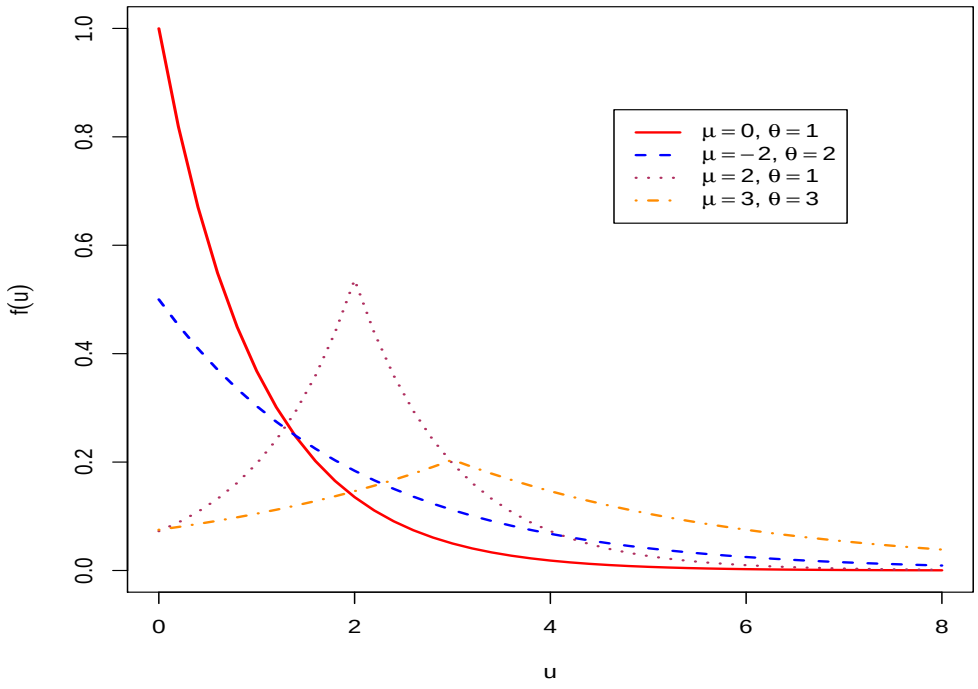
$$\varphi_u(t) = \begin{cases} c(\mu)\left[\frac{e^{\iota t\mu}}{1+t^2\theta^2} - \frac{e^{-\mu/\theta}}{2(1+\iota t\theta)}\right] & \mu \geq 0 \\ \frac{1}{1-\iota t\theta} & \mu < 0 \end{cases}.$$

The proof follows easily from the standard formula $\varphi_u(t) = E[e^{\iota tu}]$. When $\mu \geq 0$, the first bracketed term corresponds to the usual characteristic function for a symmetric Laplace random variable, while the second term controls for the truncation (the level of asymmetry). When μ is large (relative to θ) and positive, the bracketed second term is small (asymmetry is low), and the first term dominates (symmetry is high). As μ moves towards 0, the distribution becomes

more asymmetric and the second term dominates until $\mu = 0$ at which point $c(\mu) = 2$, and the characteristic function becomes that of an exponential distribution. For $\mu < 0$, the characteristic function is exponential. As $\theta \to 0$ for $\mu \geq 0$, the characteristic function of u is that of a degenerate random variable at μ.

Figure 1 plots several variants of the truncated Laplace density and shows that with $\mu \leq 0$ the distribution is exponential. We consider the $\mu \leq 0$ case below and show that the distribution of inefficiency conditional on the composed error is also exponential. In the normal-exponential stochastic frontier this conditional distribution is truncated-normal. Therefore, when the unconditional distribution of inefficiency is exponential, the conditional distribution of inefficiency is a truncated version of the two-sided error. It would be interesting to see if this result is generalizable.

Figure 1: Density of Truncated at 0 Laplace for various combinations of θ and μ.



2.1 The Laplace-Exponential Model: $\mu \leq 0$

To fix ideas we consider the case where $\mu < 0$, so that the distribution of inefficiency is exponential. This case is also important if one is interested in making estimation performance comparisons to the normal-exponential model, as we do in the sequel. To simplify notation, define constants $\lambda_+ = \gamma\theta/(\gamma+\theta)$ and $\lambda_- = \gamma\theta/(\gamma-\theta)$. Then we have the following results on the density of the composed error:

$$f_\varepsilon(\varepsilon, \mu < 0) = \begin{cases} \frac{1}{2\gamma\theta}\lambda_+ e^{-\varepsilon/\gamma}, & \varepsilon \geq 0 \\ \frac{1}{2\gamma\theta}\left[(\lambda_+ + \lambda_-)e^{\varepsilon/\theta} - \lambda_- e^{\varepsilon/\gamma}\right] & \varepsilon < 0, \quad \theta \neq \gamma \\ \frac{1}{2\gamma\theta}\left(\lambda_+ - \varepsilon\right)e^{\varepsilon/\gamma}, & \varepsilon < 0, \quad \theta = \gamma. \end{cases} \tag{5}$$

The case where $\gamma = \theta$ is provided for completeness given that λ_- does not exist in this instance. However, the $\gamma = \theta$ case can be easily derived by application of l'Hospital's rule to the $\gamma \neq \theta$ case. The point $\gamma = \theta$ is a (bounded) continuity point in the space of γ and θ in the density function. The first part of (5) ($\varepsilon \geq 0$) corresponds to the case where $v \geq u \geq 0$. Here, $f_\varepsilon(\varepsilon, \mu < 0)$ is a rescaled exponential distribution with parameter γ, implying that the Laplace portion of the convolution dominates the exponential portion. Alternatively, when $\varepsilon < 0$, the situation is more complicated, because it is not clear whether v or u dominates.

When $\mu < 0$ and $\varepsilon \geq 0$ the conditional distribution of inefficiency is exponential and per Lemma 1, is not a function of ε,

$$f_u(u|\varepsilon, \mu < 0) = \lambda_+^{-1} e^{-u/\lambda_+} \text{ for } \varepsilon \geq 0.$$

This result implies that when v is relatively large and u is small, implying ε is positive, there will be numerous firms that have the smallest conditional mean inefficiency. It is important to emphasize that this distribution is NOT $f_u(u|\varepsilon \geq 0, \mu < 0)$, which is obviously not a function of ε.[8] It is $f_u(u|\varepsilon)$ *evaluated* at any $\varepsilon \geq 0$. When ε is negative we have that the conditional mean of inefficiency depends on ε:

$$E\left[u|\varepsilon, \mu < 0\right] = \begin{cases} \lambda_+ & \varepsilon \geq 0 \\ \frac{e^{\varepsilon/\gamma}\lambda_-^2 + e^{\varepsilon/\theta}\left[\varepsilon(\lambda_- - \lambda_+) - \lambda_-^2 + \lambda_+^2\right]}{2\gamma\theta f_\varepsilon(\varepsilon, \mu<0)}, & \varepsilon < 0, \end{cases} \tag{6}$$

[8] In fact, $f_u(u|\varepsilon < 0, \mu < 0)$ is not a function of ε either.

and the conditional median, m, is:

$$m[u|\varepsilon,\mu<0]=\begin{cases}\lambda_+\ln(2), & \varepsilon\geq 0\\ -\lambda_-\ln\left[1-(\gamma-\theta)f_\varepsilon(\varepsilon,\mu<0)e^{-\varepsilon/\gamma}\right] & \left(\begin{array}{ll}\varepsilon<0, & m[u|\varepsilon,\mu<0]<-\varepsilon\\ \varepsilon<0, & m[u|\varepsilon,\mu<0]\geq-\varepsilon.\end{array}\right.\\ -\lambda_+\ln\left[(\gamma+\theta)f_\varepsilon(\varepsilon,\mu<0)e^{\varepsilon/\gamma}\right] & \end{cases} \tag{7}$$

where $f_\varepsilon(\varepsilon,\mu<0)$ corresponds to the density of ε in equation (5).[9] To make equation 7 operational, use the first part of the equation ($\lambda_+\ln(2)$) for any $\varepsilon\geq 0$. For any $\varepsilon<0$ calculate the second part of the equation and check if the condition $m[u|\varepsilon,\mu<0]<-\varepsilon$ is satisfied. If so, the median calculation for that realization of $\varepsilon<0$ is complete. If not, then it must be true that $m[u|\varepsilon,\mu<0]\geq-\varepsilon$, and the third part of the equation is used to calculate the condition median. The second part of the equation is based on a left-tail probability of the conditional density, and the third part of the equation is based on a right-tail probability.

The conditional median function, $m[u|\varepsilon]$, may be particularly relevant to the Laplace stochastic frontier model, since the limiting case (as $\theta\to 0$) is the Laplace regression which yields the LAD estimator, the estimator of the conditional median of y_i.[10] We do not formally prove that λ_+ and $\lambda_+\ln(2)$ are minima (in ε) for the conditional mean and median function (respectively), but it is certainly borne out in simulations (proving this would be equivalent to showing that the conditional mean/median is monotonic in ε, a well-known fact in the normal-truncated normal model). The constant conditional mean/median for $\varepsilon\geq 0$ implies that the proposed model has the potential for ties for the least inefficient firms in the sample. This is a meaningful result, for if we believe that an industry tends to have many highly efficient firms, then the Laplace-exponential model may be more appropriate that the normal-exponential model, which produces conditional mean/median ties with probability zero. If we substitute $\varepsilon=0$ into the conditional mean and median formulae for the $\varepsilon<0$ cases (above), we get the constant conditional mean and median results, so $\varepsilon=0$ is a continuity point in the conditional mean/median function. We now consider the Laplace-truncated Laplace model.

[9] See Appendix A for the $\gamma=\theta$ case, which we exclude here and throughout the rest of the paper.

[10] For that matter, in any parametric stochastic frontier model, the conditional median may be more informative than the conditional mean given that the conditional distribution of inefficiency is skewed.

2.2 The Laplace-Truncated Laplace model: $\mu \geq 0$

Consider the case where $\mu \geq 0$ so that the distribution of inefficiency is no longer exponential. Then we have the following result on the density of the composed error:

$$f_\varepsilon(\varepsilon, \mu \geq 0) = \begin{cases} \frac{c(\mu)}{4\theta\gamma}\left[(\lambda_+ + \lambda_-)\, e^{-(\mu+\varepsilon)/\gamma} - \lambda_- e^{-\mu/\theta} e^{-\varepsilon/\gamma}\right], & \varepsilon \geq 0, \\ \frac{c(\mu)}{4\theta\gamma}\left[(\lambda_+ + \lambda_-)\, e^{-|\mu+\varepsilon|/\gamma} - \lambda_+ e^{-\mu/\theta} e^{\varepsilon/\gamma} + (\lambda_+ - \lambda_-)\, e^{-|\mu+\varepsilon|/\theta}\right], & \varepsilon < 0 \end{cases}$$

When $\mu = 0$, the density above reduces to that in equation (5) where $\mu < 0$, so $\mu = 0$ is a continuity point in the distribution of the composed error (as is the $\gamma = \theta$ case). Therefore, the equation above holds generally with the following parameterization of μ:

$$\mu_* = \begin{cases} \mu, & \mu \geq 0 \\ 0, & \mu < 0. \end{cases}$$

Then, a general formula for the distribution of the composed error is:

$$f_\varepsilon(\varepsilon) = \begin{cases} \frac{c(\mu_*)}{4\theta\gamma}\left[(\lambda_+ + \lambda_-)\, e^{-(\mu_*+\varepsilon)/\gamma} - \lambda_- e^{-\mu_*/\theta} e^{-\varepsilon/\gamma}\right], & \varepsilon \geq 0, \\ \frac{c(\mu_*)}{4\theta\gamma}\left[(\lambda_+ + \lambda_-)\, e^{-|\mu_*+\varepsilon|/\gamma} - \lambda_+ e^{-\mu_*/\theta} e^{\varepsilon/\gamma} + (\lambda_+ - \lambda_-)\, e^{-|\mu_*+\varepsilon|/\theta}\right], & \varepsilon < 0. \end{cases} \tag{8}$$

Again, when $\varepsilon \geq 0$ the conditional distribution of u is not a function of ε:

$$f_u(u|\varepsilon) = \frac{c(\mu_*)}{\gamma\theta f_\varepsilon(0)} e^{-\frac{|u-\mu_*|}{\theta} - \frac{u}{\gamma}}, \quad \varepsilon \geq 0$$

in general. When $\mu_* = 0$ (corresponding to *any* $\mu < 0$) and $\varepsilon \geq 0$, the conditional distribution above reduces to the exponential: $f_u(u|\varepsilon, \mu_* = 0) = \lambda_+^{-1} e^{-u/\lambda_+}, \quad \varepsilon \geq 0.$

The conditional mean function implied by equation (8) is:

$$E\left[u|\varepsilon\right] = \begin{cases} \frac{c(\mu_*)}{4\theta\gamma f_\varepsilon(0)}\left\{\left[\mu_*(\lambda_+ + \lambda_-) + \lambda_+^2 - \lambda_-^2\right] e^{-\mu_*/\gamma} + \lambda_-^2 e^{-\mu_*/\theta}\right. & , \quad \varepsilon \geq 0, \\ \frac{c(\mu_*)}{4\theta\gamma f_\varepsilon(\varepsilon)}\left\{\left[\mu_*(\lambda_+ + \lambda_-) + \lambda_+^2 - \lambda_-^2\right] e^{-(\mu_*+\varepsilon)/\gamma}\right. & \\ \quad \left. + \left[\varepsilon(\lambda_- - \lambda_+) + \lambda_-^2 - \lambda_+^2\right] e^{-(\mu_*+\varepsilon)/\theta} + \lambda_+^2 e^{-\mu_*/\theta} e^{\varepsilon/\gamma}\right. & , \quad \varepsilon < 0,\ \mu_* \geq -\varepsilon, \\ \frac{c(\mu_*)}{4\theta\gamma f_\varepsilon(\varepsilon)}\left\{\left[\mu_*(\lambda_+ + \lambda_-) - \lambda_+^2 + \lambda_-^2\right] e^{(\mu_*+\varepsilon)/\gamma}\right. & \\ \quad \left. + \left[\varepsilon(\lambda_- - \lambda_+) - \lambda_-^2 + \lambda_+^2\right] e^{(\mu_*+\varepsilon)/\theta} + \lambda_+^2 e^{-\mu_*/\theta} e^{\varepsilon/\gamma}\right. & , \quad \varepsilon < 0,\ \mu_* < -\varepsilon. \end{cases} \tag{9}$$

Hence, the conditional mean function is constant when $\varepsilon \geq 0$ for all values of the distributional parameters μ_*, γ and θ. (See Appendix A for the $\gamma = \theta$ case.)

When $\mu \geq 0$ and $\varepsilon \geq 0$ the conditional median is:

$$m[u|\varepsilon] = \begin{cases} \lambda_- \ln\left[1 + 2\frac{f_\varepsilon(0)}{c(\mu_*)}(\gamma - \theta) e^{\mu_*/\theta}\right], & \varepsilon \geq 0,\ m[u|\varepsilon] < \mu_*, \\ -\lambda_+ \ln\left[2\frac{f_\varepsilon(0)}{c(\mu_*)}(\gamma + \theta) e^{-\mu_*/\theta}\right], & \varepsilon \geq 0,\ m[u|\varepsilon] \geq \mu_* \end{cases} \tag{10}$$

Notice in the last case, when $\mu_* = 0$, we get the result $m[u|\varepsilon] = \lambda_+ \ln(2)$, which is the exponential result when $\varepsilon \geq 0$.[11] The $m[u|\varepsilon] < \mu_*$ result above is based on the left-tail probability and the $m[u|\varepsilon] \geq \mu_*$ is based on the right-tail probability of the conditional distribution of u (See Appendix A for the $\gamma = \theta$ case). In practice, calculate the first part of the formula for any $\varepsilon \geq 0$. If the condition $m[u|\varepsilon] < \mu_*$ is satisfied, then the calculation is complete. Otherwise, use the second part of the equation to calculate the conditional median for that realization of $\varepsilon \geq 0$.

When $\mu \geq 0$ and $\varepsilon < 0$ the conditional median is:

$$m\left[u|\varepsilon\right] = \begin{cases} \lambda_+ \ln\left[1 + 2\frac{f_\varepsilon(\varepsilon)}{c(\mu_*)}(\gamma+\theta)e^{\mu_*/\theta}e^{-\varepsilon/\gamma}\right], & \varepsilon < 0, m[u|\varepsilon] < \min(\mu_*, -\varepsilon), \\ \lambda_- \ln\{2\frac{f_\varepsilon(\varepsilon)}{c(\mu_*)}(\gamma-\theta)e^{\mu_*/\theta}e^{\varepsilon/\gamma} \\ \quad -e^{-\varepsilon/\lambda_-}(\lambda_+/\lambda_- - 1) + e^{2\varepsilon/\gamma}\lambda_+/\lambda_- \end{cases}, \quad \varepsilon < 0, -\varepsilon < m[u|\varepsilon] < \mu_*, \\ \begin{cases} -\lambda_- \ln\{-2\frac{f_\varepsilon(\varepsilon)}{c(\mu_*)}(\gamma-\theta)e^{-\mu_*/\theta}e^{-\varepsilon/\gamma} \\ \quad +e^{-\mu/\lambda_-}(\lambda_+/\lambda_- + 1) - e^{-2\mu/\gamma}\lambda_+/\lambda_- \end{cases}, \quad \varepsilon < 0, -\varepsilon \geq m[u|\varepsilon] \geq \mu_*, \\ -\lambda_+ \ln\left[2\frac{f_\varepsilon(\varepsilon)}{c(\mu_*)}(\gamma+\theta)e^{-\mu_*/\theta}e^{\varepsilon/\gamma}\right], & \varepsilon < 0, m[u|\varepsilon] \geq \max(\mu_*, -\varepsilon). \end{cases} \tag{11}$$

The first three results are based on left tail probabilities and the last is based on the right tail probability of the conditional distribution of u (See Appendix A for the $\gamma = \theta$ cases). In practice, calculate each part of equation (11) in order, and stop when the relevant condition [e.g., $m[u|\varepsilon] < \min(\mu_*, -\varepsilon)$] is satisfied.

2.3 Estimation

The model in (1) may be estimated via corrected ordinary least squares (COLS). That is estimate the model using ordinary least squares with intercept ($x_{1i} = 1$) to get $\widehat{\beta}_j$, $j = 2, ..., k$, which are consistent for β_j, $j = 2, ..., k$ as $n \to \infty$. The intercept $\widehat{\beta}_1$ is consistent for $\beta_1 - E(u)$, but it can be corrected based on the moments of the OLS residuals. To do so, we only need the parametric assumption on the distribution of u and moments conditions on v. The parameters γ, θ and μ_* may then be consistently estimated from moments of the residuals using the parametric assumption on the distribution of v.[12]

Since the distribution of the convolution $v - u$ is absolutely continuous (and bounded), the fully parametric model can also be consistently estimated via the likelihood principle using equation

[11] There are no analogous results for $\mu_* = 0$ in the first case, because it would violate the condition $m[u|\varepsilon] < \mu_*$.

[12] COLS is a conditional mean interpretation of the model in (1). Alternatively, one could estimate LAD, in which case the estimates are the conditional median function.

(8). However for simplicity and for comparative purposes, we focus our discussion on the Laplace-exponential model where $\mu_* = 0$, so the log-likelihood is:

$$\ln L(\varepsilon_i|\gamma,\theta,\beta) = const. - \sum_{i:\varepsilon_i\geq 0}[\ln(\gamma+\theta)+\varepsilon_i/\gamma] + \sum_{i:\varepsilon_i<0}\left(\ln[(\gamma-\theta)(e^{\varepsilon_i/\gamma}-e^{\varepsilon_i/\theta})+(\gamma+\theta)e^{\varepsilon_i/\theta}\right].$$

The score and Hessian are derived in Appendix B. When maximizing the function, we recommended using a variety of alternative starting values, since for small values of θ and γ the likelihood function can be fairly flat (although the parameters remain identified). For the simulation and application that follow we use a variety of gradient and non-gradient optimization methods (including Nelder-Mead and Particle Swarm) and find similar performance across all methods. Lastly, in some simulations a maximum is found where the Hessian is not invertible, leading to problems constructing estimates of the variance-covariance matrix of the estimated parameters.

It should be noted that the conditional mean predictor of technical inefficiency is $E[u|\varepsilon]$ with the estimation residual, $\hat{\varepsilon} = y - x\hat{\beta}$, substituted for the composed error. Given that $E[u|\hat{\varepsilon}]$ is a mean and u is bound from below by 0, the conditional mean can never equal zero (the same arguments apply to the conditional median function, $m[u|\hat{\varepsilon}]$). Therefore, the Laplace-truncated Laplace model will never predict that firms are fully efficient. However, this can be said of all stochastic frontier models which assume an absolutely continuous distribution for inefficiency. For a counter-example of this see the zero-inefficiency models of Kumbhakar, Parmeter and Tsionas (2013) and Rho and Schmidt (2013).

2.4 A LAD Stationary Point

Waldman's (1982) classic result of a stable stationary point at the OLS estimator for the normal-half normal likelihood extends to the Laplace model, albeit at the LAD estimator. LAD poses well-known analytic difficulties related to the non-differentiability of the likelihood.[13]

Lemma 5 *The point* $\left(\hat{\beta}_{LAD},\hat{\gamma},0\right)$ *is a stationary point for the Laplace-exponential likelihood, where* $\hat{\beta}_{LAD}$ *is the LAD estimator of* β*, and* $\hat{\gamma}$ *is the maximum likelihood estimator of the scale parameter for the Laplace distribution.*

[13] See Koenker (2005) for a discussion of non-uniqueness and instability of the LAD estimator.

Taking the limit as $\theta \to 0$ of the score function given in Appendix B yields:

$$\lim_{\theta \to 0} S(\gamma, \theta, \beta) = \lim_{\theta \to 0} \begin{bmatrix} \frac{\partial \ln L}{\partial \beta} \\ \frac{\partial \ln L}{\partial \gamma} \\ \frac{\partial \ln L}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \gamma^{-1} \sum_{i=1}^{n} sign(\varepsilon_i) x_i \\ -n\gamma^{-1} - \gamma^{-2} \sum_{i=1}^{n} |\varepsilon_i| \\ \gamma^{-1}(n_2 - n_1) \end{bmatrix}, \tag{12}$$

where n_1 is the cardinality of the set of positive residuals, and n_2 is the cardinality of the set of negative residuals. Notice that the first component of $S(\gamma, \theta, \beta)$ is precisely the LAD condition for the vector of contrasts. The second component is the MLE of the scale parameter of the Laplace distribution ($\hat{\gamma} = n^{-1} \sum_{i=1}^{n} |\hat{\varepsilon}_i|$), while the third component is zero, given that $\hat{\beta}_{LAD}$ ensures $n_1 = n_2$ for even n.

As a curiosity, the Hessian matrix as $\theta \to 0$ evaluated at the stationary point produces,

$$\lim_{\theta \to 0} H(\gamma, \theta, \beta) = n\gamma^{-2} \begin{bmatrix} \mathbf{0}_{k \times k} & \mathbf{0}_{k \times 1} & \mathbf{0}_{k \times 1} \\ \mathbf{0}'_{k \times 1} & -1 & 0 \\ \mathbf{0}'_{k \times 1} & 0 & 1 \end{bmatrix}, \tag{13}$$

which is clearly indefinite, as expected.

The indefiniteness of the Hessian at the stationary point induces instability of the likelihood function. This is in contrast to the stable Waldman (1982) result. In simulations, when the sample size is small, we find that occasionally the LAD stationary point is a local maximizer of the likelihood function. However, as the sample size increases, LAD is no longer the local maximizer. This result is interesting in light of Simar and Wilson (2009), who show that even for sample sizes as large as 1,000,000, the normal-half normal model can produce convoluted errors with the wrong skew, and a local maximum at OLS. In contrast, more observations from the Laplace-exponential model assists in moving the solution away from the stationary point. We surmise that this is due to the fact that the indefiniteness of the Hessian in no way depends on skewness.

3 Monte Carlo Experiments

We conduct two sets of experiments. In the first set of experiments, we examine the performance of the model in equation (1) under misspecification of the distribution of v. In particular, we generate data from a normal-exponential model, but fit a Laplace-exponential model to the data.

We then generate data from a Laplace-exponential model, but fit a normal-exponential model. We find for a variety of signal-to-noise ratios, $V(u)/V(v)$, and sample sizes, n, that Laplace-exponential estimation very often (but not always) outperforms normal-exponential estimation, when v is misspecified.

In a second set of experiments, we let $V(u)/V(v) = 0.2$, so that both models produce incorrectly skewed residuals with high probability and repeatedly calculate maximum likelihood estimates of the variance of inefficiency from simulated draws. We find that the Waldman (1982) result holds for the normal-exponential model. That is, when the OLS residuals are positively skewed, the normal-exponential MLE of the variance of inefficiency is zero. This may imply that the normal exponential model possesses a stationary point at $\theta = 0$. However, this does not appear to be the case for the Laplace-exponential model, which produces non-zero MLE estimates of the variance, when the skew of the LAD residuals is positive.

3.1 Misspecification Experiments

We consider sample sizes of $n = 100$, 200, 400 and 800, and signal-to-noise ratios of $V(u)/V(v) = 10^{-1}$, $10^{-3/4}$, $10^{-1/2}$, $10^{-1/4}$, 10^{0}, $10^{1/4}$, $10^{1/2}$, $10^{3/4}$ and 10^{1} for $V(v) + V(u) = 1$. As is common practice (see, Olson, Schmidt and Waldman, 1980), we ignore regressors in equation (1), so our data generation and estimation is for the model $y = v - u$. For the normal-exponential model, $v \sim N(0, \sigma^2)$ and the signal to noise ratio is θ/σ. For the Laplace-exponential model, the variance of v is $2\gamma^2$, so the signal to noise ratio is $\theta/\left(\sqrt{2}\gamma\right)$. Maximum likelihood estimation of the misspecified model is done by differential evolution (Storn and Price, 1997), which is a genetic based algorithm.[14] Mean squared error of the variance parameters θ and γ (or σ) is estimated using 1,000 simulations.

Monte Carlo results are in Table 1. The first column of the table contains the various signal-to-noise ratios. Reading across the table for each signal-to-noise ratio, the MSE for either θ, γ or σ are reported for various sample sizes and parametric assumptions. The first set of columns (Normal - Laplace Estimates) contains the MSE for θ and γ for experiments where data are generated from a normal-exponential (NE) model, but a Laplace-exponential (LE) model is estimated. For

[14] Differential evolution was performed using the `DEoptim` package in R with default tolerance of 1e-8, a local-to-best evolution strategy, cross-over probability of 0.5, a step size of 0.8 and 500 iterations.

Table 1: Mean Squared Error Results For Misspecified Models.

n		$n = 100$	$n = 200$	$n = 400$	$n = 800$		$n = 100$	$n = 200$	$n = 400$	$n = 800$
$V(u)/V(v)$		Normal Data Laplace Estimates					Laplace Data Normal Estimates			
10^{-1}	θ	0.0127	0.0096	0.0063	0.0044	θ	0.0119	0.0066	0.0038	0.0020
	γ	0.0535	0.0495	0.0473	0.0460	σ	0.0831	0.0812	0.0836	0.0863
$10^{-3/4}$	θ	0.0143	0.0088	0.0053	0.0040	θ	0.0146	0.0089	0.0050	0.0031
	γ	0.0570	0.0514	0.0498	0.0496	σ	0.0814	0.0758	0.0775	0.0781
$10^{-1/2}$	θ	0.0113	0.0058	0.0040	0.0028	θ	0.0141	0.0077	0.0040	0.0024
	γ	0.0568	0.0534	0.0522	0.0508	σ	0.0668	0.0653	0.0679	0.0667
$10^{-1/4}$	θ	0.0089	0.0045	0.0022	0.0012	θ	0.0106	0.0056	0.0029	0.0014
	γ	0.0500	0.0463	0.0451	0.0445	σ	0.0604	0.0599	0.0548	0.0557
10^{0}	θ	0.0090	0.0049	0.0023	0.0012	θ	0.0107	0.0046	0.0024	0.0012
	γ	0.0335	0.0310	0.0303	0.0295	σ	0.0502	0.0455	0.0430	0.0419
$10^{1/4}$	θ	0.0100	0.0048	0.0024	0.0012	θ	0.0102	0.0053	0.0024	0.0012
	γ	0.0184	0.0170	0.0160	0.0156	σ	0.0287	0.0254	0.0234	0.0229
$10^{1/2}$	θ	0.0099	0.0049	0.0024	0.0012	θ	0.0096	0.0054	0.0026	0.0012
	γ	0.0086	0.0077	0.0069	0.0065	σ	0.0126	0.0114	0.0098	0.0099
$10^{3/4}$	θ	0.0096	0.0049	0.0026	0.0013	θ	0.0093	0.0049	0.0024	0.0012
	γ	0.0038	0.0030	0.0026	0.0024	σ	0.0053	0.0041	0.0040	0.0036
10^{1}	θ	0.0108	0.0051	0.0025	0.0013	θ	0.0102	0.0047	0.0026	0.0012
	γ	0.0016	0.0011	0.0010	0.0008	σ	0.0022	0.0016	0.0013	0.0012

All results are for 1000 Monte Carlo Simulations. "Normal Data Laplace Estimates" is a normal-exponential data generation process estimated by a Laplace-exponential model. "Laplace Data Normal Estimates" is a Laplace-exponential data generation process estimated by a normal-exponential model. For each value of $V(u)/V(v)$, the first row corresponds to the MSE of θ and the second row corresponds to either the MSE of γ (when estimating a Laplace-exponential model) or the MSE of σ (when estimating a normal-exponential model).

example, when the signal-to-noise ratio is 10^{-1} (a relatively noisy experiment), $n = 100$, and we fit a LE model to NE data, we have $MSE(\hat{\theta}) = 0.0127$ (first row) and $MSE(\hat{\gamma}) = 0.0535$ (second row). As the sample size increases to $n = 800$ the $MSE(\hat{\theta})$ and $MSE(\hat{\gamma})$ decrease to 0.0044 and 0.0460, respectively. Obviously, when the data are NE and the model is LE, $MSE(\hat{\gamma})$ is calculated from differences in the maximum likelihood estimates, $\hat{\gamma}$, and the true parameter value, σ. The second set of columns (Laplace Data - Normal Estimates) contains the MSE for θ and σ for experiments where data are LE, but an NE model is fit. For example, when the signal-to-noise ratio is 10^{-1} (a relatively noisy experiment) and $n = 100$, we have $MSE(\hat{\theta}) = 0.0119$ (first row) and $MSE(\hat{\sigma}) = 0.0831$(second row). Here, $MSE(\hat{\sigma})$ is calculated from differences in the maximum likelihood estimates, $\hat{\sigma}$, and the true parameter value, γ.

The results in Table 1 are interesting. First, when we fit an LE model to NE data, the MSEs

of the estimated parameters are always decreasing in sample size. However, this is not always the case when we fit an NE model to LE data. For example, when the signal-to-noise is 10^{-1} and the data are NE, the $MSE(\hat{\gamma})$ for the LE fitted model is always decreasing in sample size: 0.0535, 0.0495, 0.0473 and 0.0460 for sample sizes, 100, 200, 400, 800, respectively. However, this is not the case for LE data and an NE fitted model, where the $MSE(\hat{\sigma})$ is increasing from 0.0812 to 0.0836 to 0.0863 as we move from sample sizes of 200, 400 and 800, respectively. This is not to say that estimating a LE model on NE is consistent, but is does seem to produce better estimates (in terms of MSE) than the alternative. It should be noted that this "inconsistency problem" only occurs in estimating γ as $\hat{\sigma}$; the estimates of θ are always decreasing in MSE as the sample size grows. This is somewhat reassuring, since we are only misspecifying the distribution of v (parameterized by γ or σ). Also, the problem only occurs in the noisiest experiments with signal-to-noise ratios of 10^{-1}, $10^{-3/4}$, $10^{-1/2}$ and $10^{-1/4}$. It occurs only once when the signal-to-noise is $10^{1/4}$, but the difference between $MSE(\hat{\sigma})$ of 0.0098 ($n = 400$) and 0.0099 ($n = 800$) is probably due to statistical noise.

Secondly, we see in Table 1 that the MSE's associated with fitting an LE model to NE data are often smaller then when fitting an NE model to LE data. For example, when the signal to noise is $10^{-3/4}$ and $n = 100$, we see that $MSE(\hat{\gamma}) = 0.0570$ when fitting an LE model to NE data, while $MSE(\hat{\sigma}) = 0.0814$ when fitting an NE model to LE data. In fact, it is always the case that $MSE(\hat{\gamma}) < MSE(\hat{\sigma})$ across all combinations of n and signal-to-noise. When making these types of comparisons in estimating θ the results are mixed. For example, in the noisiest experiment ($V(u)/V(v) = 10^{-1}$), $MSE(\hat{\theta})$ is smaller when fitting a NE model to LE data, than when fitting an LE model to NE data (compare 0.0119 to 0.0127, 0.0066 to 0.0096, 0.0038 to 0.0063 and 0.0020 to 0.0044). However, in the $V(u)/V(v) = 10^{-1/3}$ experiments the opposite occurs (compare 0.0106 to 0.0089, 0.0056 to 0.0045, 0.0029 to 0.0022 and 0.0014 to 0.0012). Some of these differences may be due to statistical noise, but in certain cases they clearly are not.

In summary, our experiments suggest that fitting an LE model to NE data is always better for estimating the variance of (the misspecified component), but fitting an NE model to LE data is often better for estimating the variance of inefficiency (21 out of 36 experiments or about 80% of the time). Of course estimating both variances accurately is important when estimating the conditional

mean predictor of inefficiency. In this regard fitting an LE model to NE data appears to be more reliable: of the 72 MSE comparisons across the two panels of Table 1, the MSE associated with fitting an LE model to NE data are smaller 51 times or about 70% of the time. This suggest that if faced with making a guess at a normal or Laplace distribution for v, Laplace may be the better choice from the standpoint of the mean squared error of the variance parameters.

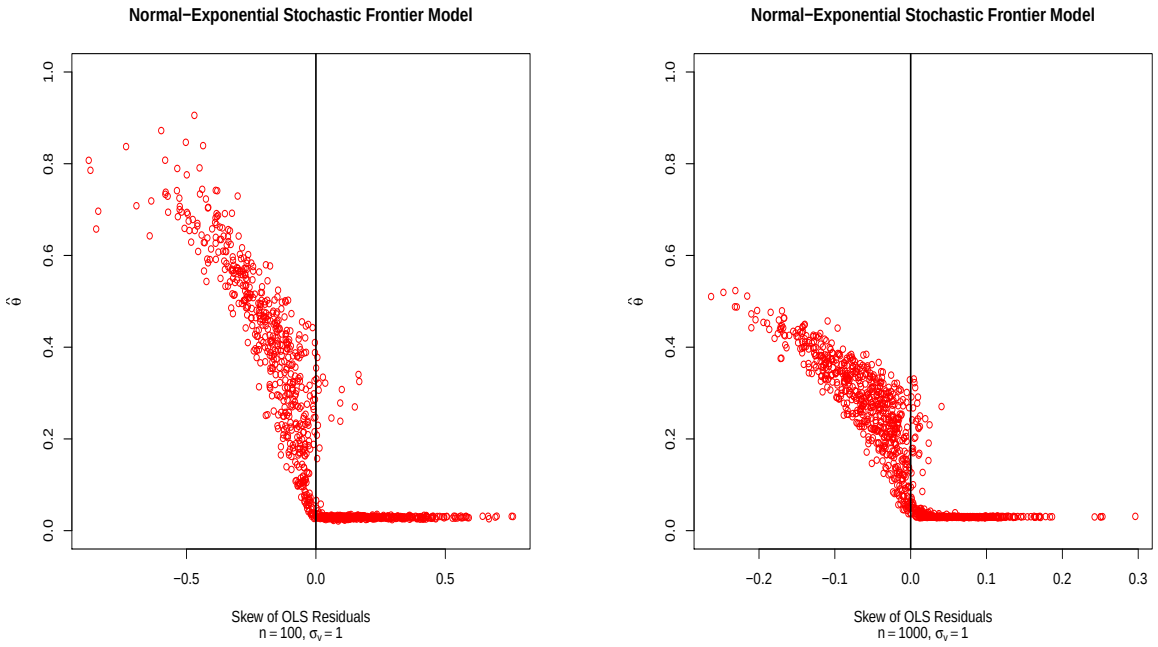
3.2 Wrong Skew Experiments

Similar to the previous experiment, we draw data from both the NE and LE models, but with $V(u)/V(v) = 0.2$ and $n = 100$ and 1000. Maximum likelihood estimates of θ and the skew of the OLS and LAD residuals are recorded for each draw. Figures 2 and 3 show this relationship for 1,000 draws. It is clear from figure 2 that the NE model produces $\hat{\theta} = 0$ when the skew of the OLS residuals in wrong (positive). This suggests that Waldman's (1982) stationary-point result holds for the normal-exponential model as well as for the normal-half normal model. Further, as is consistent with the insights of Simar and Wilson (2009), for a signal-to-noise ratio of 0.2, even for samples of size 1,000, nearly 40% of random draws from a normal-exponential model generate the wrong-skew.

Figure 3, suggest no relationship between the LAD residuals' skew and $\hat{\theta}$, which confirms our results on the Hessian in section 2.4. Additionally, for $n = 1000$ nearly all of the estimates of θ are non-zero, suggesting that the perceived instability of the likelihood function around the stationary point dissipates quickly as the sample size increase. For example, in only 12 of the 1000 simulations did the Laplace-exponential maximum likelihood estimator produce an estimator that was 0. Compare this to the almost 40% of the simulations for the normal-exponential maximum likelihood estimator which produced 0 estimates.

While the lack of a stationary point based on the skewness of the OLS or LAD residuals does not solve the "wrong skew" problem *per se*, it suggests that a Laplace-truncated Laplace frontier model (which nests the LE model) may produce non-zero $\hat{\theta}$ when the NE (or normal-truncated normal) model does not. In the empirical application that follows, a Laplace-Truncated Laplace model does not suffer from the "wrong skew" problem of the normal-exponential model.

Figure 2: Skew of OLS Residuals and MLE of Inefficiency Variance, Normal-Exponential.



(a) $n = 100$ (b) $n = 1000$

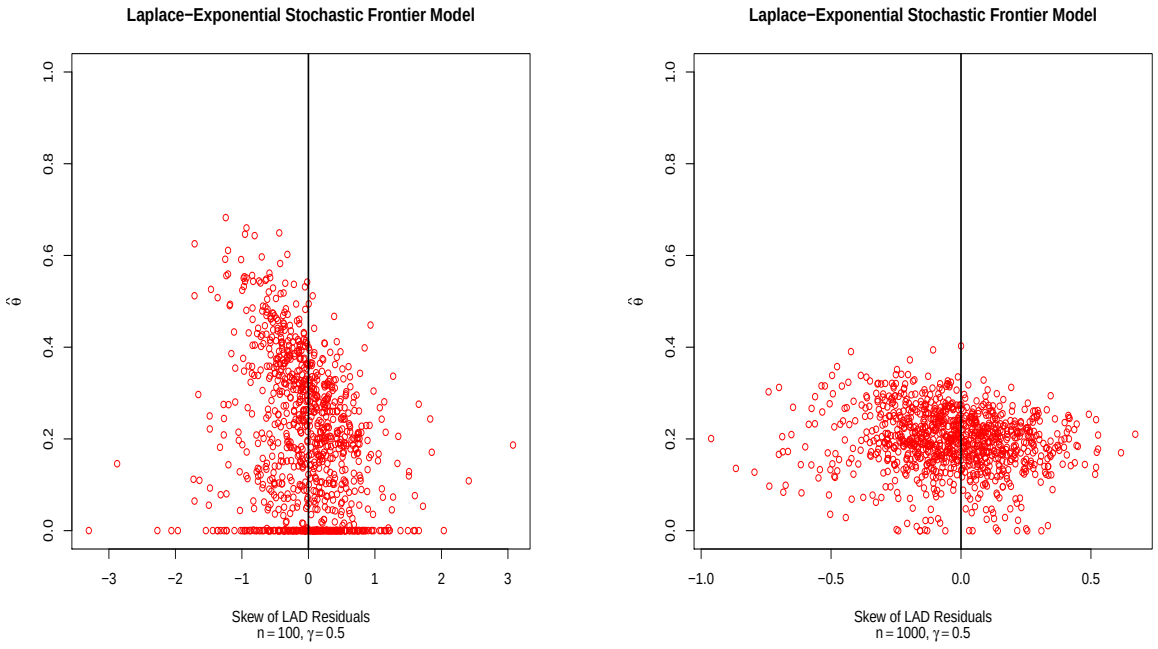
4 Brief Application to US Airlines Data

To illustrate the utility of the Laplace-truncated Laplace model, we estimate a stochastic cost function using the US Airlines data from Greene's "Econometric Analysis" textbook, Edition 7, Table F6.1.[15] The dataset are a panel of 90 observations of 6 airlines over 15 years (1970-1984) and consist of: *Costs* (in \$1,000), *Output* (in revenue passenger miles, index number), *Price* (the price of fuel), and *Load* (load factor, the average capacity utilization of the fleet). We ignore the panel structure and estimate OLS, the normal-exponential model, the Laplace-truncated Laplace model, and the Laplace exponential model with the pooled cross-section, based on the following Cobb-Douglas cost specification:

$$\ln Cost_i = \alpha + \beta_1 \ln Output_i + \beta_2 \ln Fuel_i + \beta_3 Load_i + u_i + v_i. \tag{14}$$

[15]Data are available on Prof. Greene's NYU website where he states, "These data are a subset of a larger data set provided to the author by Professor Moshe Kim. They were originally constructed by Christensen Associates of Madison, Wisconsin."

Figure 3: Skew of LAD Residuals and MLE of Inefficiency Variance, Laplace-Exponential.



(a) $n = 100$ (b) $n = 1000$

Other models were fitted (e.g., Trans-log), but only the Cobb-Douglas produced OLS residuals with the wrong skew, which is what we desire for the purposes of illustration.

Estimation results are in Table 2. The coefficients for each cost input are statistically significant and are fairly stable across our four models. For example, the coefficients for Output are 0.883, 0.884, 0.891, and 0.894 for OLS, the normal-exponential model, the Laplace-truncated Laplace model, and the Laplace exponential model, respectively.[16] The maximum likelihood estimates of the distributional parameters are in the last three rows of the table. The normal-exponential model produces an estimate of $\hat{\theta} = 0.000$, implying that the "wrong skew" issue exists in the normal-exponential model and that zero-inefficiency may be a stationary point in the normal-exponential likelihood function. The Laplace-truncated Laplace model produces a significant and non-zero estimate of $\hat{\theta} = 0.063$, indicating that the model does not suffer from the wrong skew issue that plagues the normal-exponential. However, the Laplace-truncated Laplace model does not produce a significant estimate of the location parameter, $\hat{\mu} = 0.893$, so we estimate the Laplace-exponential

[16]Standard errors are based on numerical calculation of the Hessian at the parameter estimates.

Table 2: Airline Costs Functions, 1970-1984, Pooled Cross-Section.

	Ordinary Least Squares	Normal-Exponential	Laplace-Truncated Laplace	Laplace-Exponential
Intercept	9.517**	9.537**	8.772**	9.632**
	(0.229)	(0.303)	(0.750)	(0.216)
ln(*Output*)	0.883**	0.884**	0.891**	0.894**
	(0.013)	(0.013)	(0.012)	(0.011)
ln(*Fuel*)	0.454**	0.453**	0.441**	0.439**
	(0.020)	(0.020)	(0.020)	(0.019)
Load	-1.628**	-1.655**	-1.559**	-1.546**
	(0.345)	(0.346)	(0.342)	(0.254
$\hat{\gamma}$	-	0.015**	0.063**	0.086**
		(0.006)	(0.027)	(0.012)
$\hat{\theta}$	-	0.000	0.063**	0.043*
		(0.006)	(0.026)	(0.022)
$\hat{\mu}$	-	-	0.893	-
			(0.757)	

** - significant at 5% level, * - significant at 10% level. Sample size is 90.

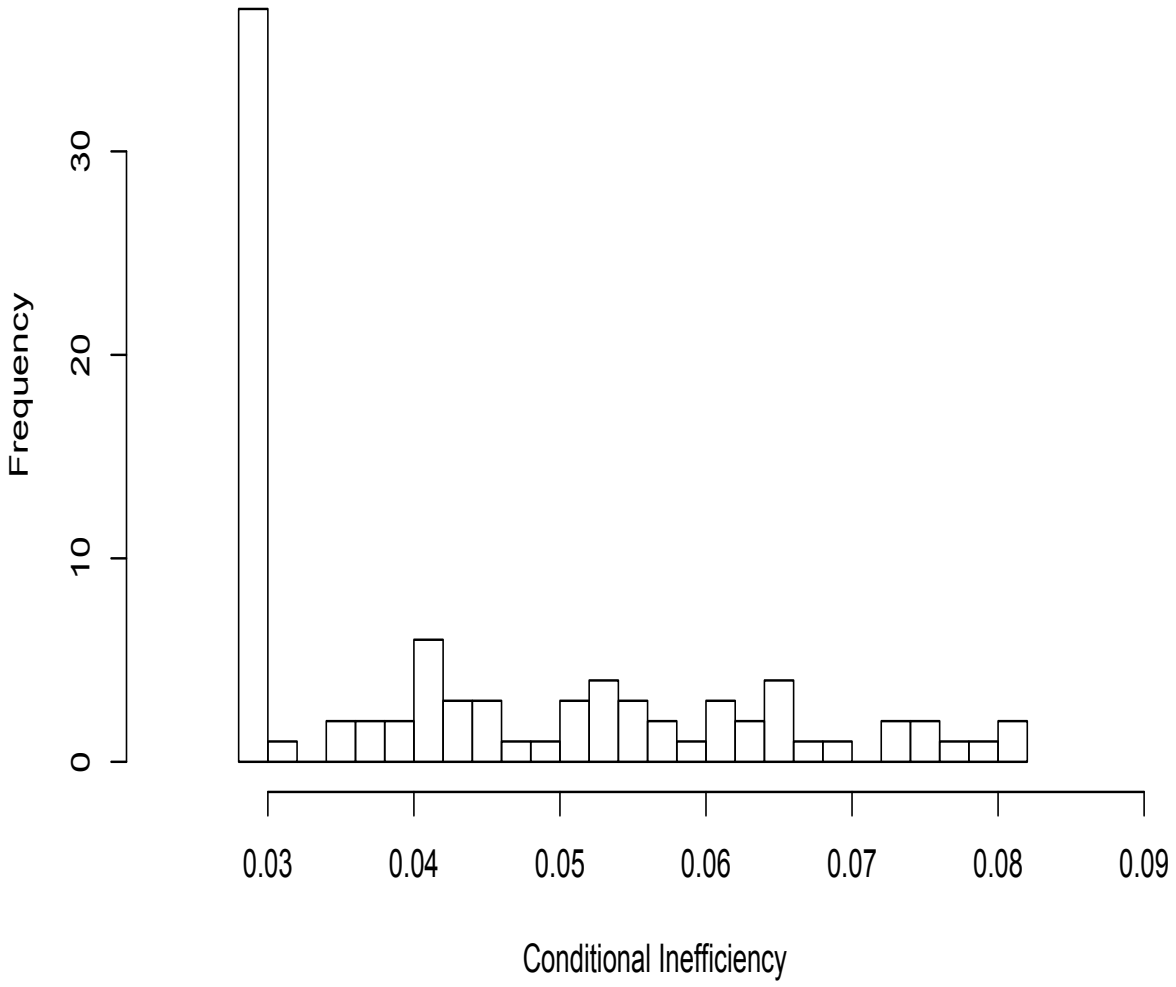
($\mu = 0$) version of the model.

The (final) Laplace-exponential model, produces significant estimates of $\hat{\gamma} = 0.086$ and $\hat{\theta} = 0.043$. Residuals for the Laplace-exponential model were used to calculate the conditional mean of inefficiency, $E[u|\varepsilon_i]$, in Equation 6 (with $\mu_* = 0$ and $\varepsilon_i = \hat{\varepsilon}_i$) for each of the 90 observations. Twenty-six of the 90 observations had positive residuals, and constant minimal values of the conditional mean equal to 0.0287. The maximal value of the conditional mean is 0.0820, the average value is 0.0438 and the median value is 0.0403. The distribution of conditional mean scores is in Figure 4. If we trim the 26 "efficient" firms (conditional mean equal to 0.02867), then the remaining 64 firms have an average conditional mean of 0.0500 and median of 0.0500.

5 Conclusions

The Laplace stochastic frontier model performs well under misspecification, produces non-zero estimates of the variance of inefficiency when the OLS and the LAD residuals have the wrong skew, and the constant conditional mean (when $\varepsilon \geq 0$) makes the model ideally suited for industries

Figure 4: Histogram of conditional inefficiency estimates for the Laplace-Exponential model.



with many firms near the efficient frontier. Interestingly, opponents of the stochastic frontier model claim that the models are just estimating noise as inefficiency; the proposed model seems to accommodate this notion by setting the conditional mean of inefficiency to a constant minimal

value when observations of the estimation error are large ($\varepsilon \geq 0$) and setting this mean to a non-minimal value when estimation error is small ($\varepsilon < 0$). In other words, the Laplace frontier model differentiates between "noisy results" and "less noisy results" in characterizing inefficiency.

Our simulation suggest that a Laplace error may be preferred to a normal error when the model is misspecified. There are other cases in the literature when a Laplace error may be preferred. Magnus, Powell and Prüfer (2009) use a Laplace prior (as opposed to a normal prior) because it has bounded risk. Meister (2004) shows that in a deconvolution setting, if one assumes that the error distribution is normal, when in fact it is Laplace, the loss of the density estimator is infinite, whereas in the reverse setting the loss is finite, implying that when one does not have concrete knowledge of the error distribution it is better to assume Laplace. It may be interesting to revisit inefficiency density deconvolution in the frontier model in the style of Horrace and Parmeter (2011), but with Laplace errors. This is currently being considered by the authors.

References

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

[2] Almanidis, P., J. Qian, and R.C. Sickles, 2014, Stochastic frontier models with bounded inefficiency, in Festschrift in Honor of Peter Schmidt Econometric Methods and Applications, eds. R. C. Sickles and W. C. Horrace, Springer, New York, 47-82.

[3] 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.

[4] 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, 153-170.

[5] Carree, M.A. 2002, Technological inefficiency and the skewness of the error component in stochastic frontier analysis. Economics Letters 77, 101-107.

[6] Cazals, C., J.P. Florens and L. Simar, 2002, Nonparametric frontier estimation: a robust approach. Journal of Econometrics 106, 1-25.

[7] Cuesta, R.A., 2000, A production model with firm-specific temporal variation in technical efficiency: with application to Spanish dairy farms. Journal of Productivity Analysis 13, 139-58.

[8] Fan, J., 1992, Deconvolution with supersmooth distributions. The Canadian Journal of Statistics 20, 155-169

[9] Feng, Q., Horrace, W.C., 2013, Alternative technical efficiency measures: skew, bias and scale. Journal of Applied Econometrics forthcoming.

[10] Flores-Lagunes A., W.C. Horrace and K.E. Schnier, 2007, Identifying technically efficient fishing vessels: a non-empty, minimal subset approach. Journal of Applied Econometrics 22, 729-745.

[11] Green, A. and D. Mayes,1991, Technical inefficiency in manufacturing industries. Economic Journal 101, 523-538.

[12] Green, W.H., 1995, LIMDEP Version 7.0 UserÕs Manual, New York: Econometric Software, Inc.

[13] Greene, W.H., 2005, Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of Econometrics 126, 269-303.

[14] Horrace, W.C., 2005, On ranking and selection from independent truncated normal distributions. Journal of Econometrics 126, 335-354.

[15] Horrace, W.C. and C.F. Parmeter, 2011, Semiparametric deconvolution with unknown error variance. Journal of Productivity Analysis 35, 129-141.

[16] Horrace W.C., S. Richards-Shubik and I. A. Wright, 2014, Expected efficiency ranks from parametric stochastic frontier models. Empirical Economics, forthcoming.

[17] Horrace, W.C. and P. Schmidt, 1996, Confidence statements for efficiency estimates from stochastic frontier models. Journal of Productivity Analysis 7, 257-82.

[18] Horrace, W.C. and P. Schmidt, 2000, Multiple comparisons with the best, with economic applications. Journal of Applied Econometrics 15, 1-26.

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

[20] Kim, Y. and P. Schmidt, 2008, Marginal comparisons with the best and the efficiency measurement problem. Journal of Business & Economic Statistics 26, 253-260.

[21] Koenker, R., 2005, Quantile Regression, Econometric Society Monographs, Cambridge University Press.

[22] Kumbhakar, S.C. and C.A.K. Lovell, 2000, Stochastic Frontier Analysis, Cambridge University Press, Cambridge.

[23] Kumbhakar, S.C., C.F. Parmeter and E.G. Tsionas, 2013, A zero inefficiency stochastic frontier model. Journal of Econometrics 172, 66-76

[24] Magnus, J.R., and O. Powell and P. Prüfer, 2010, A comparison of two model averaging techniques with an application to growth empirics. Journal of Econometrics 154, 139-153.

[25] Meister, A., 2004, On the effect of misspecifying the error density in a deconvolution problem. The Canadian Journal of Statistics 32, 439-449.

[26] Nguyen, N. B. 2010, Estimation of Technical Efficiency in Stochastic Frontier Analysis. Doctoral Dissertation, Bowling Green State University.

[27] Olson J.A., P. Schmidt, D.M. Waldman, 1980, A Monte Carlo study of estimators of stochastic frontier production functions. Journal of Econometrics 13, 67-82.

[28] Rho S., P. Schmidt, 2013, Are All Firms Inefficient? Journal of Productivity Analysis forthcoming.

[29] Simar, L. and P.W. Wilson, 2009, Inferences from cross-sectional, stochastic frontier models. Econometric Reviews 29, 62-98.

[30] Storn, R. and K. Price 1997, Differential Evolution – A simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization 115, 341-359.

[31] Timmer, C.P., 1971, Using a probabilistic frontier production function to measure technical efficiency. Journal of Political Economy 79, 776-794.

[32] Tsionas, E.G., 2007, Efficiency measurement with the Weibull stochastic frontier. Oxford Bulletin of Economics and Statistics 69, 693-706.

[33] Waldman, D.M., 1982, A stationary point for the stochastic frontier likelihood, Journal of Econometrics 18, 275-279.

[34] Wheat P., W.H. Greene and A.S.J. Smith, 2013, Understanding prediction intervals for firm specific inefficiency scores from parametric. Journal of Productivity Analysis forthcoming.

[35] Wilson, P.W., 1993, Detecting outliers in deterministic nonparametric frontier models with multiple outputs. Journal of Business & Economic Statistics 11, 319-323.

A Derivations for $\gamma = \theta$

Even though λ_- does not exist when $\gamma = \theta$, it is a continuity point in line 2 of equation (5). To see this rewrite the bracketed term in in the second line of equation (5) as $\lambda_+ e^{\varepsilon/\theta} + \lambda_- \left(e^{\varepsilon/\theta} - e^{\varepsilon/\gamma}\right)$. As $\theta \to \gamma$, the term $\left(e^{\varepsilon/\theta} - e^{\varepsilon/\gamma}\right) \to 0$ faster than $\lambda_- \to \infty$. Application of L'Hopital's rule shows that as $\theta \to \gamma$, $\lambda_- \left(e^{\varepsilon/\theta} - e^{\varepsilon/\gamma}\right) \to -\varepsilon e^{\varepsilon/\gamma}$, and the third line in equation (5) results. For completeness we provide results for the case where $\theta = \gamma$ for the most general Laplace-truncated Laplace model (equations (8), (9) and (11)):

$$f_\varepsilon(\varepsilon) = \begin{cases} \frac{c(\mu_*)}{4\gamma\theta}\left(\mu_* + \lambda_+\right) e^{-(\mu_* + \varepsilon)/\gamma}, & \varepsilon \geq 0,\ \theta = \gamma, \\ \frac{c(\mu_*)}{4\gamma\theta}\left[2\lambda_+ e^{-|\mu_* + \varepsilon|/\gamma} - \lambda_+ e^{-(\mu_* - \varepsilon)/\gamma} + |\mu_* + \varepsilon| e^{-|\mu_* + \varepsilon|/\gamma}\right] & \varepsilon < 0,\ \theta = \gamma \end{cases},$$

and

$$E(u|\varepsilon) = \begin{cases} \frac{c(\mu_*)}{4\gamma\theta f_\varepsilon(0)}\left(0.5\mu_*^2 + \mu_*\lambda_+ + \lambda_+^2\right) e^{-\mu_*/\gamma} & \varepsilon \geq 0,\ \gamma = \theta \\ \frac{c(\mu_*)}{4\gamma\theta f_\varepsilon(\varepsilon)}\left(\left[0.5(\mu_*^2 - \varepsilon^2) + \lambda_+(\mu_* - \varepsilon)\right] e^{-(\mu_* + \varepsilon)/\gamma} + \lambda_+^2 e^{-(\mu_* - \varepsilon)/\gamma}\right) & \mu_* \geq -\varepsilon,\ \gamma = \theta \\ \frac{c(\mu_*)}{4\gamma\theta f_\varepsilon(\varepsilon)}\left(\left[0.5(\varepsilon^2 - \mu_*^2) + \lambda_+(\mu_* - \varepsilon)\right] e^{(\mu_* + \varepsilon)/\gamma} + \lambda_+^2 e^{-(\mu_* - \varepsilon)/\gamma}\right) & \mu_* < -\varepsilon,\ \gamma = \theta \end{cases}.$$

The conditional median results are

$$m[u|\varepsilon] = \begin{cases} 2\gamma\theta \frac{f_\varepsilon(0)}{c(\mu_*)} e^{\mu_*/\theta}, & \varepsilon \geq 0,\ m[u|\varepsilon] < \mu_*, \\ \gamma\theta f_\varepsilon(\varepsilon) e^{-\varepsilon/\gamma}, & \varepsilon < 0,\ m[u|\varepsilon] < -\varepsilon,\ \mu < 0 \\ 2\gamma^2 \frac{f_\varepsilon(\varepsilon)}{c(\mu_*)} e^{(\mu_* + \varepsilon)/\gamma} - \lambda_+(\frac{\gamma}{2} e^{2\varepsilon/\gamma} - 1) - \varepsilon, & \varepsilon < 0, -\varepsilon < m[u|\varepsilon] < \mu_*, \\ 2\gamma^2 \frac{f_\varepsilon(\varepsilon)}{c(\mu_*)} e^{-(\mu_* + \varepsilon)/\gamma} - \lambda_+(\frac{\gamma}{2} e^{-2\mu_*/\gamma} - 1) + \mu, & \varepsilon < 0, -\varepsilon \geq m[u|\varepsilon] \geq \mu_* \end{cases} \tag{15}$$

B Score and Hessian of Laplace-Exponential Likelihood Function

First, to condense on notation we first derive the first and second derivatives of (5), with respect to our unknown parameter. First, let

$$\ln A = \ln \left(\frac{e^{\varepsilon/\gamma} - e^{\varepsilon/\theta}}{\gamma - \theta} + \frac{e^{\varepsilon/\theta}}{\gamma + \theta} \right) \tag{16}$$

Doing so yields

$$\mathcal{L}_\gamma = \frac{\partial \ln A}{\partial \gamma} = \frac{4\theta\gamma^3 e^{\varepsilon/\theta} - (\theta + \gamma)^2 e^{\varepsilon/\gamma}(\gamma^2 + \gamma\varepsilon - \theta\varepsilon)}{\gamma^2(\gamma - \theta)(\theta + \gamma)\left((\theta + \gamma)e^{\varepsilon/\gamma} - 2\theta e^{\varepsilon/\theta}\right)},$$

$$\mathcal{L}_\theta = \frac{\partial \ln A}{\partial \theta} = \frac{\theta e^{\varepsilon/\gamma}(\gamma + \theta)^2 - 2e^{\varepsilon/\theta}\left(-\gamma^2\varepsilon + \theta^3 + \theta^2\varepsilon + \gamma^2\theta\right)}{\theta(\theta - \gamma)(\gamma + \theta)\left(2\theta e^{\varepsilon/\theta} - e^{\varepsilon/\gamma}(\gamma + \theta)\right)}$$

$$\mathcal{L}_{\gamma,\gamma} = \frac{\partial^2 \ln A}{\partial \gamma^2} = \frac{\frac{\varepsilon^2 e^{\varepsilon/\gamma}}{\gamma^4(\gamma-\theta)} + \frac{2\varepsilon e^{\varepsilon/\gamma}}{\gamma^3(\gamma-\theta)} + \frac{2\varepsilon e^{\varepsilon/\gamma}}{\gamma^2(\gamma-\theta)^2} + \frac{2(e^{\varepsilon/\gamma}-e^{\varepsilon/\theta})}{(\gamma-\theta)^3} + \frac{2e^{\varepsilon/\theta}}{(\theta+\gamma)^3}}{\frac{e^{\varepsilon/\gamma}-e^{\varepsilon/\theta}}{\gamma-\theta} + \frac{e^{\varepsilon/\theta}}{\theta+\gamma}} - \frac{\left(-\frac{\varepsilon e^{\varepsilon/\gamma}}{\gamma^2(\gamma-\theta)} - \frac{e^{\varepsilon/\gamma}-e^{\varepsilon/\theta}}{(\gamma-\theta)^2} - \frac{e^{\varepsilon/\theta}}{(\theta+\gamma)^2}\right)^2}{\frac{e^{\varepsilon/\gamma}-e^{\varepsilon/\theta}}{\gamma-\theta} + \frac{e^{\varepsilon/\theta}}{(\theta+\gamma)^2}},$$

and

$$\begin{aligned}\mathcal{L}_{\theta,\theta} = \frac{\partial^2 \ln A}{\partial \theta^2} = \; & \frac{-\varepsilon^2 e^{\varepsilon/\theta}}{\theta^4(\gamma-\theta)} + \frac{\varepsilon^2 e^{\varepsilon/\theta}}{\theta^4(\gamma+\theta)} - \frac{2\varepsilon e^{\varepsilon/\theta}}{\theta^3(\gamma-\theta)} + \frac{2\varepsilon e^{\varepsilon/\theta}}{\theta^3(\gamma+\theta)} \\ & + \frac{2\varepsilon e^{\varepsilon/\theta}}{\theta^2(\gamma-\theta)^2} + \frac{2\varepsilon e^{\varepsilon/\theta}}{\theta^2(\gamma+\theta)^2} + \frac{2(e^{\varepsilon/\gamma}-e^{\varepsilon/\theta})}{(\gamma-\theta)^3} + \frac{2e^{\varepsilon/\theta}}{(\gamma+\theta)^3}\Bigg)\Bigg(\frac{e^{\varepsilon/\gamma}-e^{\varepsilon/\theta}}{\gamma-\theta} + \frac{e^{\varepsilon/\theta}}{\gamma+\theta}\Bigg)\Bigg(\\ & - \frac{\varepsilon e^{\varepsilon/\theta}}{\theta^2(\gamma-\theta)} - \frac{\varepsilon e^{\varepsilon/\theta}}{\theta^2(\gamma+\theta)} + \frac{e^{\varepsilon/\gamma}-e^{\varepsilon/\theta}}{(\gamma-\theta)^2} - \frac{e^{\varepsilon/\theta}}{(\gamma+\theta)^2}\Bigg)^2 \Bigg/ \frac{e^{\varepsilon/\gamma}-e^{\varepsilon/\theta}}{\gamma-\theta} + \frac{e^{\varepsilon/\theta}}{\gamma+\theta}\Bigg)^2.\end{aligned}$$

To calculate $\mathcal{L}_{\gamma,\theta}$, we denote the numerator of $\mathcal{L}_\gamma$ as $\mathcal{L}_\gamma^N$ and the denominator as $\mathcal{L}_\gamma^D$

$$\mathcal{L}_{\gamma,\theta} = \frac{\partial^2 \ln A}{\partial \gamma \partial \theta} = \frac{\mathcal{L}_\gamma^D \frac{\partial \mathcal{L}_\gamma^N}{\partial \theta} - \mathcal{L}_\gamma^N \frac{\partial \mathcal{L}_\gamma^D}{\partial \theta}}{\left(\mathcal{L}_\gamma^D\right)^2},$$

where

$$\frac{\partial \mathcal{L}_\gamma^N}{\partial \theta} = \gamma^{-2}\left(4\gamma^3 e^{\varepsilon/\theta}(\varepsilon-\theta) - \theta(\theta+\gamma)(-2\gamma^2+\gamma\varepsilon+\theta\varepsilon)e^{\varepsilon/\gamma}\right)$$

and

$$\frac{\partial \mathcal{L}_\gamma^D}{\partial \theta} = \gamma^{-2}\theta\Big((\theta+\gamma)\left(3\gamma^3 - \gamma^2(\theta+\varepsilon) + \theta^2\varepsilon\right)e^{\varepsilon/\gamma} - 4\theta\gamma^3 e^{\varepsilon/\theta}\Big)\Big($$

Lastly, note that

$$\mathcal{L}_\beta = \frac{\partial \ln A}{\partial \beta} = A^{-1}\frac{\partial A}{\partial \beta}$$

where

$$\frac{\partial A}{\partial \beta} = -\frac{e^{\varepsilon/\gamma}\frac{x}{\gamma} - e^{\varepsilon/\theta}\frac{x}{\theta}}{\gamma-\theta} - \frac{e^{\varepsilon/\theta}\frac{x}{\theta}}{\gamma+\theta} = A_\beta$$

The score vector of the Laplace-Exponential likelihood function is

$$S(\Theta) = \begin{bmatrix} \gamma^{-1}\sum\limits_{i\in\{\varepsilon_i>0\}} x_i + \sum\limits_{i\in\{\varepsilon_i<0\}} \mathcal{L}_{\beta,i} \\ -n_1(\gamma+\theta)^{-1} + \gamma^{-2}\sum\limits_{i\in\{\varepsilon_i>0\}} \varepsilon_i + \sum\limits_{i\in\{\varepsilon_i<0\}} \mathcal{L}_{\gamma,i} \\ -n_1(\gamma+\theta)^{-1} + \sum\limits_{i\in\{\varepsilon_i<0\}} \mathcal{L}_{\theta,i} \end{bmatrix}, \tag{17}$$

where $\Theta = (\gamma, \theta, \beta)$. The Hessian is

$$H(\Theta) = \begin{bmatrix} \sum\limits_{i\in\{\varepsilon_i<0\}} A_i^{-1}\left(A_{\beta,\beta,i} - A_i^{-1}A_{\beta,i}A'_{\beta,i}\right) & \sum\limits_{i\in\{\varepsilon_i<0\}} A_i^{-1}\left(A_{\beta,\gamma,i} - A_i^{-1}A_{\beta,i}A_{\gamma,i}\right) & \sum\limits_{i\in\{\varepsilon_i<0\}} A_i^{-1}\left(A_{\beta,\theta,i} - A_i^{-1}A_{\beta,i}A_{\theta,i}\right) \\ \sum\limits_{i\in\{\varepsilon_i<0\}} A_i^{-1}\left(A_{\beta,\gamma,i} - A_i^{-1}A_{\beta,i}A_{\gamma,i}\right) & n_1(\gamma+\theta)^{-2} - 2\gamma^{-3}\sum\limits_{i\in\{\varepsilon_i>0\}} \varepsilon_i + \sum\limits_{i\in\{\varepsilon_i<0\}} \mathcal{L}_{\gamma,\gamma,i} & n_1(\gamma+\theta)^{-2} + \sum\limits_{i\in\{\varepsilon_i<0\}} \mathcal{L}_{\gamma,\theta,i} \\ \sum\limits_{i\in\{\varepsilon_i<0\}} A_i^{-1}\left(A_{\beta,\theta,i} - A_i^{-1}A_{\beta,i}A_{\theta,i}\right) & n_1(\gamma+\theta)^{-2} + \sum\limits_{i\in\{\varepsilon_i<0\}} \mathcal{L}_{\gamma,\theta,i} & n_1(\gamma+\theta)^{-2} + \sum\limits_{i\in\{\varepsilon_i<0\}} \mathcal{L}_{\gamma,\gamma,i} \end{bmatrix}.$$

We have also used the notation

$$A_\gamma = e^{\varepsilon/\theta}\left((\gamma-\theta)^{-2}-(\gamma+\theta)^{-2}\right) - e^{\varepsilon/\gamma}(\gamma-\theta)^{-1}\left[\frac{\varepsilon}{\gamma^2}+(\gamma-\theta)^{-1}\right]$$

$$A_\theta = e^{\varepsilon/\theta}(\gamma-\theta)^{-1}\left[\frac{\varepsilon}{\theta^2}+(\gamma-\theta)^{-1}\right] - e^{\varepsilon/\theta}\left[(\gamma-\theta)^{-2}+(\gamma+\theta)^{-2}+(\gamma+\theta)^{-1}\frac{\varepsilon}{\theta^2}\right],$$

$$A_{\beta\gamma} = (\gamma-\theta)^{-2}\left[e^{\varepsilon/\gamma}\frac{x}{\gamma}-e^{\varepsilon/\theta}\frac{x}{\theta}\right]+(\gamma+\theta)^{-2}e^{\varepsilon/\theta}\frac{x}{\theta}+(\gamma-\theta)^{-1}e^{\varepsilon/\gamma}\frac{x}{\gamma^2}\left[\frac{\varepsilon}{\gamma}+1\right]$$

$$A_{\beta\theta} =(\gamma-\theta)^{-2}\left[e^{\varepsilon/\gamma}\frac{x}{\gamma}-e^{\varepsilon/\theta}\frac{x}{\theta}\right]+(\gamma+\theta)^{-2}e^{\varepsilon/\theta}\frac{x}{\theta}+\left[(\gamma+\theta)^{-1}-(\gamma-\theta)^{-1}\right]e^{\varepsilon/\theta}\frac{x}{\theta^2}\left[\frac{\varepsilon}{\theta}+1\right]$$