

Stationary Points for Parametric Stochastic Frontier Models

William C. Horrace* Ian A. Wright†

August 25, 2016

Abstract

The results of Waldman (1982) on the Normal-Half Normal stochastic frontier model are generalized using the theory of the Dirac delta (Dirac, 1930), and distribution-free conditions are established to ensure a stationary point in the likelihood as the variance of the inefficiency distribution goes to zero. Stability of the stationary point and "wrong skew" results are derived or simulated for common parametric assumptions on the model. Identification is discussed.

Keywords: *Inefficiency Estimation; Ordinary Least Squares; Singular Distribution; Dirac Delta, Generalized Function*

JEL Codes: C21, D24

Preliminary and incomplete. Do not cite without author permission

*Department of Economics, Syracuse University, Syracuse, NY, 13244-1020. whorrace@maxwell.syr.edu.

†Corresponding author, Department of Economics, Northeastern University, Boston, MA 02115-5000. i.wright@northeastern.edu.

1 Introduction

Parametric Stochastic Frontier Analysis (SFA) allows for production function estimation while accounting for inefficiency in a cross-section of firms.¹ Specifically, SFA models a firm's output as a function of its inputs plus a random error. Aigner, Lovell and Schmidt (1977), hereafter ALS, specify a composed error $\varepsilon = v - u$, where v represents random fluctuations in the production frontier and where $u \geq 0$ (independent of v) represents random inefficiency. Typically, u is called "signal," v is called "noise," and the model is parameterized in terms of a "signal to noise" ratio of the variance components. Estimation proceeds by making distributional assumptions on the error components and calculating (or searching for) the maximum likelihood estimator (MLE). ALS specify a normal distribution for noise, $v \sim N(0, \sigma_v^2)$, a half normal distribution for signal, $u \sim |N(0, \sigma_u^2)|$, so that the "Normal-Half Normal" stochastic frontier model (the N-HN model) has signal to noise ratio, $\lambda = \frac{\sigma_u}{\sigma_v}$.² The Half Normal (HN) specification for u implies that its skew is positive so that skew of ε is negative. However, in practice it often happens that the skew of the Ordinary Least Squares (OLS) residuals is positive, which implies that the Maximum Likelihood Estimator (MLE) of σ_u is zero in the N-HN model. This is called the "wrong skew" problem, and all rigorous treatments of the issue in the literature have been for the N-HN specification. See for example, Olson, Schmidt and Waldman (1980), Waldman (1982), Simar and Wilson (2010) and Feng, Horrace and Wu (2013).³

Waldman (1982) analyzes the wrong skew problem for the N-HN model, showing that: (1) OLS is a stationary point in the parameter space of the likelihood, and (2) a sufficient condition for OLS ($\sigma_u = 0$) to be a local maximum (a stable solution) is that the sign of the skew of the OLS residuals is positive. Therefore, if we *a priori* believe that there is inefficiency in the population of firms, wrong skew of the OLS residuals may be problematic. Theoretically, wrong skew of the OLS residuals creates problems for inference,

¹This paper is concerned with production function estimation, but the analysis can be applied to cost functions as well.

²ALS also consider an exponential distributional assumption on the inefficiency distribution, leading to a Normal-Exponential model. In the N-HN model the variance of the pretruncated distribution of inefficiency is σ_u^2 . The variance of the post-truncated distribution is $V(u) \neq \sigma_u$. The distinction is important in what follows.

³Almanidis and Sickles (2012) examine the behavior of the likelihood for the Normal-Doubly Truncated Normal specification to understand how it produces a MLE that is asymptotically efficient relative to OLS regardless of the skew of the OLS residuals. Our primary focus is understanding the model under more general assumptions on the error components.

because the Hessian of the likelihood is singular. Empirical "solutions" to the wrong skew problem include pulling another sample (which is often not practical) or re-specifying the distribution of inefficiency.

In his analysis, Waldman exploits the "signal to noise" parameterization of the N-HN model by setting $\lambda = 0$ in the likelihood and in the first-order conditions (F.O.C.) of the maximization problem. In the N-HN case the likelihood is finite at $\lambda = 0$, and it reduces to that of OLS, making the analysis tractable. A problem with this approach is that when $\sigma_u = 0$ the distribution of inefficiency is singular. It so happens that the singularity in the N-HN model is not problematic in determining the behavior of the likelihood in the neighborhood OLS. This fortuitous outcome does not occur in general, and it is entirely possible that, when $\sigma_u = 0$, the likelihood and its F.O.C. are undefined for other parametric specifications of the model. Therefore, we need to appeal to a more general theory that examines the limiting behavior of the likelihood function, based on a singular distribution for inefficiency and an unspecified distribution for noise. This "distribution free" analysis is the primary contribution of the paper.

We exploit the so-called "sifting" property of the Dirac measure (Dirac, 1930) to examine the model under very general assumptions on the inefficiency and noise distributions.⁴ We show that a stationary point exists, as long as the inefficiency distribution can be represented as a "delta sequence" that converges to a Dirac delta located at the origin as its variance shrinks. Therefore, our analysis generalizes the "N-HN stationary point" result of Waldman (1982) without any distributional assumptions on the error components. In particular, we show that under weak assumptions the likelihood and its F.O.C. reduce to those based solely on the noise distribution. For example, if the noise distribution is zero-mean normal, then OLS is a stationary point for *any* specification of the inefficiency distribution that is continuous and converges to a Dirac delta at the origin. If the resulting Dirac delta is not located at the origin, then MLE is not identified. This may suggest that empiricists restrict their distributional choices for u to classes of distributions that have a Dirac delta representation located at the origin. The half normal, exponential and gamma distributions

⁴See Kobayashi (1991, 2008), Kobayashi and Shi (2005), Frieden (1983) and Arley and Buch (1950) for modern treatments of the Dirac delta.

possess this feature, as do the truncated normal (TN) and doubly truncated normal distributions (DTN), when the pre-truncated mean (μ) is non-positive. When the pre-truncated mean is positive, the Dirac delta is not centered at the origin so MLE is not identified as $\sigma_u \rightarrow 0$. That is, the location of the singularity is an unknown parameter, but there is zero variation to identify it. This is, perhaps, further evidence that the pre-truncated mean of inefficiency is only weakly identified in these models (Almanidis, Qian and Sickles, 2014, p64).

Given the existence of the distribution-free stationary point, we then explore its stability for common parameterizations on the distribution of inefficiency by examining the behavior of the Hessian in the neighborhood of the stationary point. We consider the following cases: the Normal-Truncated Normal model (N-TN) due to Stevenson (1980), the Normal-Exponential model (N-E) due to ALS and Meeusen and Van den Broeck (1977), and the Normal-Doubly Truncated Normal model (N-DTN) due to Almanidis, Qian and Sickles (2014).⁵ Our results show that in all cases, the stationary point is the OLS solution (due to normality of the noise distribution), and in most cases the solution is stable (i.e., it is a local maximum on the $\sigma_u = 0$ edge of the parameter space). We also explore the wrong skewness issue in each case. For the N-TN and N-DTN models with a non-positive pre-truncated mean of inefficiency, OLS is a stable stationary point when the OLS residuals have the wrong skew: the usual Waldman result. In particular, when the pre-truncated mean is non-positive, the behavior of the likelihood in the neighborhood of OLS for these models is identical to that of the N-HN model as $\sigma_u \rightarrow 0$. Furthermore, the OLS intercept is identified in the sense that the expectation of u goes to zero as its variance shrinks.⁶ When the pre-truncated mean is positive and σ_u goes to 0, all firms in the sample are inefficiency and their individual "draws" from the inefficiency "distribution" are identical. Perhaps this is a wholly unrealistic situation which argues for inefficiency distributions that collapse to the origin, implying that all firms are efficient.

⁵General results on the stability of the stationary point are not forthcoming (as we shall see), so we focus on the traditional parametric assumptions for our stability analysis.

⁶In the DTN case, when the pre-truncated mean is non-positive the skew of the distribution is positive. When the pre-truncated mean is positive, there is scope for the distribution to have negative skew, however, the model is not identified as the variance of inefficiency goes to zero, therefore, we do not discuss this case.

For the N-E model, we establish stability of OLS, but cannot establish a theoretical relationship between its stability and the skew of its residuals. However, simulations suggest that the wrong skew is indeed a sufficient condition for stability.

Generally speaking, as the variance of the inefficiency distribution goes to zero all of the aforementioned models are observationally equivalent (up to the location of the mass point of the resulting Dirac delta), and OLS becomes the local maximizer of the likelihood when the noise distribution is normal. This may have additional implications for empiricists. Our results suggest that re-specifying the inefficiency distribution (when the wrong skew arises) may be a fruitless exercise when the population variance of inefficiency is very small or zero. There may not be enough signal from the inefficiency distribution to identify its parameters regardless of its specification, even when the sample size is fairly large. In this sense, even the DTN distribution may not be immune to the wrong skew problem. The N-DTN model can accommodate the wrong skew, but it may be impossible to estimate its parameters when the wrong skew arises and the population variance of inefficiency is small or zero.

These nuances of empirical implementations of parametric SFA underscore the difficulties of the implicit "signal-to-noise" deconvolution problem that the composed error model presents.⁷ These difficulties are exacerbated when there is only a cross-section of data to aid in estimating the model's parameters. Our findings suggest that when faced with incorrectly skewed OLS residuals (particularly in large samples) empiricists should either admit that inefficiency does not exist in the population or (if another sample is not available) use the inferential procedures of Simar and Wilson (2010).

The paper is organized as follows. The next section establishes the stationary point. Section 3 provides theoretical stability results for the N-E model and for the N-TN and N-DTN models when the pre-truncated mean is non-positive. In the latter two cases we establish the relationship between the skew of the OLS residuals and the stability of OLS. Section 4 provides simulated evidence of the stationary point for each

⁷See Horrace and Parmeter (2011) for a discussion of deconvolution in cross-sectional SFA.

model. Section 5 summarizes and concludes.

2 Limiting Behavior of the Likelihood Function

The cross-sectional stochastic frontier model of ALS (1977) is:

$$y_i = x_i' \beta + \varepsilon_i, \quad i = 1, \dots, n \quad (1)$$

where y_i is a single output (typically in logarithms), x_i is a $k \times 1$ vector of inputs with first element equal to 1 for all i , β is a $k \times 1$ vector of unknown parameters, and $\varepsilon_i = v_i - u_i$ represents random shocks to the production process. The v_i are random fluctuations in the production frontier for each firm i , and the u_i are random inefficiency draws for each firm i . Without specific distributional assumptions on the error terms, the basic assumptions of the ALS model are:

Assumption 1 v_i and u_i are independent random variables.

Assumption 2 v_i and u_i are independent of random variable x_i , and $n^{-1} \sum_i x_i' x_i$ is a positive definite matrix.

Assumption 3 $v_i \in \mathbb{R}$ has zero-mean probability density $f_v(v, \sigma_v)$ has continuous second derivatives with respect to v and unknown scale parameter $\sigma_v > 0$.

Assumption 4 $u_i \geq 0$ has probability density $f_u(u, \sigma_u)$, continuous in both u and unknown scale parameter $\sigma_u > 0$.

These are generally accepted assumptions throughout the literature regardless of the parametric form of the distributions in Assumptions 3 and 4. The density of u may contain other parameters with compact

support that we suppress for know.⁸ Given these assumptions, the composed error has continuous density:

$$f_\varepsilon(\varepsilon) = \int_0^\infty f_v(\varepsilon + u, \sigma_v) f_u(u, \sigma_u) du.$$

The likelihood function is:

$$L(y, x, \beta, \sigma_v, \sigma_u) = \prod_{i=1}^n f_\varepsilon(y_i - x'_i \beta) = \prod_{i=1}^n \int_0^\infty f_v(y_i - x'_i \beta + u_i, \sigma_v) f_u(u_i, \sigma_u) du_i. \quad (2)$$

Interest centers on the behavior of the likelihood function at the point $\sigma_u = 0$. In general $f_u(u, 0)$ may not be well-defined, so plugging $\sigma_u = 0$ into the likelihood may not always be feasible in understanding this behavior.⁹ Therefore, to understand the likelihood one must consider its behavior as $\sigma_u \rightarrow 0$. Based on Assumptions 1-4, we have:

$$\lim_{\sigma_u \rightarrow 0} L(y, x, \beta, \sigma_v, \sigma_u) = \prod_{i=1}^n \lim_{\sigma_u \rightarrow 0} \int_0^\infty f_v(y_i - x'_i \beta + u_i, \sigma_v) f_u(u_i, \sigma_u) du_i. \quad (3)$$

To understand the limiting behavior of the likelihood is to understand the limiting behavior of the product of the integrals on the RHS of equation 3, which is governed by the limiting behavior of f_u under the product of n integrals. The challenge is finding conditions on the integrands to allow the limit to being interchanged with the integral operations. The problem is that the limiting behavior of the integrands in equation 3 does not lend itself to the usual pointwise convergence arguments, because as $\sigma_u \rightarrow 0$, the distribution f_u is singular for the usual specifications of inefficiency.¹⁰ Instead, we must appeal to the theory of the Dirac

⁸For example, in ALS the distribution of u is known up to σ_u and can either be half normal, $u \sim |N(0, \sigma_u^2)|$, or exponential, $f_u(u, \sigma_u) = \sigma_u^{-1} e^{-u/\sigma_u}$, leading to the N-HN and the N-E models, respectively. The N-TN and N-DTN models have additional unknown parameters in the density of u and the likelihood function. Although the additional parameters make for a much richer class of models, they make estimation more difficult in general. Moreover, these additional parameters are not identified when $\sigma_u \rightarrow 0$.

⁹This is certainly the case for the N-TN and the N-DTN models (when the pre-truncated mean is negative), and for the N-E model.

¹⁰For probability densities that have Lebesgue measure, the dominated convergence theorem can be used to allow the interchanging of the limit and the integral. However, if the sequence of probability densities converge to a Dirac delta, then dominated convergence is not applicable.

delta.

Initially considered by Cauchy (1815), popularized by Dirac (1930) and formalized by Schwartz (1957), the Dirac delta has been used extensively in physics and engineering. Heuristically, the Dirac delta is the symmetric "function": $\delta(t) = 0$ for $t \neq 0$, but it is infinite at $t = 0$, while satisfying the distributional property $\int_{-\infty}^{\infty} \delta(t) dt = 1$. It is everywhere zero except for a singularity at $t = 0$, yet the area under the curve is unity. This apparent contradiction arises because the Dirac delta is not a function *per se*, but an equivalence class of functions that serves as the representation of a limiting process that is useful under the Riemann integral.¹¹ To this end, define the sequence of regular distribution functions on \mathbb{R} :

$$h_m(t) \equiv mh(mt), m = 1, 2, 3, \dots, \quad (4)$$

where h is an ordinary function satisfying:

$$\int_{-\infty}^{\infty} h(t) dt = 1.$$

If $h_m(t)$ converges in distribution to the singular distribution $\delta(t)$ as $m \rightarrow \infty$, then h_m is called a "delta sequence." If h_m is a delta sequence, and $g(t)$ is any arbitrary function that is continuous in the neighborhood of $t = 0$, then we have the so-called "sifting" property under the Riemann integral:

$$\lim_{m \rightarrow \infty} \int_{-\infty}^{\infty} g(t) h_m(t) dt = g(0). \quad (5)$$

(For a proof of equation 5 see Hoskins and Pinto, 2005, p.64) That is, under very general conditions on the integrand, we can exchange the limit and integration operations in equation 5, returning the function g evaluated at 0. Therefore, if f_u can be written as a delta sequence in equation 4, then equation 5 will

¹¹Technically it is a *generalized function*. See Frieden (1983) or Arley and Buch (1950) for a measure theoretic definition of the Dirac delta.

hold, and it may be used to show that the limit of the likelihood in equation 3 is strictly a function of $f_v(y_i - x_i'\beta + u_i, \sigma_v)$ evaluated as $u_i = 0$. The only complication is establishing an appropriate mapping between the index m (in equation 4) and σ_u (in equation 3). Fortunately, this is not difficult to do as both m and σ_u are scale parameters in the sequence of distributions, so limiting behavior under $m \rightarrow \infty$ or $\sigma_u \rightarrow 0$ will be identical. To this end, define the sequence of non-negative constants $\{\sigma_{u,m}\}_{m=1,2,\dots}$ converging to 0, then we require the assumption:

Assumption 5 *The sequence of distributions $f_{u,m}(u, \sigma_u) \equiv mf_u(mu, \sigma_u)$ $m = 1, 2, 3, \dots$ is a delta sequence, such that*

$$\lim_{m \rightarrow \infty} f_{u,m}(u, \sigma_u) = \lim_{m \rightarrow \infty} mf_u(mu, \sigma_u) = \lim_{\sigma_{u,m} \rightarrow 0} f_u(u, \sigma_{u,m})$$

is a Dirac delta centered at the origin.

Assumption 5 restricts f_u to a class of non-negative functions that behave identically as either $m \rightarrow \infty$ in $f_{u,m}(u, \sigma_u)$ or as $\sigma_{u,m} \rightarrow 0$ in $f_u(u, \sigma_{u,m})$. Not surprisingly, Assumption 5 will hold if the inefficiency distribution belongs to a class of scalable distributions satisfying:

$$f_u(u, \sigma_u) = f_u^*(u/\sigma_u)/\sigma_u, \tag{6}$$

where f_u^* is the "standardized" version of f_u , in which case Assumption 5 holds trivially. Important single-parameter families of distributions satisfy 6 and (hence) Assumption 5, including the half-normal and exponential distributions. For example, the half-normal family of distributions is:

$$f_u^{HN}(u, \sigma_u) = \frac{1}{\sigma_u} \sqrt{\frac{2}{\pi}} \exp\left[-\frac{u^2}{2\sigma_u^2}\right],$$

and its delta sequence representation is simply:

$$f_u^{HN}(u, \sigma_{u,m}) = \frac{1}{\sigma_{u,m}} \sqrt{\frac{2}{\pi}} \exp\left[-\frac{u^2}{2\sigma_{u,m}^2}\right], m = 1, 2, 3\dots$$

which converges in distribution to a Dirac delta at the origin. To see this, notice that for every $u \neq 0$ the exponent term converges to zero faster than the term $1/\sigma_{u,m} \rightarrow \infty$, while at $u = 0$ the exponent equals one, so $1/\sigma_{u,m} \rightarrow \infty$ causes the singularity at the origin. The exponential distribution has delta sequence representation:

$$f_u^E(u, \sigma_{u,m}) = \frac{1}{\sigma_{u,m}} \exp\left(-\frac{u}{\sigma_{u,m}}\right), m = 1, 2, 3\dots$$

which converges in distribution to a Dirac delta at the origin (using arguments similar to those above).

Therefore, Assumption 5 is satisfied with $\sigma_{u,m}$ substituted for σ_u in f_u^E .

Assumption 5, along with equation 5, can be used to handle the interchange of the limit and the integral in equation 3, leading to the following result.

Lemma 1 *Let $L_m \equiv L(y, x, \beta, \sigma_v, \sigma_{u,m})$, then under Assumptions 1-5:*

$$\lim_{\sigma_{u,m} \rightarrow 0} L_m \equiv L_\infty(y, x, \beta, \sigma_v) = \prod_{i=1}^n f_v(y_i - x_i' \beta, \sigma_v).$$

To prove this, apply the sifting property to equation 3. Lemma 1 states that as the sequence $\sigma_{u,m} \rightarrow 0$, we ignore $f_{u,m}$ in the likelihood of equation 3, leading to the likelihood, L_∞ , based solely on $f_v(y_i - x_i' \beta + u_i, \sigma_v)$, evaluated at $u_i = 0$. Therefore, the likelihood of equation 2 converges in distribution to the likelihood of the model in equation 1, evaluated at $u_i = 0$.¹² Not surprisingly, if the likelihood converges according to Lemma 1, then the limit of the F.O.C. of the likelihood in equation 3 will converge to the F.O.C. of the likelihood L_∞ . That is:

¹²Horrace and Parmeter (2016) study the limiting behavior of the characteristic function convolution, imply the likelihood results in Lemma 1. Understanding the behavior of the *density* convolution is more natural to consider here, as it provides a direct link to the likelihood and first-order and second-order conditions which are the objects of interest.

Lemma 2 *Under Assumptions 1-5, we have:*

$$\lim_{\sigma_{u,m} \rightarrow 0} \frac{\partial L_m}{\partial \beta} = \frac{\partial L_\infty}{\partial \beta} = \sum_{j=1}^n \frac{\partial f_v(y_j - x'_j \beta, \sigma_v)}{\partial \beta} \prod_{i \neq j} f_v(y_i - x'_i \beta, \sigma_v),$$

$$\lim_{\sigma_{u,m} \rightarrow 0} \frac{\partial L_m}{\partial \sigma_v^2} = \frac{\partial L_\infty}{\partial \sigma_v^2} = \sum_{j=1}^n \frac{\partial f_v(y_j - x'_j \beta, \sigma_v)}{\partial \sigma_v^2} \prod_{i \neq j} f_v(y_i - x'_i \beta, \sigma_v),$$

and the solutions to $\partial L_\infty / \partial \beta = 0$ and $\partial L_\infty / \partial \sigma_v^2 = 0$ are stationary points in the parameter space of the likelihood as $\sigma_{u,m} \rightarrow 0$.

To prove this, use Leibniz rule to pass the partial derivative with respect to β or σ_v through the definite integrals in equation 2 and apply the product rule to get (for example):

$$\frac{\partial L_m}{\partial \beta} = \sum_{j=1}^n \int_0^\infty \frac{\partial f_v(y_j - x'_j \beta + u_j, \sigma_v)}{\partial \beta} f_u(u_j, \sigma_{u,m}) du_j \prod_{i \neq j} \int_0^\infty f_v(y_i - x'_i \beta + u_i, \sigma_v) f_u(u_i, \sigma_{u,m}) du_i.$$

Since f_v and its first derivative are continuous at $u_i = 0$, we can applying the sifting property of equation 5 to obtain the desired result.

Lemma 2 generalizes the Waldman (1982) result to any convolution $f_v * f_u$ satisfying Assumptions 1-5. Obviously, as $\sigma_{u,m} \rightarrow 0$, the F.O.C. $\partial L_m / \partial \sigma_{u,m}$ vanishes, as the optimum is conditional on $\sigma_u = 0$. In this sense σ_u becomes a nuisance parameter in the optimization, and its F.O.C. is irrelevant. If f_u is a function of additional parameters (beyond σ_u), then those parameters must vanish in the limit to satisfy Assumption 5, causing the F.O.C. of equation 2 with respect to those additional parameters to be irrelevant. They are free parameters in the optimization as $\sigma_{u,m} \rightarrow 0$, as long as they do not violate Assumption 5. For example, if the limit of $f_u(u, \mu, \sigma_{u,m})$ (say) as $\sigma_{u,m} \rightarrow 0$ is a Dirac delta centered at $\mu \in \mathbb{R}$ (say), then Assumption 5 is violated, Lemmas 1 and 2 may not hold in general, and the F.O.C. with respect to μ will not vanish in the limit. The Lemma implies that different parametric specifications for the inefficiency distribution are

observationally equivalent as $\sigma_u \rightarrow 0$.¹³

Lemma 2 can be extended to higher-order partial derivatives with respect to the parameters of f_v (as long as the derivatives exist). For example, the second order conditions (S.O.C.) will have a similar property in the limit:

Lemma 3 *Under Assumptions 1-5 we have:*

$$\lim_{\sigma_{u,m} \rightarrow 0} \frac{\partial^2 L_m}{\partial \beta \partial \beta'} = \frac{\partial^2 L_\infty}{\partial \beta \partial \beta'}, \quad \lim_{\sigma_{u,m} \rightarrow 0} \frac{\partial^2 L_m}{\partial (\sigma_v^2)^2} = \frac{\partial^2 L_\infty}{\partial (\sigma_v^2)^2}, \quad \lim_{\sigma_{u,m} \rightarrow 0} \frac{\partial^2 L_m}{\partial \beta \partial \sigma_v} = \frac{\partial L_\infty}{\partial \beta \partial \sigma_v}.$$

The proof is similar to that of Lemma 2.¹⁴ Therefore, the limit of the Hessian of the likelihood in equation 2 (with respect to β and σ_v) is equal to the Hessian of the likelihood in Lemma 1 as $\sigma_{u,m} \rightarrow 0$. Therefore, if v is distributed $N(0, \sigma_v^2)$, then $\partial^2 \ln L_\infty / \partial \beta \partial \beta' = -\sum_i x_i x_i' / \sigma_v^2$, $\partial^2 \ln L_\infty / \partial (\sigma_v^2)^2 = -n/2\sigma_v^4$, and $\partial^2 \ln L_\infty / \partial \beta \partial \sigma_v = 0$, corresponding to the 2×2 Hessian of the OLS estimator. That is, let the complete Hessian from equation 2 be H , and let the submatrix of second partial derivatives and cross-partial for β and σ_v be H_{β, σ_v} , then under Assumptions 1-5, this submatrix converges to H_{OLS} , the negative semi-definite Hessian matrix for OLS:

$$\lim_{\sigma_{u,m} \rightarrow 0} H_{\beta, \sigma_v}(y, x, \beta, \sigma_v, \sigma_{u,m}) = H_{OLS} = \begin{bmatrix} -\sum_i x_i x_i' / \sigma_v^2 & \mathbf{0} \\ \mathbf{0}' & -n/2\sigma_v^4 \end{bmatrix},$$

where $\mathbf{0}$ is a column vector of zeros with length equal to the row dimension of x_i . This generalizes the Hessian results of Waldman (1982) for the 2×2 submatrix associated with β and σ_v .

The complete Hessian will have additional rows and columns corresponding to the second-order derivatives and cross-partial derivatives of the likelihood with respect to σ_u (or any additional parameters in

¹³See Rothenberg (1971) for a discussion on the concept of observational equivalence.

¹⁴In particular, any derivatives of the likelihood with respect to the parameters of f_v will be sums and products of integrals, whose integrands contain the products of derivatives of f_v convoluted with delta sequence, $f_{u,m}$, so that all those integrals converge to exactly those same derivatives of f_v evaluated at $u_i = 0$.

f_u that vanish as $\sigma_u \rightarrow 0$). Unfortunately, derivation of the limiting form of the complete Hessian is not possible under Assumptions 1-5. Derivatives of the likelihood with respect to any parameters of f_u cause the distribution to no longer be a delta sequence, and Assumption 5 is violated. Therefore, we derive these results for specific parametric assumptions on f_u in section 3. To that end we now discuss the limiting behavior of additional parametric forms.

2.1 Two-Parameter Families of Distributions

Things get only slightly more complicated when we move to a two-parameter inefficiency distribution, like the truncated normal (TN) family,

$$f_u^{TN}(u, \mu, \sigma_u) = \frac{\frac{1}{\sigma_u \sqrt{2\pi}} e^{-\left(\frac{u-\mu}{2\sigma_u}\right)^2}}{1 - \int_{-\infty}^0 \frac{1}{\sigma_u \sqrt{2\pi}} e^{-\left(\frac{u-\mu}{2\sigma_u}\right)^2} du} = \frac{\frac{1}{\sigma_u} \phi\left(\frac{u-\mu}{\sigma_u}\right)}{1 - \Phi\left(\frac{-\mu}{\sigma_u}\right)},$$

where ϕ and Φ are the density and distribution function (respectively) of a standard normal random variable and $\mu \in \mathbb{R}$. Unlike the HN and exponential families, the TN distribution does not satisfy 6 *per se*. However, its delta sequence representation is:

$$f_u^{TN}(u, \mu, \sigma_{u,m}) = \frac{\frac{1}{\sigma_{u,m}} \phi\left(\frac{u-\mu}{\sigma_{u,m}}\right)}{1 - \Phi\left(-\frac{\mu}{\sigma_{u,m}}\right)}. \quad (7)$$

In particular, in order for $f_{u,m}$ to be a sequence of distribution functions, we must scale the definite integral in the denominator to correspond to the scaling in the numerator, so that the area under each distribution in the sequence is unity.¹⁵

When $\mu > 0$, the denominator of equation 7 converges to 1 as $\sigma_{u,m} \rightarrow 0$, and can be ignored. Hence, the truncated normal distribution converges in distribution to a Dirac delta centered at $\mu > 0$ by the same arguments used to show that the half-normal distribution converges to a Dirac delta centered at *its*

¹⁵The problem is not that the TN distribution does not satisfy equation 6. The problem is our choice of notation for the density, which fails to capture the fact that the definite integral in the denominator must be scaled.

pre-truncated mean ($\mu = 0$). That is, in the truncated normal case we have:

$$\lim_{\sigma_{u,m} \rightarrow 0} f_u(u, \mu > 0, \sigma_{u,m})$$

converges to a Dirac delta, $\delta(u - \mu)$, centered at $\mu > 0$, so that Assumption 5 is violated and the likelihood converges in distribution to:

$$\begin{aligned} \lim_{\sigma_{u,m} \rightarrow 0} L(y, x, \beta, \sigma_v, \mu > 0, \sigma_{u,m}) &= \prod_{i=1}^n f_v(y_i - x_i' \beta + \mu, \sigma_v) \\ &\neq L_\infty. \end{aligned}$$

Therefore, when $\mu > 0$ and the x 's contain an intercept, the parameter μ is not identified as $\sigma_{u,m} \rightarrow 0$.

When $\mu < 0$, the denominator in equation 7 converges to 0 as $\sigma_{u,m} \rightarrow 0$. In this case, we use L'Hopital's rule to show that the delta sequence converges to a Dirac delta centered at the origin, implying that Assumption 5 and (hence) Lemmas 1-3 hold. See the Appendix for a proof that Assumption 5 is satisfied in this case. Therefore, when $\mu < 0$,

$$\begin{aligned} \lim_{\sigma_{u,m} \rightarrow 0} L(y, x, \beta, \sigma_v, \mu < 0, \sigma_{u,m}) &= \prod_{i=1}^n f_v(y_i - x_i' \beta, \sigma_v) \\ &= L_\infty. \end{aligned}$$

In the limit, the likelihood is no longer a function of $\mu < 0$. However, in practice the likelihood is still a function of μ , so its estimator will not be identified as $\sigma_{u,m} \rightarrow 0$.

Greene (1990) considers the gamma distributed stochastic frontier model, where

$$f_u(u, k, \sigma_u) = \frac{\frac{1}{\sigma_u^k} u^{k-1} e^{-u/\sigma_u}}{\Gamma(k)},$$

with delta sequence representation:

$$f_u(u, k, \sigma_{u,m}) = \frac{\left(\frac{1}{\sigma_{u,m}}\right)^k u^{k-1} e^{-u/\sigma_{u,m}}}{\Gamma(k)},$$

and Assumption 5 holds using similar arguments.

2.2 Three-Parameter Families of Distributions

The only three-parameter family commonly considered in the stochastic frontier literature is the doubly-truncated normal (DTN):

$$f_u^{DTN}(u, \mu, \sigma_u, B) = \frac{\frac{1}{\sigma_u} \phi\left(\frac{u-\mu}{\sigma_u}\right)}{\Phi\left(\frac{B-\mu}{\sigma_u}\right) - \Phi\left(\frac{-\mu}{\sigma_u}\right)}; \quad u \in [0, B], \quad B > 0. \quad (8)$$

See Almanidis, Qian and Sickles (2014). Here, the delta sequence:

$$f_u^{DTN}(u, \mu, \sigma_{u,m}, B) = \frac{\frac{1}{\sigma_{u,m}} \phi\left(\frac{u-\mu}{\sigma_{u,m}}\right)}{\Phi\left(\frac{B-\mu}{\sigma_{u,m}}\right) - \Phi\left(-\frac{\mu}{\sigma_{u,m}}\right)} \quad (9)$$

behaves like that of the truncated normal with no complications, because the limit of the denominator is equal to 1. As in the truncated normal case of equation 7, when $\mu > 0$ the doubly truncated normal sequence converges to a Dirac delta centered at the minimum of the pre-truncated mean and the upper bound ($\min[\mu, B]$). Otherwise it converges to a Dirac delta centered at zero. See the Appendix for a proof. In either case, neither the μ nor B estimates are identified.

2.3 Limiting Distributional Equivalence

To summarize, the exponential, gamma, half-normal, truncated normal, and doubly truncated normal distributions are all scalable families with delta sequence representations as $\sigma_u \rightarrow 0$. The implication for the

stochastic frontier model is that these common inefficiency distributions possess a observational equivalence in the limit (up to location). That is, they all converge (in the distributional sense) to a Dirac delta with a singularity at zero, except for the truncated normal and doubly truncated normal with a positive pre-truncated mean ($\mu > 0$), which converge to Dirac deltas, but with singularities at $\mu > 0$ and $\min[\mu, B]$, respectively. Therefore, the sifting property of equation 5 causes the likelihood functions and F.O.C. under each specification to be the identical (up to location). In all cases the parameters associated with the inefficiency distribution are not identified.

Assuming that v is normally distributed, if the inefficiency distribution belongs to a single-parameter family, like the half-normal or exponential families, then the MLE is OLS as $\sigma_u \rightarrow 0$, and OLS (including an intercept) is identified and is a local maximum. If the inefficiency distribution is from a two (or greater) parameter family (gamma, truncated normal, doubly truncated normal), then none of the distributional parameters are identified as $\sigma_u \rightarrow 0$, as we have seen. Therefore, if we suspect that a particular sample is marked with low levels of inefficiency (i.e., as σ_u close to zero), then we may be better served estimating a model with a single-parameter assumption for inefficiency.

The departure point for empirical investigations of the stochastic frontier model is the N-HN model, and empiricists are advised to first estimate OLS (which assumes $\sigma_u = 0$), and to check the skew of the OLS residuals. If the skew of the residuals is wrong (and some level of inefficiency is anticipated), then the half normal model is not identified, and the prescription is to pull another sample or re-specify the model with, perhaps, an alternative distribution for inefficiency. The observational equivalence of the various parametric models implies that re-specifying the model may be a waste of time when some inefficiency variability is anticipated, but when that variability is quite low (σ_u close to 0). Assuming that pulling a new sample is infeasible, this argues for using the bagging approach of Simar and Wilson (2010) when the wrong skew problem arises. Our analysis of the U.S. Airlines data from Greene's *Econometric Analysis* textbook, seventh edition, Table F6.1 confirms this result. The point estimates of an airline cost function are identical for the

three models: OLS, N-E and N-TN. For the N-TN model, μ is not identified and the Hessian is singular, precluding calculation of standard errors. The Hessian is also singular in the N-E model.¹⁶

We now consider the stability of the OLS solutions for the most common models as $\sigma_u \rightarrow 0$. In some cases we provide theoretical proof of a stable solution and show that this solution occurs when the OLS residuals having the wrong skew. In other cases there is no theoretical solution forthcoming. In these cases the paper considers simulated evidence of the stability of the solution and its relation to the wrong skew of the OLS residuals.

3 Stability as $\sigma_u \rightarrow 0$

We now always assume $v \sim N(0, \sigma_v^2)$. The Hessians are derived for some common specifications of the inefficiency distribution and are used to show that the OLS solution is stable (a local maximum) under Assumption 1-5. Since the gamma stochastic frontier is rarely considered, we ignore it in what follows. The likelihoods for the N-TN and N-DTN models are:

$$\begin{aligned} \text{N-TN: } \ln L^{N-TN}(y, \beta, \sigma^2, \lambda, \mu) &= -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{\sum_i^n (y - x' \beta + \mu)^2}{2\sigma^2} \\ &\quad - n \ln \left[1 - \Phi\left(\frac{-\mu(\lambda^{-2} + 1)^{\frac{1}{2}}}{\sigma}\right) \right] + \sum_i^n \ln \left[1 - \Phi\left(\frac{(y - x' \beta)\lambda - \mu\lambda^{-1}}{\sigma}\right) \right], \end{aligned} \quad (10)$$

$$\begin{aligned} \text{N-DTN: } \ln L^{N-DTN}(y, \beta, \sigma^2, \lambda, \mu, B) &= -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{\sum_i^n (y - x' \beta + \mu)^2}{2\sigma^2} - n \ln \left[\Phi\left(\frac{(B - \mu)(\lambda^{-2} + 1)^{\frac{1}{2}}}{\sigma}\right) \right. \\ &\quad \left. - \Phi\left(\frac{-\mu(\lambda^{-2} + 1)^{\frac{1}{2}}}{\sigma}\right) \right] + \sum_i^n \ln \left[\Phi\left(\frac{(B + (y - x' \beta))\lambda + (B - \mu)\lambda^{-1}}{\sigma}\right) - \Phi\left(\frac{(y - x' \beta)\lambda - \mu\lambda^{-1}}{\sigma}\right) \right], \end{aligned} \quad (11)$$

where $\sigma^2 = \sigma_v^2 + \sigma_u^2$. Then we have the following theorem:

Theorem 1 *If $v \sim N(0, \sigma_v^2)$ and u has either a truncated normal or a doubly truncated normal distribution*

¹⁶Results of this empirical analysis are available from the authors upon request.

with $\mu \leq 0$, then Assumptions 1-5 are satisfied and OLS is a stationary point in the likelihood function as $\sigma_u \rightarrow 0$. Furthermore, "wrong skewness" of the OLS residuals is a sufficient condition for the stationary point to be a local maximum.

The proof is in the appendix. Theorem 1 generalizes the result of Waldman (1982) to the N-TN and N-DTN models. The condition that $\mu \leq 0$ ensures that the limiting distribution is a Dirac delta function at zero. Not surprisingly, our results on the behavior of the likelihood in the neighborhood of OLS nests and are identical to Waldman's result. That is:

$$\Delta \ln L_{\infty}^{N-HN} = \Delta \ln L_{\infty}^{N-TN} = \Delta \ln L_{\infty}^{N-DTN} = \frac{1}{6} \frac{\sum_i e_i^3}{\hat{\sigma}_v^3} \frac{2}{\sqrt{2\pi}} \frac{\pi - 4}{\pi} \gamma^3 + o(\gamma^4), \quad (12)$$

where e_i is the OLS residual, $\hat{\sigma}_v^3 = (\sum_i e_i^2/n)^{3/2}$, and γ is a small, positive number, representing a perturbation of the likelihood away from OLS. Since $(\pi - 4)/\pi < 0$, $\Delta \ln L$ is the opposite sign of $\sum_i e_i^3$, the skew of the OLS residuals. If this skew is "correct" (negative), then the likelihood increases ($\Delta \ln L > 0$) as we move away from OLS (i.e., as σ_u becomes positive). If the skew is wrong (positive), OLS is a local maximum.¹⁷

For the N-E model, the likelihood is:

$$\text{N-E: } \ln L^E(y, \beta, \sigma_v, \sigma_u,) = -n \ln(\sigma_u) + \sum_i^n \ln[1 - \Phi(\frac{(y - x'\beta)}{\sigma_v} + \frac{\sigma_v}{\sigma_u})] + \sum_i^n (\frac{(y - x'\beta)}{\sigma_u} + \frac{\sigma_v}{2\sigma_u^2}) \quad (13)$$

and we have that:

Theorem 2 *If $v \sim N(0, \sigma_v^2)$ and u has an exponential distribution, then Assumptions 1-5 are satisfied and OLS is a stationary point in the likelihood function as $\sigma_u \rightarrow 0$.*

¹⁷For the N-DTN model: $\text{plim}(\frac{1}{n}\sum e^3) = E[\varepsilon - E(\varepsilon)]^3 = -E[u - E(u)]^3 < 0$ for $\sigma_u^3 \neq 0$ and $\mu \leq 0$, and skew (u) is positive if $(B < 2\mu)$, which implies that the skew (ε) is negative, see Almanidis, Qian and Sickle (2014). For the N-TN model: $\text{plim}(\frac{1}{n}\sum e^3) = E[\varepsilon - E(\varepsilon)]^3 = -E[u - E(u)]^3 < 0$ for $\sigma_u^3 \neq 0$, skew (u) is positive which implies that the skew (ε) is negative, see Horrace (2015).

For the N-E model we show that as $\lim_{\sigma_{u,m} \rightarrow 0} \Delta \ln L(y, x, \beta, \sigma_v, \sigma_{u,m}) = 0$ and that there is no theoretical relationship between the skew of OLS residuals and the MLE of σ_u (later).

4 Simulations

Comprehensive simulations verified Theorems 1 and 2. In particular, our experiments confirmed that OLS is a stable solution in the N-TN and N-DTN models with non-positive pre-truncated mean (μ), when OLS residuals have the wrong skew. Even though we do not provide theoretical proof, this relationship held in our experiments when the pre-truncated mean was also positive. The results for N-TN and N-DTN models were consistent for a wide range of parameterizations and are available from the authors by request. Since we could not establish this relationship for the N-E model, it is our present focus.¹⁸ We impose the standard restriction that the total variance equals one ($\sigma^2 = \sigma_u^2 + \sigma_v^2 = 1$), and select relatively small values for the signal to noise ratio ($\lambda = \frac{\sigma_u}{\sigma_v} = 0.25, 0.50, 1.0$) and sample size ($n = 50, 100$) to ensure that there are sufficient cases where the OLS residuals have the wrong skew. The N-E DGP is $y = 3 + v - u$. The OLS estimates of σ_v^2 and β are used for the starting values in the optimization. When the OLS residuals have the correct (negative) skew, then the starting value of σ_u is set to the negative of this skew. When the OLS residuals have the wrong (positive) skew, then the starting value of σ_u is set to an arbitrarily small, positive value, implying that the starting value of the log-likelihood is quite large (see equation 13). However, this did not pose any convergence problem.

The simulations for the N-E model are in Figures 1-3. Each figure contains two plots of the MLE of σ_u as a function of the skew of the OLS residuals, with each circle representing one of 1,000 simulation draws. In all figures the MLE is always positive when the skew is negative and zero when the skew is non-negative, implying that a stability relationship exists even though we were unable to find one theoretically.

¹⁸The simulations are conducted in Matlab 7.4.0 version. We used unrestricted MLE to estimate all three models. The function *fiminuc* is used to maximize the log-likelihoods. This uses the BFGS Quasi-Newton method with mixed quadratic and cubic line search; it uses the BFGS formula for updating an approximation for the Hessian.

For example, Figure 1a is for the case $\lambda = 0.25$, $n = 50$. In this example, 27.5% of the simulation draws have the wrong (positive) skew of the OLS residuals, and the MLE is always 0 to the right of the origin. Figure 1b increases the sample size to $n = 100$, and the proportion of simulations draws with wrong (positive) skew decreases to 18.3%. Note that the cloud of estimates with the correct (negative) skew becomes more dense as n increases. The results are similar in Figures 2 and 3 except the signal to noise ratios have increased to $\lambda = 0.50$ and 1.0 (respectively), so that the frequency of "wrong skew" draws declines. For example, at a sample size of $n = 100$, as the signal to noise ratio increases from 0.5 to 1.0, the proportion of wrong skew decreases from 5.2% to 0.6%, respectively.

Figure 4 contains two similar plots for the N-TN model with a positive pre-truncated mean of inefficiency. Here, the data generation process excludes the intercept, as including it causes a variety of convergence problems in both the N-TN model and the N-DTN model (not reported here). The OLS estimate of σ_v was used as the starting value and the N-HN normal estimate for σ_u and μ set at zero.¹⁹ The figure contains the parameterizations $\lambda = 0.25$, $\mu = 0.1$ and $\sigma_u = 0.2425$ with $n = 50$ in Figure 4a and $n = 100$ in Figure 4b, causing the proportion of simulation draws with the wrong skew to decrease from 46.8% to 45.8% (respectively). It appears that even with a positive pre-truncated mean of inefficiency, the stability relationship holds. The simulation results for N-TN and N-DTN models were similar to Figure 4 for a wide range of parameterizations. It appears that wrong skew is always associated with OLS being a stable stationary point.

5 Conclusions

We show that the inefficiency distribution is important in determining the behavior of the likelihood of the composed error when the variance goes to zero. First, it must have a delta sequence representation, which essentially means that it must be from a scalable class of continuous distributions that converge in

¹⁹For completeness we point out that setting $B = \mu = 0$ as the starting values in the N-DTN model created convergence problem, so we used the maximal OLS residual as the starting value for B .

distribution to a Dirac delta. For example, the Chi-squared distribution would be an inappropriate choice, while the exponential family of distributions (e.g., the normal or exponential distribution) seems appropriate. Second, the location of the resulting Dirac delta determines identification of the MLE, as variance goes to zero. Third, any parameters associated with the inefficiency distribution will not be identified. Without parametric assumptions on the error components, we show that a stationary point may exist under very weak assumptions. Understanding stability of the stationary point is less clear as it requires derivatives with respect to the parameters of the inefficiency distribution, which causes the resulting function to no longer have a delta sequence representation. Therefore, stability analysis requires specific parametric forms for the error components. Waldman's (1982) full suite of results holds for the N-TN and N-DTN models for a non-positive pre-truncated mean. Furthermore, simulations suggest that it also holds when the pre-truncated mean is positive. We were unable to find a theoretical result relating the skew of OLS residuals to the MLE in the N-E model, however simulations suggest that such a relationship exists.

Our results on observational equivalence suggest that when the wrong skew problem arises due to a small population value of σ_u , re-specifying the inefficiency distribution may not solve the problem, and empiricists may be better served used the bagging technique of Simar and Wilson (2010). We have also argued that single-parameter families are preferred when inefficiency variance is closed to zero. Future research might look into the behavior of MLEs for the N-DTN model if the true inefficiency distribution has negative skew but the sample skew of the OLS residuals is negative (the wrong skew).

References

- [1] Aigner, D.J., Lovell C.A. K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6, 21-37.
- [2] Almanidis, P., Sickles, R.C., 2012, The skewness problem in stochastic frontier models: fact or fiction?

- in *Exploring Research Frontiers in Contemporary Statistics and Econometrics: A Festschrift in Honor of Leopold Simar*, Ingrid Van Keilegom and Paul Wilson (eds.), Springer Publishing: New York, 201-227.
- [3] Almanidis, P., Qian J., Sickles, R.C., 2014. Stochastic frontier models with bounded inefficiency. In *Festschrift in Honor of Peter Schmidt: Econometric Methods and Application*, Sickles, R., Horrace, W.C. (Eds). New York: Springer Science + Business Media.
- [4] Arley, N., Buch, K.R., 1950. *Introduction to the theory of probability and statistics*, New York: John Wiley & Sons, Inc., London: Chapman & Hall, Limited
- [5] Battese, G.E., Coelli, T.J., 1988. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics* 38, 387-399
- [6] Battese, G.E., Coelli, T.J., 1992. Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *Journal of Productivity Analysis* 3, 153-169
- [7] Battese, G.E., Coelli, T.J., 1995. A model for technical inefficiency effects in stochastic frontier production function for Panel data. *Empirical Economics* 20, 325-332.
- [8] Berger, L.R., Casella, G., 2002. *Statistical inference*, 2nd Edition. California: Duxbury.
- [9] Bracewell, R.N., 2000. *The Fourier Transform and its Applications*, 3rd Edition. New York: McGraw-Hill Higher Education.
- [10] Cauchy, A.-L., 1815. *Theorie de la propagation des ondes a la surface d'un fluide pesant d'une profondeur indefinie* (published 1827, with additional Notes). *Oeuvres* (1) 1, 4-318.
- [11] Cheng, S., 2006. Differentiation under the integral sign with weak derivatives. <http://www.gold-saucer.org/math/diff-int/diff-int.pdf>.
- [12] Chiang, C., Wainwright, K. , 2005. *Fundamental methods of mathematical economics*, 4th Edition. New York: McGraw-Hill/Irwin.

- [13] Degroot, M.H., Schervish M.J., 2002. Probability and statistics, 3rd Edition. New York: Adison-Wesley.
- [14] Dirac, P.A.M., 1930. The Principles of Quantum Mechanics. Oxford University Press, UK.
- [15] Feng, Q., Horrace W.C., Wu , G.L., 2013. Wrong skewness and finite sample correction in parametric stochastic frontier models. Department of Economics, Syracuse University, unpublished manuscript.
- [16] Frieden, B.R., 1983. Probability, statistical optics, and data testing, New York: Springer-Verlag Berlin Heidelberg.
- [17] Greene, W.H., 1990. A gamma-distributed stochastic frontier model. *Journal of Econometrics* 46,141-163.
- [18] Greene, W.H., 1986-2011. LIMDEP.
- [19] Griffiths, D.J., 2005. Introduction to quantum mechanics, 2nd Edition, New Jersey: Pearson Prentice Hall.
- [20] Horrace, W.C., Schmidt, P., 1996. Confidence statement for efficiency estimates from stochastic frontier models. *Journal of Productivity Analysis* 7, 257-282.
- [21] Horrace, W.C., 2005. On ranking and selection from independent truncated normal distributions. *Journal of Econometrics* 126, 335-354.
- [22] Horrace, W.C., Parmeter, C., 2011. Semiparametric deconvolution with unknown error variance. *Journal of Productivity Analysis* 35, 129-141.
- [23] Horrace, W.C., Parmeter, C., (forthcoming). A Laplace stochastic frontier model. *Econometric Reviews*.
- [24] Horrace, W.C., 2015. Moments of the truncated normal distribution. *Journal of Productivity Analysis* 43, 133-138.

- [25] Hoskins, R.F., Pinto, J.S., 2005. Theories of Generalised Functions: Distributions, Ultradistributions and Other Generalised Functions. Woodhead Publishing, Oxford, UK.
- [26] Jondrow, J., Lovell, C.A.K., Materov, I.S., Schmidt, P., 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics* 19, 233-238.
- [27] Kobayashi, M., 2009. Testing for jumps in the stochastic volatility models. *Mathematics and Computers in Simulation* 79, 2597-2608.
- [28] Kobayashi, M., 1991. Testing for autocorrelated disturbances in nonlinear regression analysis simulation. *Econometrica* 59, 1153-1159.
- [29] Kobayashi, M., Shi, X., 2005. Testing for EGARCH against stochastic volatility models. *Journal of Times Series Analysis* 26, 135-150.
- [30] Lee, L., Pitt, M., 1981. The measurement and sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics* 9, 43-64.
- [31] Lehmann, E.L., 1999. Elements of large-sample theory, New York: Springer Science + Business Media Inc.
- [32] Meeusen, W., Van den Broeck, J., 1977. Efficient estimation from Cobb-Douglas production functions with composed error. *International Economic Review* 18, 435-444.
- [33] Olson, J.A., Schmidt, P., Waldman, D., 1980. A Monte carlo study of estimators of stochastic frontier production functions. *Journal of Econometrics* 13, 67-82.
- [34] Ritter, C., Simar, L., 1997. Pitfalls of normal-gamma stochastic frontier models. *Journal of Productivity Analysis* 8, 167-182.
- [35] Rothenberg, J.T., 1971. Identification in parametric models. *Econometrica* 39, 577-591.

- [36] Schmidt, P., Sickles R.C., 1984. Production frontiers and panel data. *Journal of Business and Economics Statistics* 2, 367-374.
- [37] Schwartz, L., 1957, *Théorie des Distributions I*, Hermann, Paris
- [38] Simar, L., Wilson P.W., 2010. Inference from cross-sectional, stochastic frontier models. *Econometric Reviews* 29, 62-98.
- [39] Stevenson, R.E., 1980. Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics* 13, 57-66.
- [40] Waldman, D.M., 1982. A stationary point for the stochastic frontier likelihood. *Journal of Econometrics* 18, 275-279.

Appendix: Proofs of Lemma and Theorems

This appendix provides all the proofs for the lemma and theorems in the text.

A Proof that the DTN and TN converge to a Dirac delta

Proof.

Since the DTN nests the TN distribution, so we only need prove this for the DTN density:

$$f_u^{DTN}(u, \mu, \sigma_u, B) = \frac{\frac{1}{\sigma_u} \phi\left(\frac{u-\mu}{\sigma_u}\right)}{\Phi\left(\frac{B-\mu}{\sigma_u}\right) - \Phi\left(\frac{-\mu}{\sigma_u}\right)}; u \in [0, B], B > 0 \quad (1)$$

The limiting behavior of the density as $\sigma_u \rightarrow 0$ is governed by the limiting behaviors of the numerator and the denominator. We consider three cases. First, if $\mu \in (0, B)$, then the limit of the denominator:

$$\lim_{\sigma_u \rightarrow 0} \left\{ \Phi\left(\frac{B-\mu}{\sigma_u}\right) - \Phi\left(\frac{-\mu}{\sigma_u}\right) \right\} \quad (2)$$

is a finite and positive constant. Therefore, the limiting behavior of the DTN is dictated solely by limiting behavior of the numerator:

$$\lim_{\sigma_u \rightarrow 0} \frac{1}{\sigma_u} \phi\left(\frac{u-\mu}{\sigma_u}\right). \quad (3)$$

Since the numerator (divided by the finite and positive limit of the denominator) is proportional to the density of a $N(\mu, \sigma_u^2)$ random variate with $\mu \in (0, B)$, then the limit of the DTN density in this case is a Dirac delta with mass point at μ .

The two remaining cases to consider are $\mu \leq 0$ and $\mu \geq B$. In both cases the limits of the denominator in 2 and the numerator in 3 equal zero. Taking derivatives of these expressions with respect to σ_u , and applying L'Hopital's rule yields:

$$\lim_{\sigma_u \rightarrow 0} f_u^{DTN}(u, \mu, \sigma_u, B) = \lim_{\sigma_u \rightarrow 0} \frac{\frac{1}{\sigma_u^2} \left[\left(\frac{u-\mu}{\sigma_u}\right)^2 - 1 \right] \phi\left(\frac{u-\mu}{\sigma_u}\right)}{\frac{1}{\sigma_u} \left[\frac{-\mu}{\sigma_u} \phi\left(\frac{-\mu}{\sigma_u}\right) - \frac{B-\mu}{\sigma_u} \phi\left(\frac{B-\mu}{\sigma_u}\right) \right]}. \quad (4)$$

Some algebra on equation 4 yields:

$$\lim_{\sigma_u \rightarrow 0} f_u^{DTN}(u, \mu, \sigma_u, B) = \lim_{\sigma_u \rightarrow 0} \frac{\frac{1}{\mu} \exp\left(-\frac{u(u-2\mu)}{2\sigma_u^2}\right) \left[1 - \frac{(u-\mu)^2}{\sigma_u^2}\right]}{\frac{B-\mu}{\mu} \exp\left(-\frac{B(B-2\mu)}{2\sigma_u^2}\right) + 1}.$$

For $\mu \leq 0$ the limit of the denominator above equals 1, so we need only evaluate the limit of the numerator. That is:

$$\lim_{\sigma_u \rightarrow 0} \frac{1}{\mu} \exp\left(-\frac{u(u-2\mu)}{2\sigma_u^2}\right) \left[1 - \frac{(u-\mu)^2}{\sigma_u^2}\right].$$

In general the limit of the exponential term dominates the limit of the bracketed term. For $\mu \leq 0$ the limit of the exponential term is 0, except for $u = 0$ when it equals 1. When $u = 0$ two things may occur. First, if $\mu < 0$, then $\frac{1}{\mu}$ is a negative constant and the bracketed term goes to negative infinity in the limit, so the numerator goes to positive infinity in the limit. Second, if $\mu = 0$, then $\frac{1}{\mu} \rightarrow \infty$ and the bracketed term

equals 1, so (again) the numerator goes to positive infinity in the limit. Therefore, when $\mu \leq 0$, $\lim_{\sigma_u \rightarrow 0} f_u$ is a Dirac delta centered at 0.

For the $\mu \geq B$ case algebra on equation 4, yields:

$$\lim_{\sigma_u \rightarrow 0} f_u^{DTN}(u, \mu, \sigma_u, B) = \lim_{\sigma_u \rightarrow 0} \frac{\frac{1}{\mu-B} \exp\left(-\frac{u(u-2\mu)-B(B-2\mu)}{2\sigma_u^2}\right) \left[1 + \frac{(u-\mu)^2}{\sigma_u^2}\right]}{-\frac{\mu}{\mu-B} \exp\left(\frac{B(B-2\mu)}{2\sigma_u^2}\right) + 1}. \quad (5)$$

For $\mu \geq B$ the limit of the denominator above equals 1, so again we need only evaluate the limit of the numerator. That is:

$$\lim_{\sigma_u \rightarrow 0} \frac{1}{\mu-B} \exp\left(-\frac{u(u-2\mu)-B(B-2\mu)}{2\sigma_u^2}\right) \left[1 + \frac{(u-\mu)^2}{\sigma_u^2}\right]. \quad (6)$$

Again, in general the limit of the exponential term dominates the $\frac{1}{\mu-B}$ term and the limit of the bracketed term. For $\mu \geq B$ (and noting that $u \leq B$) the limit of the exponential term is 0, except for $u = B$ when it equals 1. When $u = B$ two things can occur. First, if $\mu > B$, then $\frac{1}{\mu-B}$ is a positive constant and the bracketed term goes to positive infinity in the limit, so the numerator goes to positive infinity in the limit. Second, if $\mu = B$, then $\frac{1}{\mu-B} \rightarrow \infty$ and the bracketed term equals 1, so (again) the numerator goes to positive infinity in the limit. Therefore, when $\mu \geq B$ it must be true that $\lim_{\sigma_u \rightarrow 0} f_u$ is a Dirac delta centered on B .

To summarize we have that as $\sigma_u \rightarrow 0$, then $f_u^{DTN}(u, \mu, \sigma_u, B)$ converges to a Dirac delta centered on:

$$\begin{array}{ll} 0 & \mu \leq 0 \\ \min(\mu, B) & \mu > 0 \end{array}$$

For the TN let $B \rightarrow \infty$, and we have that the TN distribution converges to a Dirac delta centered on:

$$\begin{array}{ll} 0 & \mu \leq 0 \\ \mu & \mu > 0 \end{array}$$

B Proof of Theorem 1

The loglikelihood for the N-DTN model $\ln L = \ln L^{N-DTN}$ is:

$$\begin{aligned} \ln L = & -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{\sum_i^n (y_i - x_i' \beta + \mu)^2}{2\sigma^2} - n \ln \Phi \left[\left(\frac{(B - \mu)(\lambda^{-2} + 1)^{\frac{1}{2}}}{\sigma} \right) - \Phi \left(\frac{-\mu(\lambda^{-2} + 1)^{\frac{1}{2}}}{\sigma} \right) \right] \\ & + \sum_i^n \ln \left[\Phi \left(\frac{(B + (y_i - x_i' \beta))\lambda + (B - \mu)\lambda^{-1}}{\sigma} \right) - \Phi \left(\frac{(y_i - x_i' \beta)\lambda - \mu\lambda^{-1}}{\sigma} \right) \right] \end{aligned}$$

where:

$$\begin{aligned} \lambda &= \frac{\sigma_u}{\sigma_v}, \\ \sigma^2 &= \sigma_u^2 + \sigma_v^2, \\ \frac{1}{\sigma_u} &= \frac{(\lambda^{-2} + 1)^{\frac{1}{2}}}{\sigma}. \end{aligned}$$

Let:

$$\begin{aligned}
A_1 &= \frac{(B - \mu)(\lambda^{-2} + 1)^{\frac{1}{2}}}{\sigma} \\
A_2 &= \frac{-\mu(\lambda^{-2} + 1)^{\frac{1}{2}}}{\sigma} \\
A_3 &= \frac{(B + (y_i - x'_i\beta))\lambda + (B - \mu)\lambda^{-1}}{\sigma} \\
A_4 &= \frac{(y_i - x'_i\beta)\lambda - \mu\lambda^{-1}}{\sigma}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\ln L^{N-DTN} &= -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{\sum_i^n (y_i - x'_i\beta + \mu)^2}{2\sigma^2} \\
&\quad - n \ln[\Phi(A_1) - \Phi(A_2)] + \sum_i^n \ln[\Phi(A_3) - \Phi(A_4)].
\end{aligned} \tag{7}$$

A few useful facts when $\mu \leq 0$.

Fact 1:

$$\lim_{\lambda \rightarrow 0, \mu \leq 0} \frac{\phi(A_1)}{\phi(A_2)} = \lim_{\lambda \rightarrow 0, \mu \leq 0} \exp \left[-\frac{B(B - 2\mu)^2(\lambda^{-2} + 1)}{2\sigma^2} \right] = 0 \tag{8}$$

Fact 2:

$$\lim_{\lambda \rightarrow 0, \mu \leq 0} \frac{\phi(A_3)}{\phi(A_4)} = \lim_{\lambda \rightarrow 0, \mu \leq 0} \exp \left[-\frac{B^2\lambda^4 + 2\varepsilon_i B\lambda^4 + B(B - 2\mu) + 2B(B - \mu)\lambda^2 + 2\varepsilon_i B\lambda^2}{\lambda^2\sigma^2} \right] = 0 \tag{9}$$

It follows we also exploit the fact that Lemma 2 implies that the stationary point will be OLS. So $\sum_{i=1}^n e_i = 0$ and $\sum_{i=1}^n e_i x_i = 0$ are satisfied. We derive the F.O.C. for the N-DTN model and take limits as $\lambda \rightarrow 0$ for the case $\mu \leq 0$. Lemma 2 ensures that the limit of the F.O.C. w.r.t β and σ^2 produces OLS estimates $\hat{\beta} = (\sum_i^n x_i x'_i)^{-1} \sum_i^n x_i y_i$ and $\hat{\sigma}^2 = n^{-1} \sum_i^n e_i^2$. The derivatives of $\ln L^{N-DTN}$ with w.r.t λ , μ and B are.

$$\begin{aligned}
\frac{\partial \ln L}{\partial \lambda} &= -n \frac{\phi(A_1) \frac{(B-\mu)}{\sigma} (\lambda^{-2} + 1)^{-\frac{1}{2}} (-\lambda^{-3}) - \phi(A_2) \frac{\mu}{\sigma} (\lambda^{-2} + 1)^{-\frac{1}{2}} (\lambda^{-3})}{\Phi(A_1) - \Phi(A_2)} \\
&\quad + \sum_i^n \frac{\phi(A_3) \frac{(B+(y_i-x'_i\beta))-(B-\mu)\lambda^{-2}}{\sigma} - \phi(A_4) \frac{(y_i-x'_i\beta)+\mu\lambda^{-2}}{\sigma}}{\Phi(A_3) - \Phi(A_4)} \\
\frac{\partial \ln L}{\partial \mu} &= -\frac{\sum_i^n (y - x'\beta + \mu)}{\sigma^2} - \frac{n}{\sigma} \frac{\phi(A_1)(-\lambda^{-2} + 1)^{\frac{1}{2}} - \phi(A_2)(-\lambda^{-2} + 1)^{\frac{1}{2}}}{\Phi(A_1) - \Phi(A_2)} \\
&\quad + \sum_i^n \frac{\phi(A_3)(-\frac{\lambda^{-1}}{\sigma}) - \phi(A_4)(-\frac{\lambda^{-1}}{\sigma})}{\Phi(A_3) - \Phi(A_4)} \\
\frac{\partial \ln L}{\partial B} &= -\frac{n}{\sigma} \frac{\phi(A_1)(\lambda^{-2} + 1)^{\frac{1}{2}}}{\Phi(A_1) - \Phi(A_2)} + \frac{\sum_i^n \phi(A_3)(\frac{\lambda+\lambda^{-1}}{\sigma})}{\Phi(A_3) - \Phi(A_4)}
\end{aligned} \tag{10}$$

Note that $\lim_{\lambda \rightarrow 0, \mu \leq 0} \phi(A_i) = 0$, so that $\lim_{\lambda \rightarrow 0, \mu \leq 0} \Phi(A_i) = 1$, $i = 1, \dots, 4$. We evaluate the limit of the above derivatives. First, we have:

$$\begin{aligned} \lim_{\lambda \rightarrow 0, \mu \leq 0} \frac{\partial \ln L}{\partial \lambda} &= \lim_{\lambda \rightarrow 0, \mu \leq 0} n \frac{\phi(A_1) \frac{(B-\mu)}{\sigma} (\lambda^{-2} + 1)^{-\frac{1}{2}} (-\lambda^{-3}) - \phi(A_2) \frac{\mu}{\sigma} (\lambda^{-2} + 1)^{-\frac{1}{2}} (\lambda^{-3})}{\Phi(A_1) - \Phi(A_2)} \\ &\quad + \lim_{\lambda \rightarrow 0, \mu \leq 0} \sum_i^n \frac{\phi(A_3) \frac{(B+(y-x'\beta))-(B-\mu)\lambda^{-2}}{\sigma} - \phi(A_4) \frac{(y-x'\beta)+\mu\lambda^{-2}}{\sigma}}{\Phi(A_3) - \Phi(A_4)} \end{aligned}$$

The limits of the numerators and denominators of the two terms on the RHS are all zero, as the limits of ϕ and Φ dominate all other terms. That is, the limits of the two terms on the RHS are:

$$\lim_{\lambda \rightarrow 0, \mu \leq 0} n \frac{\phi(A_1) \frac{(B-\mu)}{\sigma} (\lambda^{-2} + 1)^{-\frac{1}{2}} (-\lambda^{-3}) - \phi(A_2) \frac{\mu}{\sigma} (\lambda^{-2} + 1)^{-\frac{1}{2}} (\lambda^{-3})}{\Phi(A_1) - \Phi(A_2)} = \frac{0}{0} \quad (11)$$

$$\lim_{\lambda \rightarrow 0, \mu \leq 0} \sum_i^n \frac{\phi(A_3) \frac{(B+(y-x'\beta))-(B-\mu)\lambda^{-2}}{\sigma} - \phi(A_4) \frac{(y-x'\beta)+\mu\lambda^{-2}}{\sigma}}{\Phi(A_3) - \Phi(A_4)} = \frac{0}{0}. \quad (12)$$

After judicious application of L'Hopital's rule, we have:

$$\lim_{\lambda \rightarrow 0, \mu \leq 0} \frac{\partial \ln L}{\partial \lambda} = 0. \quad (13)$$

By similar arguments we also have:

$$\lim_{\lambda \rightarrow 0, \mu \leq 0} \frac{\partial \ln L}{\partial \mu} = \lim_{\lambda \rightarrow 0, \mu \leq 0} \frac{\partial \ln L}{\partial} = 0. \quad (14)$$

For the N-TN model, we simply let $B \rightarrow \infty$ in the above derivatives. That is,

$$\lim_{\lambda \rightarrow 0, \mu \leq 0} \frac{\partial \ln L}{\partial \lambda} \lim_{\lambda \rightarrow 0, \mu \leq 0} \lim_{B \rightarrow \infty} \frac{\partial \ln L}{\partial \lambda} = 0 \quad (15)$$

and

$$\lim_{\lambda \rightarrow 0, \mu \leq 0} \frac{\partial \ln L}{\partial \mu} \lim_{\lambda \rightarrow 0, \mu \leq 0} \lim_{B \rightarrow \infty} \frac{\partial \ln L}{\partial \mu} = 0 \quad (16)$$

We now derive the Hessian for the N-DTN model. It is partially derived in Almanidis and Sickles (2012) holding B and μ fixed. Lemma 3 provides the limit of the second-order partials and cross-partial w.r.t. β and σ^2 . The rest of the Hessian is given by:

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \mu \partial \beta} &= \frac{\sum_i^n x_i}{\sigma^2} + \sum_i^n \frac{-A_3^* \phi_i(A_3) A_3 \frac{\partial A_3}{\partial \beta} + A_4^* \phi_i(A_4) A_4 \frac{\partial A_4}{\partial \beta}}{\Phi(A_3) - \Phi(A_4)} \\ &\quad - \sum_i^n \frac{[A_3^* \phi_i(A_3) - A_4^* \phi_i(A_4)] [\phi_i(A_3) \frac{\partial A_3}{\partial \beta} - \phi_i(A_4) \frac{\partial A_4}{\partial \beta}]}{[\Phi(A_3) - \Phi(A_4)]^2} \end{aligned}$$

where $A_i^* = \partial A_i^* / \partial \mu$, $i = 3, 4$.

$$\frac{\partial^2 \ln L}{\partial B \partial \beta} = \sum_i^n \frac{-A_3^* \phi_i(A_3) A_3 \frac{\partial A_3}{\partial \beta} + \phi_i(A_3) \frac{\partial A_3^*}{\partial \beta}}{\Phi(A_3) - \Phi(A_4)} - \sum_i^n \frac{\phi_i(A_3) A_3^* \left[\phi_i(A_3) \frac{\partial A_3}{\partial \beta} + \phi_i(A_4) \frac{\partial A_4}{\partial \beta} \right]}{[\Phi(A_3) - \Phi(A_4)]^2} \quad (17)$$

where $A_3^* = \partial A_3^* / \partial B$.

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \lambda \partial \beta} = \sum_i^n \frac{\phi(A_3) \left[\frac{\partial A_3^*}{\partial \beta} - A_3^* \frac{\partial A_3}{\partial \beta} A_3 \right] - \phi(A_4) \left[\frac{\partial A_4^*}{\partial \beta} - A_4^* \frac{\partial A_4}{\partial \beta} A_4 \right]}{\Phi(A_3) - \Phi(A_4)} \\ - \sum_i^n \frac{[\phi(A_3) A_3^* - \phi(A_4) A_4^*] \left[\phi(A_3) \frac{\partial A_3}{\partial \beta} - \phi(A_4) \frac{\partial A_4}{\partial \beta} \right]}{[\Phi(A_3) - \Phi(A_4)]^2}, \end{aligned}$$

where $A_i^* = \partial A_i^* / \partial \lambda$, $i = 3, 4$.

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \sigma^2 \partial B} = -n \left[\frac{\phi(A_1) \left[\frac{\partial A_1^*}{\partial B} - A_1^* \frac{\partial A_1}{\partial B} A_1 \right]}{\Phi(A_1) - \Phi(A_2)} - \frac{[\phi(A_1) A_1^* - \phi(A_2) A_2^*] \phi(A_1) \frac{\partial A_1}{\partial B}}{[\Phi(A_1) - \Phi(A_2)]^2} \right] \\ + \sum_i^n \frac{\phi(A_3) \left[\frac{\partial A_3^*}{\partial B} - A_3^* \frac{\partial A_3}{\partial B} A_3 \right]}{\Phi(A_3) - \Phi(A_4)} - \sum_i^n \frac{[\phi(A_3) A_3^* - \phi(A_4) A_4^*] \phi(A_3) \frac{\partial A_3}{\partial B}}{[\Phi(A_3) - \Phi(A_4)]^2} \end{aligned}$$

where $A_i^* = \partial A_i^* / \partial \sigma^2$, $i = 1, 2, 3, 4$.

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \mu \partial B} = -n \left[\frac{-\phi(A_1) A_1^* \frac{\partial A_1}{\partial B} A_1}{\Phi(A_1) - \Phi(A_2)} - \frac{[\phi(A_1) A_1^* - \phi(A_2) A_2^*] \phi(A_1) \frac{\partial A_1}{\partial B}}{[\Phi(A_1) - \Phi(A_2)]^2} \right] \\ + \sum_i^n \frac{-\phi(A_3) A_3^* \frac{\partial A_3}{\partial B} A_3}{\Phi(A_3) - \Phi(A_4)} - \sum_i^n \frac{[\phi(A_3) A_3^* - \phi(A_4) A_4^*] \phi(A_3) \frac{\partial A_3}{\partial B}}{[\Phi(A_3) - \Phi(A_4)]^2} \end{aligned}$$

where $A_i^* = \partial A_i^* / \partial \mu$, $i = 1, 2, 3, 4$.

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \lambda \partial B} = -n \left[\frac{\phi(A_1) \left[-A_1^* \frac{\partial A_1}{\partial B} A_1 + \frac{\partial A_1^*}{\partial B} \right]}{\Phi(A_1) - \Phi(A_2)} - \frac{[\phi(A_1) A_1^* - \phi(A_2) A_2^*] \phi(A_1) \frac{\partial A_1}{\partial B}}{[\Phi(A_1) - \Phi(A_2)]^2} \right] \\ + \sum_i^n \frac{\phi(A_3) \left[-A_3^* \frac{\partial A_3}{\partial B} A_3 + \frac{\partial A_3^*}{\partial B} \right]}{\Phi(A_3) - \Phi(A_4)} - \sum_i^n \frac{[\phi(A_3) A_3^* - \phi(A_4) A_4^*] \phi(A_3) \frac{\partial A_3}{\partial B}}{[\Phi(A_3) - \Phi(A_4)]^2} \end{aligned}$$

where $A_i^* = \partial A_i^* / \partial \lambda$, $i = 1, 2, 3, 4$.

$$\frac{\partial^2 \ln L}{\partial B^2} = -n \left[\frac{-\phi(A_1) A_1^{*2} A_1}{\Phi(A_1) - \Phi(A_2)} - \frac{(\phi(A_1) A_1^*)^2}{[\Phi(A_1) - \Phi(A_2)]^2} \right] + \sum_i^n \frac{-\phi(A_3) A_3^{*2} A_3}{\Phi(A_3) - \Phi(A_4)} - \sum_i^n \frac{(\phi(A_3) A_3^*)^2}{[\Phi(A_3) - \Phi(A_4)]^2}$$

where $A_i^* = \partial A_i^* / \partial B$, $i = 1, 3$.

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \mu^2} = & -\frac{n}{\sigma^2} - n \left[\frac{-\phi(A_1)A_1^{*2}A_1 + \phi(A_2)A_2^{*2}A_2}{\Phi(A_1) - \Phi(A_2)} - \frac{[\phi(A_1)A_1^* - \phi(A_2)A_2^*]^2}{[\Phi(A_1) - \Phi(A_2)]^2} \right] \\ & + \sum_i^n \frac{[-\phi(A_3)A_3^{*2}A_3 + \phi(A_4)A_4^{*2}A_4]}{\Phi(A_3) - \Phi(A_4)} - \sum_i^n \frac{[\phi(A_3)A_3^* - \phi(A_4)A_4^*]^2}{[\Phi(A_3) - \Phi(A_4)]^2} \end{aligned}$$

where $A_i^* = \partial A_i^* / \partial \mu$, $i = 1, 2, 3, 4$.

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \mu} = & \frac{\sum_i^n (y - x'\beta + \mu)}{\sigma^4} - n \frac{\phi(A_1) \left[\frac{\partial A_1^*}{\partial \mu} - A_1^* \frac{\partial A_1}{\partial \mu} A_1 \right] - \phi(A_2) \left[\frac{\partial A_2^*}{\partial \mu} - A_2^* \frac{\partial A_2}{\partial \mu} A_2 \right]}{\Phi(A_1) - \Phi(A_2)} \\ & + n \frac{[\phi(A_1)A_1^* - \phi(A_2)A_2^*][\phi(A_1) \frac{\partial A_1}{\partial \mu} - \phi(A_2) \frac{\partial A_2}{\partial \mu}]}{[\Phi(A_1) - \Phi(A_2)]^2} + \sum_i^n \frac{\phi(A_3) \left[\frac{\partial A_3^*}{\partial \mu} - A_3^* \frac{\partial A_3}{\partial \mu} A_3 \right] - \phi(A_4) \left[\frac{\partial A_4^*}{\partial \mu} - A_4^* \frac{\partial A_4}{\partial \mu} A_4 \right]}{\Phi(A_3) - \Phi(A_4)} \\ & - \sum_i^n \frac{[\phi(A_3)A_3^* - \phi(A_4)A_4^*][\phi(A_3) \frac{\partial A_3}{\partial \mu} - \phi(A_4) \frac{\partial A_4}{\partial \mu}]}{[\Phi(A_3) - \Phi(A_4)]^2} \end{aligned}$$

where $A_i^* = \partial A_i^* / \partial \sigma^2$, $i = 1, 2, 3, 4$.

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \lambda \partial \mu} = & -n \frac{\phi(A_1) \left[\frac{\partial A_1^*}{\partial \mu} - A_1^* \frac{\partial A_1}{\partial \mu} A_1 \right] - \phi(A_2) \left[\frac{\partial A_2^*}{\partial \mu} - A_2^* \frac{\partial A_2}{\partial \mu} A_2 \right]}{\Phi(A_1) - \Phi(A_2)} \\ & + n \frac{[\phi(A_1)A_1^* - \phi(A_2)A_2^*][\phi(A_1) \frac{\partial A_1}{\partial \mu} - \phi(A_2) \frac{\partial A_2}{\partial \mu}]}{[\Phi(A_1) - \Phi(A_2)]^2} + \sum_i^n \frac{\phi(A_3) \left[\frac{\partial A_3^*}{\partial \mu} - A_3^* \frac{\partial A_3}{\partial \mu} A_3 \right] - \phi(A_4) \left[\frac{\partial A_4^*}{\partial \mu} - A_4^* \frac{\partial A_4}{\partial \mu} A_4 \right]}{\Phi(A_3) - \Phi(A_4)} \\ & - \sum_i^n \frac{[\phi(A_3)A_3^* - \phi(A_4)A_4^*][\phi(A_3) \frac{\partial A_3}{\partial \mu} - \phi(A_4) \frac{\partial A_4}{\partial \mu}]}{[\Phi(A_3) - \Phi(A_4)]^2} \end{aligned}$$

where $A_i^* = \partial A_i^* / \partial \lambda$, $i = 1, 2, 3, 4$.

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \lambda} = & -n \frac{\phi(A_1) \left[\frac{\partial A_1^*}{\partial \lambda} - A_1^* \frac{\partial A_1}{\partial \lambda} A_1 \right] - \phi(A_2) \left[\frac{\partial A_2^*}{\partial \lambda} - A_2^* \frac{\partial A_2}{\partial \lambda} A_2 \right]}{\Phi(A_1) - \Phi(A_2)} + n \frac{[\phi(A_1)A_1^* - \phi(A_2)A_2^*][\phi(A_1) \frac{\partial A_1}{\partial \lambda} - \phi(A_2) \frac{\partial A_2}{\partial \lambda}]}{[\Phi(A_1) - \Phi(A_2)]^2} \\ & + \sum_i^n \frac{\phi(A_3) \left[\frac{\partial A_3^*}{\partial \lambda} - A_3^* \frac{\partial A_3}{\partial \lambda} A_3 \right] - \phi(A_4) \left[\frac{\partial A_4^*}{\partial \lambda} - A_4^* \frac{\partial A_4}{\partial \lambda} A_4 \right]}{\Phi(A_3) - \Phi(A_4)} \\ & - \sum_i^n \frac{[\phi(A_3)A_3^* - \phi(A_4)A_4^*][\phi(A_3) \frac{\partial A_3}{\partial \lambda} - \phi(A_4) \frac{\partial A_4}{\partial \lambda}]}{[\Phi(A_3) - \Phi(A_4)]^2} \end{aligned}$$

where $A_i^* = \partial A_i^* / \partial \sigma^2$, $i = 1, 2, 3, 4$.

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \lambda^2} = & -n \frac{\phi(A_1) \left[\frac{\partial A_1^*}{\partial \lambda} - A_1^* A_1 \frac{\partial A_1}{\partial \lambda} \right] - \phi(A_2) \left[\frac{\partial A_2^*}{\partial \lambda} - A_2^* A_2 \frac{\partial A_2}{\partial \lambda} \right]}{\Phi(A_1) - \Phi(A_2)} + n \frac{[\phi(A_1)A_1^* - \phi(A_2)A_2^*]^2}{[\Phi(A_1) - \Phi(A_2)]^2} \\ & + \sum_i^n \frac{\phi(A_3) \left[\frac{\partial A_3^*}{\partial \lambda} - A_3^* A_3 \frac{\partial A_3}{\partial \lambda} \right] - \phi(A_4) \left[\frac{\partial A_4^*}{\partial \lambda} - A_4^* A_4 \frac{\partial A_4}{\partial \lambda} \right]}{\Phi(A_3) - \Phi(A_4)} \\ & - \sum_i^n \frac{[\phi(A_3)A_3^* - \phi(A_4)A_4^*]^2}{[\Phi(A_3) - \Phi(A_4)]^2} \end{aligned}$$

where $A_i^* = \partial A_i^* / \partial \lambda$, $i = 1, 2, 3, 4$. Applying L'Hopital's rule and taking limits:

$$\lim_{\lambda \rightarrow 0, \mu \leq 0} \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = -\frac{\sum_i^n x_i x_i'}{\sigma^2}; \quad \lim_{\lambda \rightarrow 0, \mu \leq 0} \frac{\partial^2 \ln L}{\partial \beta \partial \mu} = \frac{\sum_i^n x_i}{\sigma^2}; \quad \lim_{\lambda \rightarrow 0, \mu \leq 0} \frac{\partial^2 \ln L}{\partial (\sigma^2)^2} = -\frac{n}{2\sigma^4}; \quad \lim_{\lambda \rightarrow 0, \mu \leq 0} \frac{\partial^2 \ln L}{\partial \mu^2} = -\frac{2n}{\sigma^2}$$

with all the other limits equaling zero.¹ The Hessian for $(\beta, \sigma, \mu, B, \lambda)$ evaluated at OLS $\theta^{*'} = (\hat{\beta}, \hat{\sigma}^2, \hat{\mu} = 0, \hat{\lambda} = 0)$ is:

$$H(\theta^*) = \begin{bmatrix} \frac{-\sum_i^n x_i x_i'}{\sigma_v^2} & 0 & \frac{\sum_i^n x_i}{\sigma_v^2} & 0 & 0 \\ 0 & \frac{-n}{2\sigma_v^4} & 0 & 0 & 0 \\ \frac{\sum_i^n x_i}{\sigma_v^2} & 0 & \frac{-2n}{\sigma_v^2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$H(\theta^*)$ is negative semidefinite with two zero eigenvalues. The eigenvectors associated with zero eigenvalues are:

$$z_1' = [\mathbf{0}' \quad 1 \quad 0] \quad (18)$$

and

$$z_2' = [\mathbf{0}' \quad 0 \quad 1] \quad (19)$$

where $\mathbf{0}$ is a $K + 2$ vector of zeros. In what follows we use the fact $\lim_{\lambda \rightarrow 0, \mu \leq 0} \frac{\hat{\mu}}{\lambda} = 0$. The change in the loglikelihood is evaluated based on the number of non-zero elements in z_1 and z_2 , which correspond to λ and B .

$$\begin{aligned} \Delta \ln L &= \ln L(\theta^* + z_1 \gamma_1 + z_2 \gamma_2) - \ln L(\theta^*) \\ &= -n \ln \left[\Phi \left(\frac{\gamma_1 (\gamma_2^{-2} + 1)^{\frac{1}{2}}}{\hat{\sigma}} \right) - \Phi(0) \right] + \sum_i^n \ln \Phi \left[\left(\frac{(\gamma_1 + e_i) \gamma_2 + \gamma_1 \gamma_2^{-1}}{\hat{\sigma}} \right) - \Phi \left(\frac{e_i \gamma_2}{\hat{\sigma}} \right) \right] \quad (20) \end{aligned}$$

We use a third-order Taylor series expansion of $\Delta \ln L$ around the point $(\hat{\lambda} = 0, B = B_o)$ and let $B_o \rightarrow \infty$. Recall that B is a free parameter in the F.O.C. above. If $\Delta \ln L \leq 0$, then OLS is a $\hat{\lambda} = 0$ is a stable solution; otherwise it is not and maximizing the likelihood will move the solution to away from the border of the parameter space.

¹Derivations of the results are available upon request.

$$\begin{aligned}
\Delta \ln L &\simeq \lim_{B_o \rightarrow \infty} \ln L(\hat{\lambda} = 0, B_o) + \frac{\Delta \ln L}{\partial B}(\gamma_1) + \frac{\partial \Delta \ln L}{\partial \lambda}(\gamma_2) + \frac{1}{2} \left[\frac{\partial^2 \Delta \ln L}{\partial B^2}(\gamma_1)^2 + \frac{\partial^2 \Delta \ln L}{\partial \lambda^2}(\gamma_2)^2 \right. \\
&\quad + 2 \frac{\partial^2 \Delta \ln L}{\partial B \partial \lambda}(\gamma_1)(\gamma_2) \left. \right] + \frac{1}{6} \left[\frac{\partial^3 \Delta \ln L}{\partial B^3}(\gamma_1)^2 + 3 \frac{\partial^3 \Delta \ln L}{\partial B^2 \partial \lambda}(\gamma_1)^2(\gamma_2) + 3 \frac{\partial^3 \Delta \ln L}{\partial B \partial \lambda^2}(\gamma_1)(\gamma_2)^2 \right. \\
&\quad \left. + \frac{\partial^3 \Delta \ln L}{\partial \lambda^3}(\gamma_2)^3 \right] + o(\gamma_2^4)
\end{aligned}$$

Note that $\lim_{B_o \rightarrow \infty} \frac{\partial^3 \Delta \ln L}{\partial B^3} = 0$, so the only term that needs to be evaluated is $\frac{\partial^3 \Delta \ln L}{\partial \lambda^3}$, yielding precisely the Waldman result for the N-DTN model with $\mu \leq 0$:

$$\Delta \ln L = \frac{1}{6} \frac{\Sigma e^3}{\hat{\sigma}_v^3} \frac{2}{\sqrt{2\pi}} \left[\frac{\pi - 4}{\pi} \right] \gamma_2^3 + o(\gamma_2^4). \tag{21}$$

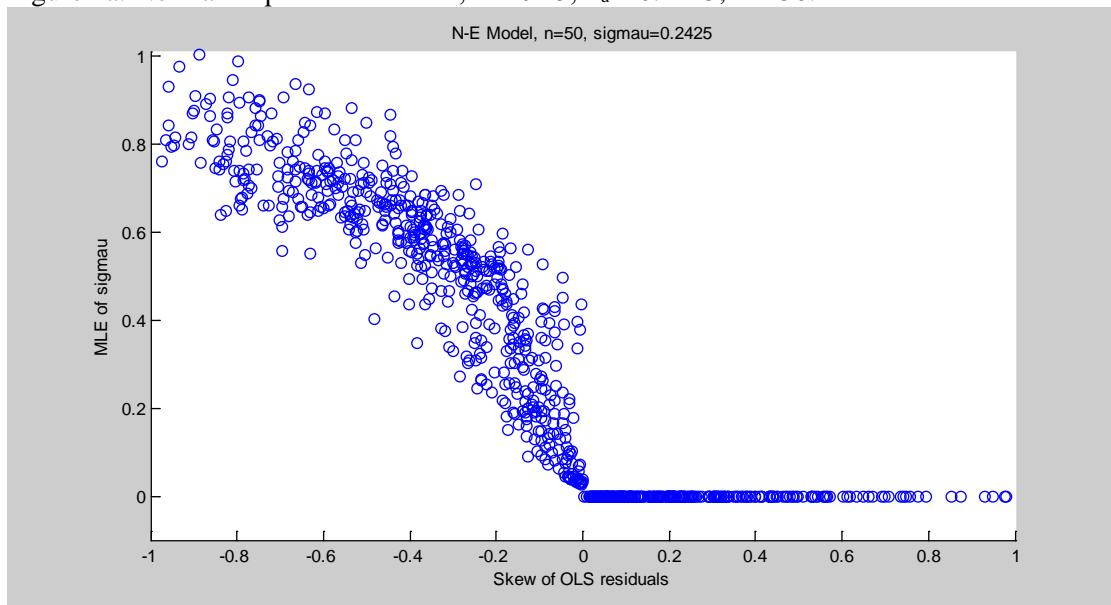
Since this approximation allows $B_o \rightarrow \infty$, it must also hold for the N-TN model. Therefore, when $\mu \leq 0$ OLS is a stable stationary point when the OLS residual skew is positive for the normal half-normal (due to Waldman), normal truncated-normal and normal doubly-truncated-normal models.

C Proof of Theorem 2

(later)

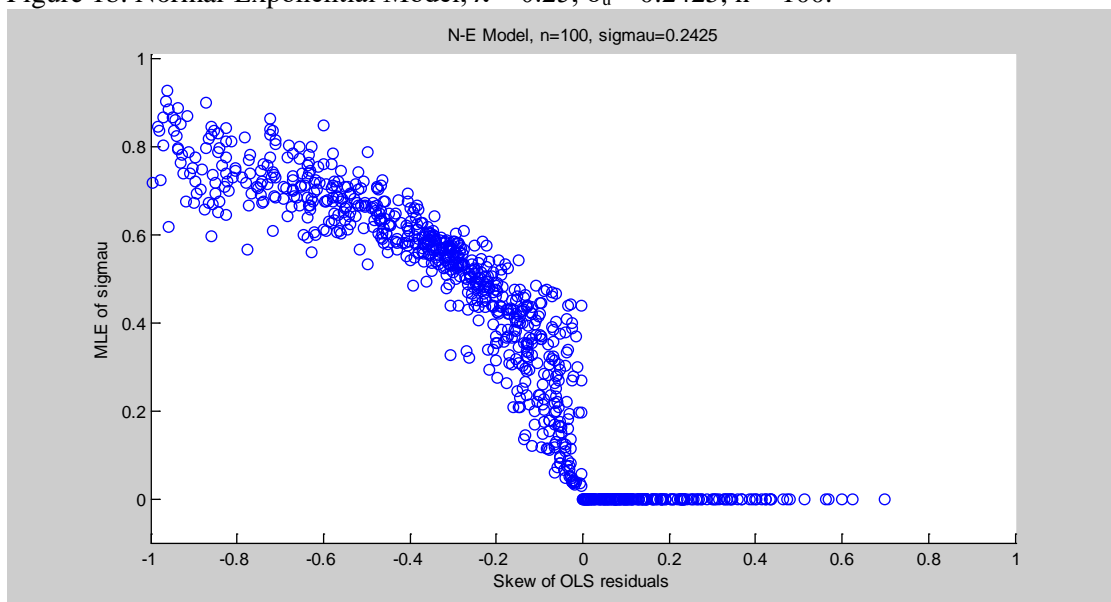
END OF APPENDIX

Figure 1a. Normal-Exponential Model, $\lambda = 0.25$, $\sigma_u = 0.2425$, $n = 50$.



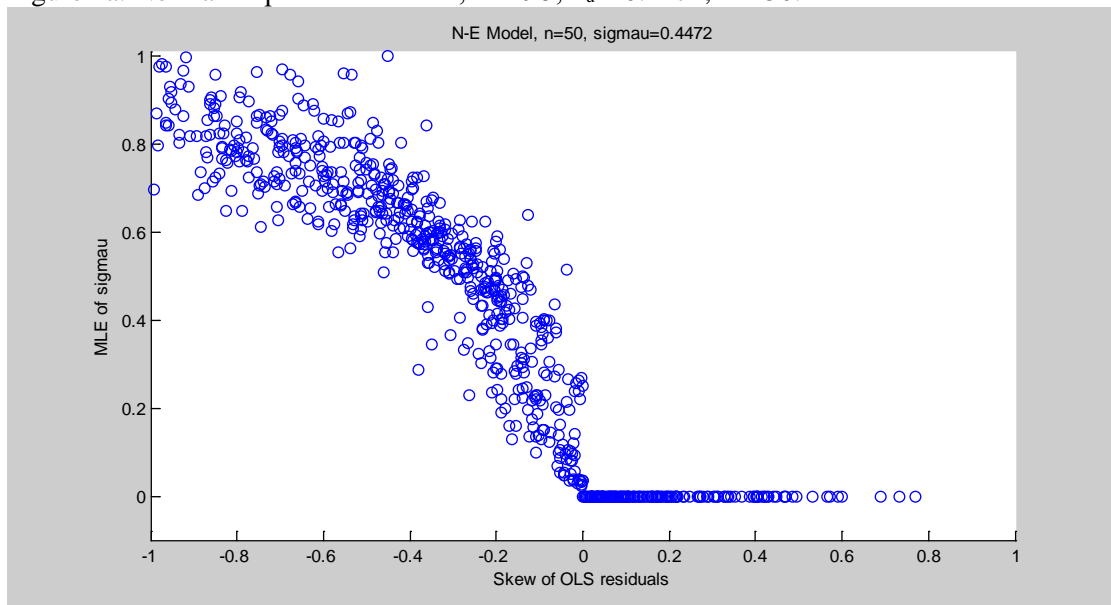
1,000 simulation draws. Frequency of wrong skew = 27.5%

Figure 1b. Normal-Exponential Model, $\lambda = 0.25$, $\sigma_u = 0.2425$, $n = 100$.



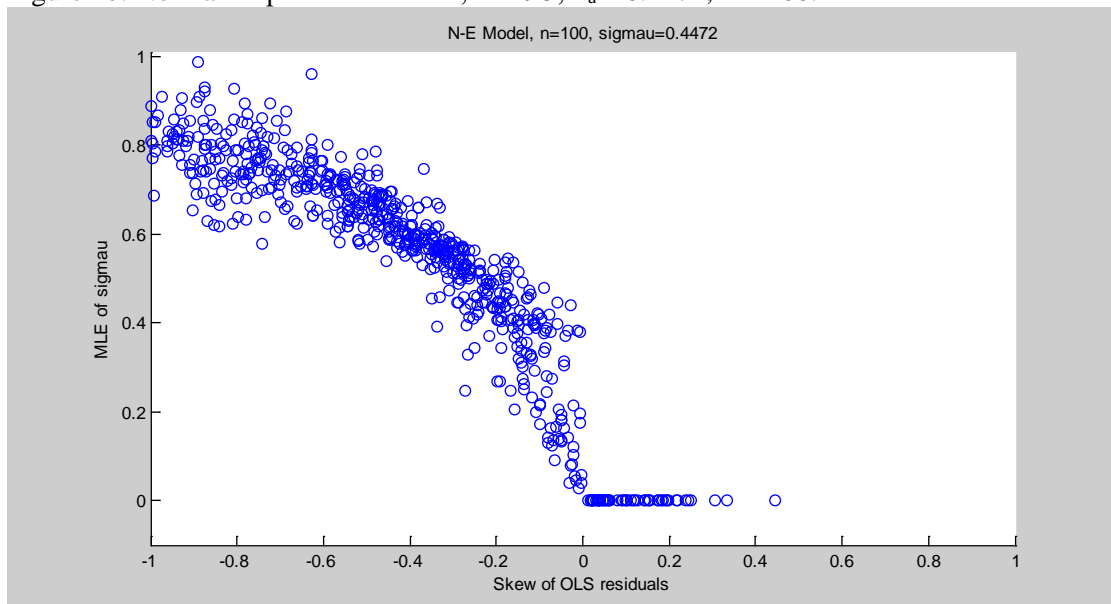
1,000 simulation draws. Frequency of wrong skew = 18.3%

Figure 2a. Normal-Exponential Model, $\lambda = 0.5$, $\sigma_u = 0.4472$, $n = 50$.



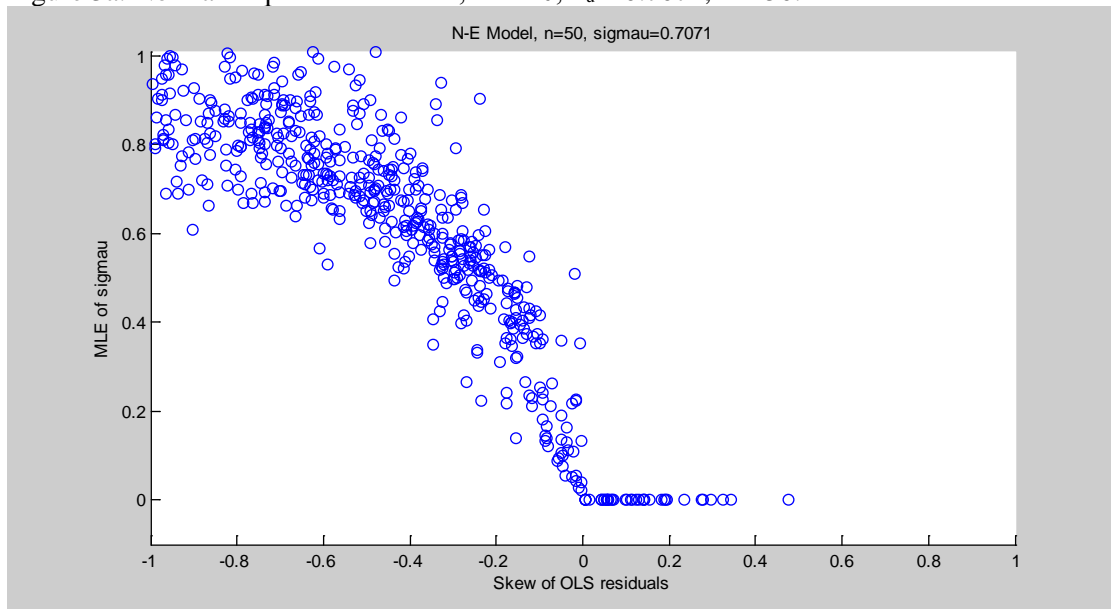
1,000 simulation draws. Frequency of wrong skew = 15.7%

Figure 2b. Normal-Exponential Model, $\lambda = 0.5$, $\sigma_u = 0.4472$, $n = 100$.



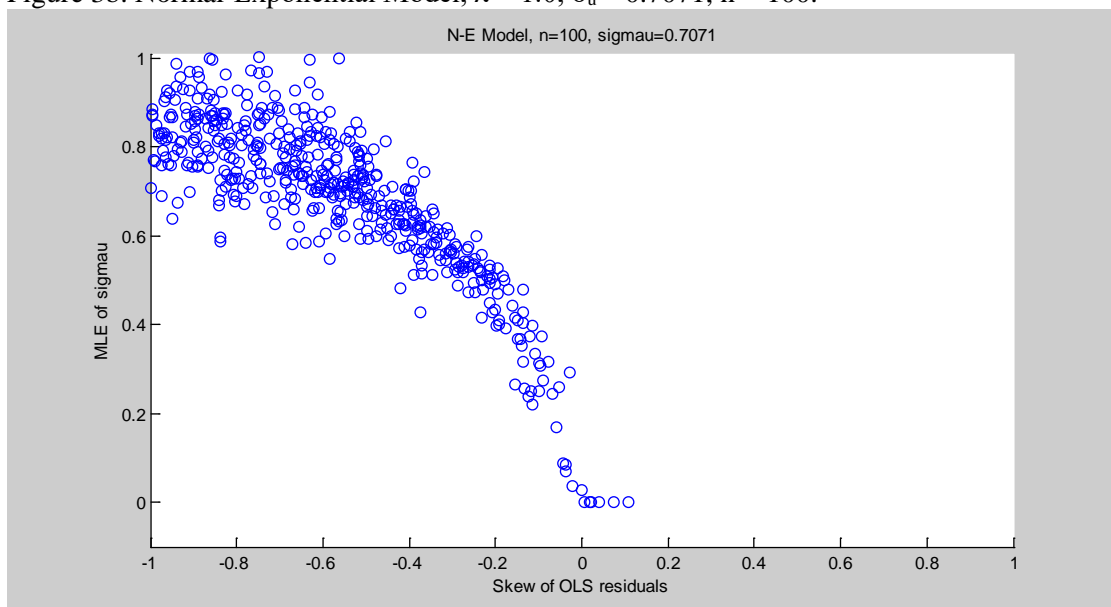
1,000 simulation draws. Frequency of wrong skew = 5.2%

Figure 3a. Normal-Exponential Model, $\lambda = 1.0$, $\sigma_u = 0.7071$, $n = 50$.



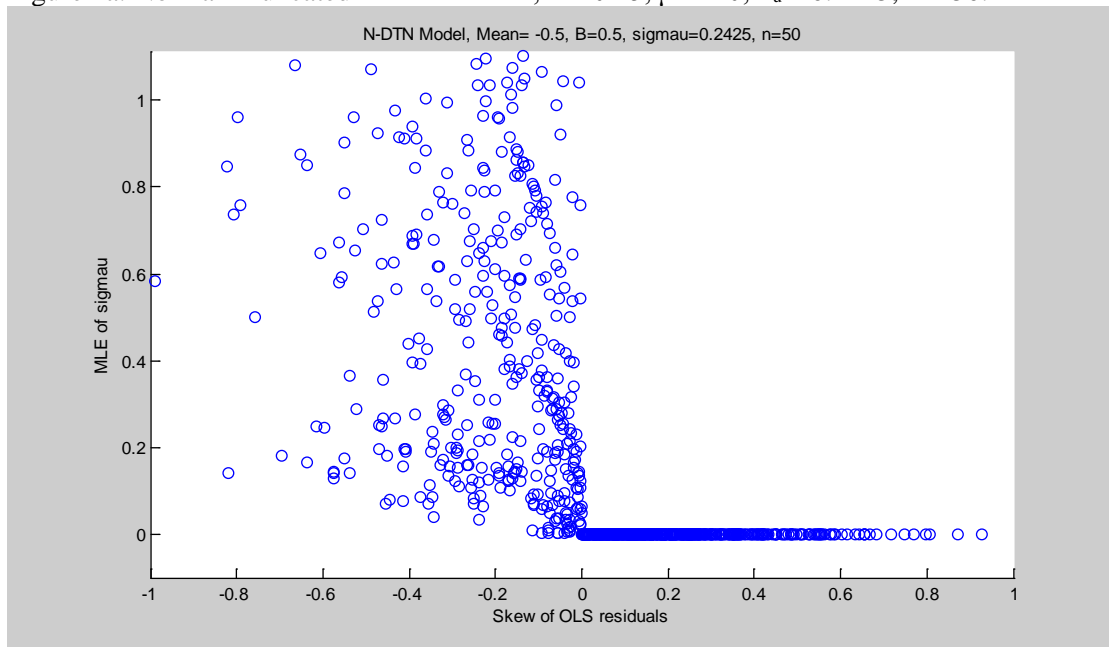
1,000 simulation draws. Frequency of wrong skew = 3.7%

Figure 3b. Normal-Exponential Model, $\lambda = 1.0$, $\sigma_u = 0.7071$, $n = 100$.



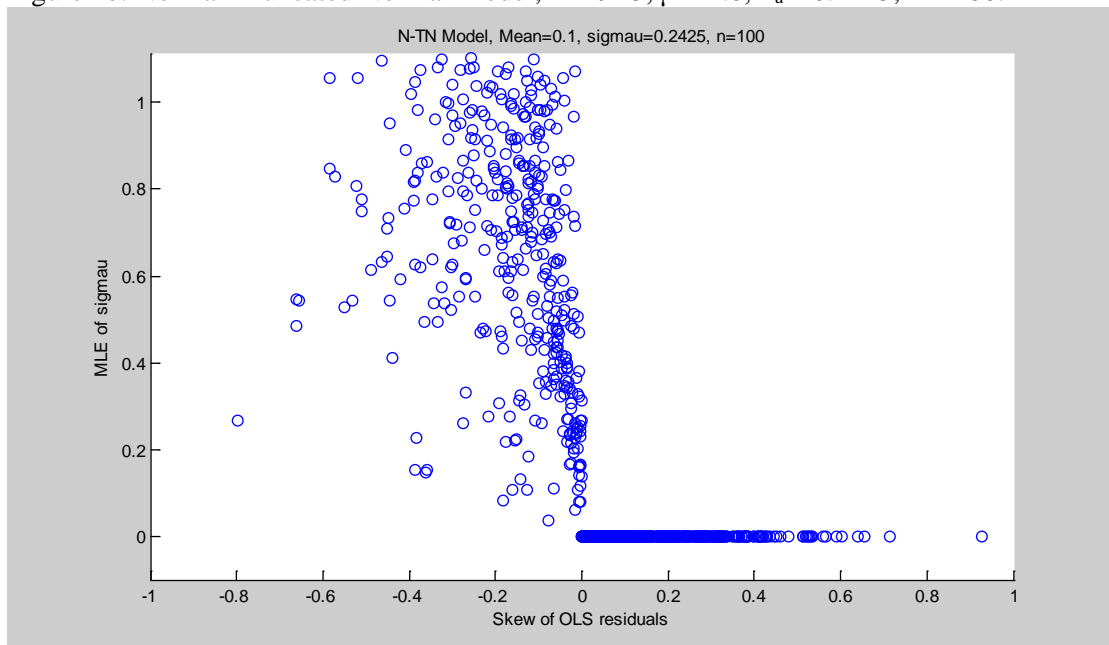
1,000 simulation draws. Frequency of wrong skew = 0.6%

Figure 4a. Normal-Truncated Normal Model, $\lambda = 0.25$, $\mu = 1.0$, $\sigma_u = 0.2425$, $n = 50$.



1,000 simulation draws. Frequency of wrong skew = 46.8%

Figure 4b. Normal-Truncated Normal Model, $\lambda = 0.25$, $\mu = 1.0$, $\sigma_u = 0.2425$, $n = 100$.



1,000 simulation draws. Frequency of wrong skew = 45.8%