



Stochastic frontier models using the Generalized Exponential distribution

Alecos Papadopoulos ¹

Accepted: 24 November 2020 / Published online: 2 February 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

We present a new, single-parameter distributional specification for the one-sided error components in single-tier and two-tier stochastic frontier models. The distribution has its mode away from zero, and can represent cases where the most likely outcome is non-zero inefficiency. We present the necessary formulas for estimating production, cost and two-tier stochastic frontier models in logarithmic form. We pay particular attention to the use of the conditional mode as a predictor of individual inefficiency. We use simulations to assess the performance of existing models when the data include an inefficiency term with non-zero mode, and we also contrast the conditional mode to the conditional expectation as measures of individual (in)efficiency.

Keywords Stochastic frontier analysis · Distribution mode · Conditional expectation · Two-tier stochastic frontier

JEL D24 · C46 · C21

1 Introduction and motivation

In the founding papers of stochastic frontier analysis (SFA), Aigner et al. (1977) and Meeusen and van den Broeck (1977), the authors considered two distributions for the inefficiency term, the Half Normal and the Exponential, creating organically the Normal-Half Normal (NHN) and the Normal-Exponential (NE) specifications for the composite error term. The first would go on and have a spectacular future outside SFA also, after Azzalini (1985) baptized it as the Skew Normal distribution and presented it to the statistical community. The second had already a well established past, since a variant of it under the name “Exponentially modified Gaussian” was being used already from the ’60s in chromatography (for a review see Grushka, 1972).

But no one is a prophet in their own land: Stevenson (1980) noted that in both specifications the inefficiency distribution had its mode at zero, imposing a specific economic/structural assumption: that the most likely occurrence was near-zero inefficiency. While such an assumption could

be justified to a degree by invoking purposeful optimizing behavior as well as the forces of competition, Stevenson linked inefficiency to managerial competence and argued that the latter is not distributed monotonically in the population (of managers at least). Based on this argument, the author proposed the use of the Truncated Normal distribution for the inefficiency term as an alternative, as well as the Gamma distribution, restricting the values of its shape parameter so as to obtain a closed-form density for the composite error term.¹ To this day, these are the main specifications that allow for a non-zero mode of the inefficiency component.² But both these distributions have issues: the maximum likelihood estimator (MLE) has often difficulties converging under a Truncated Normal specification, while the Gamma specification in its general formulation has a non closed-form density that makes it less appealing for empirical implementations. Moreover, Ritter and Simar (1997) found that the shape parameter of the Gamma distribution is weakly identified and imprecisely estimated when the sample size is not really large. But this is the very parameter that allows us to have a non-zero mode.

✉ Alecos Papadopoulos
papadopalex@aueb.gr

¹ Athens University of Economics and Business, Athens, Greece

¹ Greene (1980, 1990) studied the Normal-Gamma specification in much more depth.

² Other distributions have been proposed for the inefficiency term, see Parmeter and Kumbhakar (2014) and Stead et al. (2019).

In this paper we present production and cost stochastic frontier models as well as the two-tier extension (2TSF), where we specify the one sided component(s) to follow the Generalized Exponential distribution, which is a single-parameter distribution that has always a non-zero mode. Assuming in addition that the noise term follows a zero-mean Normal distribution, we obtain the Normal-Generalized Exponential specification (NGE).

Since concerns related to the mode of the inefficiency distribution is what generated this research, it is fitting that we will also pay particular attention to the conditional mode as a measure of individual inefficiency. These measures are sometimes called “JLMS” measures from the paper of Jondrow et al. (1982) where both the conditional expected value and the conditional mode were considered as predictors of individual inefficiency.³ In fact this paper dealt with predicting the error component of the logarithmic specification. It was Battese and Coelli (1988) that presented the conditional expectation expressions for the exponentiated error, i.e. for a prediction at the original measurement scale. We extend the approach by examining the conditional modes for the exponentiated case. In practice, the conditional expectation, being an optimal predictor under the Mean-Squared Error criterion, appears to have prevailed as the inefficiency measure of choice. But by considering the mode of the inefficiency distribution we essentially propose a more elaborate investigation of the inefficiency terms, since by having available their (marginal and conditional) distributions, we can go beyond obtaining some predictor of their value and instead form a more complete picture of their stochastic behavior. In addition, the mode always exists and it may also be easier to derive.

In most cases the regression specification used in empirical SFA papers has the dependent variable in logarithmic form, and it is for this equation that distributional assumptions are made. This implies that inefficiency measures ultimately relate to the exponentiated variables. We will focus on this model and present JLMS measures only for the exponentiated inefficiency terms. This is also a way to contain to some degree the sprawling mass of mathematical expressions, since we will develop in detail three different models.

In Section 2 we present the distribution for the one-sided error components that we will use, and provide its main properties. In Section 3, we address certain concerns that are often raised when new specifications for SF models are proposed. In Sections 4 and 5 we present the production and cost SF models respectively, and in Section 6 the two-tier stochastic frontier model. Section 7 concludes with

simulations that explore how the familiar NHN and NE specifications perform when the data come from an NGE process and vice-versa, but also, how the conditional expectation and the conditional mode fare as measures of individual inefficiency.

2 The Generalized Exponential distribution

We consider the distribution that has the following density:

$$f_u(u) = \frac{2}{\theta_u} \exp\{-u/\theta_u\} (1 - \exp\{-u/\theta_u\}), \quad (1)$$

$$\theta_u > 0, u \geq 0.$$

Note that the density is two times an Exponential density times the Exponential distribution function with the same scale parameter. Let $h_E(u; \theta_u)$ denote the Exponential density with scale parameter θ_u . Then we can write equivalently,

$$f_u(u) = 2h_E(u; \theta_u) - h_E(u; \theta_u/2). \quad (2)$$

This additive form will prove convenient in calculating expressions that involve integration, exploiting the linearity of integrals and the already known results from the NE specification. The distribution function is

$$F_u(u) = (1 - \exp\{-u/\theta_u\})^2. \quad (3)$$

There are at least three ways to obtain the above distribution. First, as a general consequence of the Probability Integral Transform that states that for every continuous random variable X with support S_X , distribution function $F_X(x)$ and density $f_X(x)$, we have that $F_X(X) \sim U(0, 1)$. This then implies that

$$E[F_X(X)] = \int_{S_X} f_X(x) F_X(x) dx = \frac{1}{2}$$

$$\Rightarrow \int_{S_X} 2f_X(x) F_X(x) dx = 1.$$

So the function $2f_X(x)F_X(x)$ is non-negative in the support of X and integrates to unity over it, therefore it is a density. We take this general result and apply it to the Exponential distribution.

Second, $2f_X(x)F_X(x)$ is the density of the maximum of two i.i.d. random variables—indeed, since its distribution function is $[F_X(x)]^2$. This representation provides a straightforward way of generating draws from the distribution for simulation purposes.

Third, it can be seen as a special case of the “Generalized Exponential” distribution introduced by Gupta and Kundu (1999), with shape parameter equal to 2 (their α), scale parameter equal to θ_u (their λ) and location parameter equal to zero (their μ). We will write $u \sim GE(2, \theta_u, 0)$, and use this name to identify our distribution.

³ This approach was first presented in Materov (1981), that is written in Russian and it is a good example of the universality of the language of mathematics.

2.1 Moments and other properties

Yet another representation of this distribution, based on results from Gupta and Kundu (1999) and for the specific values of the parameters, is the following: let e_i , $i = 1, 2$ be two i.i.d. Exponentials with scale parameter equal to 1. Then

$$u \sim \theta_u e_1 + \theta_u e_2 / 2.$$

This makes the derivation of the basic moments of u very simple by using cumulants κ , for which we have $\kappa_r(cz) = c^r \kappa_r(z)$, and under independence, $\kappa_r(e_1 + e_2) = \kappa_r(e_1) + \kappa_r(e_2)$. It follows that

$$\kappa_r(u) = (1 + 2^{-r}) \kappa_r(e) \theta_u^r,$$

$$\kappa_1(e) = \kappa_2(e) = 1, \quad \kappa_3(e) = 2, \quad \kappa_4(e) = 6.$$

Then

$$E(u) = \kappa_1(u) = \frac{3}{2} \theta_u, \quad \text{Var}(u) = \kappa_2(u) = \frac{5}{4} \theta_u^2,$$

$$\text{Skewness}(u) = \gamma_1(u) = \frac{\kappa_3(u)}{\kappa_2^{3/2}(u)} = \frac{18}{\sqrt{125}} \approx 1.61,$$

$$\text{Ex. Kurtosis}(u) = \gamma_2(u) = \frac{\kappa_4(u)}{\kappa_2^2(u)} = \frac{102}{25} = 4.08.$$

We see that the distribution has lower skewness and excess kurtosis compared to the Exponential distribution (for which they are 2 and 6 respectively), but exceeds in both the Half-Normal (0.995 and 0.869 respectively). Paired with a Normal error component, these values represent the maximum skewness and kurtosis the composed SF error can accommodate under an NGE specification. Papadopoulos and Parmeter (2021) presented the empirical skewness and kurtosis of the OLS residuals for eight representative SF empirical studies, and in none of them did they exceeded the values of the GE distribution. This is an indication that although lower than the corresponding values for the Exponential distribution, they do not restrict in practice the applicability of the GE distribution.

Further, it is a simple exercise to find the argmax of the density,

$$\text{mode}(u) = \theta_u \ln 2.$$

The mode of this distribution is equal to the *median* of an Exponential distribution with the same scale parameter, and it locates the point where 0.25 of probability mass lies to the left of it. In other words, it is equal to the 1st quartile. We also see that the mode cannot be zero, and in that sense the labeling of the distribution as “Generalized Exponential” could be considered as a misnomer, since it does not nest the Exponential.

The quantile function is

$$Q_u(p) = -\theta_u \ln(1 - \sqrt{p}), \quad p \in [0, 1),$$

from which we can obtain the median and other quantiles.

Finally, by using the additive expression eq. (2) for the density, it is easy to determine that it is log-concave (and therefore so is its distribution function). Since we will combine this distribution with a Normal random-noise component, the resulting composite error density will also be log-concave, since the Normal density is log-concave, and convolutions of log-concave functions retain the property.

3 Why use yet another distributional specification?

Now that we have familiarized ourselves with the aspiring newcomer, it is time to confront certain issues that are routinely raised in relation to distributional specifications like the one we propose here: a fully parametric specification with a log-concave density.

The first issue is the risk that, in case of distributional *mis*-specification, inference would be unreliable. But this is an argument in favor of abandoning parametric inference altogether. For those scholars that find a net benefit in using it though, increasing the number of available specifications mitigates this problem since it increases the diversity of available models and so our ability to get within tolerable distance from the true data generating process (DGP).

Still under the spectre of distributional misspecification, we can avoid this particular risk when “determinants of inefficiency” are available. Then we can model the distribution parameters of the inefficiency components as functions of observed data (as long as a valid economic argument supports it). Moreover, one can exploit the “scaling property” (see Wang and Schmidt, 2002, for the single-tier and Parmeter, 2018, for the two-tier SF model respectively), that always holds for single-parameter distributions, and, instead of using maximum likelihood estimation, implement non-linear least squares (NLLS) that does not require distributional assumptions. An issue in this approach is the finite-sample reliability of the NLLS estimates: in his simulations, Parmeter (2018) has found a persistent upward bias in the estimates of the coefficients for the determinants of inefficiency for small and medium samples, which would lead to an *under*-estimation of technical efficiency, if we were to calculate the corresponding exponentiated measure: firms appear less efficient than they actually are.⁴

⁴ The under-estimation of technical efficiency here depends on the determinants of inefficiency being positive variables, and they usually are.

A third issue relates to all SF models that assume a log-concave density for the random noise component: in standard notation, in an SF production model we would have a composite error term $\varepsilon = v - u$, $E(v) = 0$, $u \geq 0$. Ondrich and Ruggiero (2001) have proven that, if the random noise component has a log-concave density, then the expected value of the inefficiency term conditional on the composite error, $E(u|\varepsilon)$, is a monotonic function of the conditioning variable.⁵ This has the implication that as regards ranking of the observations in the sample with respect to inefficiency, we might as well use the estimation residuals instead of $E(u|\varepsilon)$, and anticipate negligible differences (under correct specification). Moreover, the OLS residuals are more robust in ranking observations according to inefficiency, since they are free of possible misspecification from distributional assumptions.

We respond here by noting that ranking may be important and more safe as an inferential result, but there is more to efficiency analysis than a beauty contest to congratulate the more efficient firms and frown at the less efficient ones. The actual quantitative measurement of inefficiency is what matters as regards the efficient use of scarce resources, and for this we need to go beyond the estimation residuals.

In any case, the result of Ondrich and Ruggiero (2001) relates to the conditional expectation. We will provide proof that it holds also for the conditional mode of the production NGE frontier, and we will obtain simulation evidence as regards differences in ranking.

From a technical point of view, the NGE specification responds to the Stevenson (1980) critique of the NHN and NE specifications by offering a more convenient and, let’s say a more “user-friendly” solution to the non-zero mode desideratum, compared to the specifications involving the Truncated Normal or the Gamma distributions: in single-tier SF models the unknown parameters of the NGE specification are only two and not three, while in the 2TSF NGE model they are only three and not five. Moreover, the fixed investment cost in time and intellectual energy that the NGE specification requires (as any new tool does) is modest.

A weakness of the GE distribution is that it does not nest the zero-mode case. This in theory is an undesirable inflexibility, but the existence of a non-zero mode for the one-sided component is mostly an issue for economic debate rather than a statistical matter. We may not “let the data decide”, but knowledge of the particular industry/market under study in each case should allow for a convincing argument and a safe choice on this matter.

From a statistical point of view, the NGE specification, in addition of having a non-zero mode, is characterized by

higher values of skewness and excess kurtosis in the composed error term compared to the NHN specification, and so it can accommodate a larger set of real-world data samples.

Finally, from an economics point of view, by representing the maximum of two i.i.d. underlying random variables it allows us to picture the economic mechanism as operating at high intensity: if notionally there are two possible inefficiency values, the GE specification “selects” the strongest one to affect the outcome. To some, this may appear as unjustified negative bias; to others it would be classified as prudential, connected informally with the concept of entropy and the tendency of things to get to a lower state of energy, which in business terms translates into disorganization and loss of efficiency.

4 The production frontier

In a production frontier setting, the original model is

$$y_i = f(\mathbf{x}_i) \exp\{v_i - u_i\}, \quad i = 1, \dots, n.$$

Typically, y is some measure of output and the regressors are production inputs. We focus on its estimation in logarithms, so the composite error term is $\varepsilon = v - u$ and we assume $v \sim N(0, \sigma_v)$, $u \sim GE(2, \theta_u, 0)$.

The additive representation of the density of u , eq. (2), means that the density of ε can be obtained as the difference of two convolutions of the Normal density with the Exponential, each being nothing else than the familiar Normal-Exponential SF specification, (see e.g. Kumbhakar and Lovell 2000, for the density formula). So we can easily obtain

$$f_\varepsilon(\varepsilon) = \frac{2}{\theta_u} [\exp\{a_u\} \Phi(b_u) - \exp\{2a_u + \frac{\sigma_v^2}{\theta_u^2}\} \Phi(b_u - \frac{\sigma_v}{\theta_u})], \tag{4}$$

where Φ is the standard Normal distribution function, and

$$a_u = \frac{\varepsilon}{\theta_u} + \frac{\sigma_v^2}{2\theta_u^2}, \quad b_u = -\left(\frac{\varepsilon}{\sigma_v} + \frac{\sigma_v}{\theta_u}\right).$$

This formula is nested in the formulas for the 2TSF model presented in Section 6, by setting there $\theta_w = 0$. The same holds for the distribution function, which in the production frontier case is

$$F_\varepsilon(\varepsilon) = \Phi\left(\frac{\varepsilon}{\sigma_v}\right) + 2 \exp\{a_u\} \Phi(b_u) - \exp\left\{2a_u + \frac{\sigma_v^2}{\theta_u^2}\right\} \Phi\left(b_u - \frac{\sigma_v}{\theta_u}\right). \tag{5}$$

⁵ The authors proved this for the variable in levels. Using intuition but also the formal results in Egozcue (2015), it also holds for $E(\exp\{-u\}|\varepsilon)$.

The distribution function will be needed in modeling sample selection bias, and also in models where regressor endogeneity is handled using Copulas rather than instrumental variables (see Tran and Tsionas 2015; Papadopoulos 2020a).

Due to the independence of the error components, the mean and variance of the composed error term are immediately obtained. Regarding skewness and excess kurtosis, Papadopoulos and Parmeter (2021) have shown that they can be expressed as follows:

$$\gamma_1(\varepsilon) = -\gamma_1(u) \left(\frac{s_u}{s_\varepsilon}\right)^3, \quad \gamma_2(\varepsilon) = \gamma_2(u) \left(\frac{s_u}{s_\varepsilon}\right)^4,$$

where the symbol s represents the standard deviation. The skewness expression depends on the assumption that the distribution of v is symmetric, while the excess kurtosis expression on the assumption that it is Normal. Their maximum values (in absolute terms) are the skewness and excess kurtosis of the one-sided component. The authors present a powerful specification test using only sample moments of OLS residuals that is suitable for distributions with constant skewness and excess kurtosis.

Past this stage, the density can be used to implement maximum likelihood estimation. Alternatively, one can apply the Corrected Least Squares estimator (COLS), equating in the first stage the sample means of the 2nd and 3rd power of the OLS residuals with the 2nd and 3rd cumulant of the composite error term respectively, which are,

$$\kappa_2(\varepsilon) = \sigma_v^2 + \frac{5}{4}\theta_u^2, \quad \kappa_3(\varepsilon) = -\frac{9}{4}\theta_u^3.$$

We note that the COLS estimator is vulnerable to the “wrong skew” issue: the case where the OLS residuals exhibit skew with the opposite sign than the one assumed. Here if $\widehat{\kappa}_3(\varepsilon) > 0$, we cannot compute θ_u which is constrained to be strictly positive. On the other hand, the simulations presented later showed that the MLE under the NGE specification does not break down when the skew is wrong: in such cases, the value of θ_u is estimated by MLE as really small but away from zero, around the values 0.03–0.04. But these are not reliable estimates, and so if the sample skew is wrong and the researcher decides to treat it as a sample problem, special treatment is required (for different such methods see Cai et al. 2020; Hafner et al. 2018; Simar and Wilson 2009).

4.1 Assessing and measuring technical (in)efficiency

For the single-output production frontier model estimated in logarithms, the standard measure of technical efficiency is Shephard’s output distance function that here equals $\exp\{-u\} \equiv q_u, q_u \in (0, 1]$ (see Sickles and Zelenyuk 2019,

pp. 20–24), which is the ratio of actual to maximum output. This is a random variable on its own. To obtain its marginal distribution, we apply the distribution method and we have

$$F_{q_u}(q_u) = \Pr(\exp\{-u\} \leq q_u) = \Pr(u \geq -\ln q_u) = 1 - F_u(-\ln q_u) \Rightarrow F_{q_u}(q_u) = 1 - \left(1 - q_u^{1/\theta_u}\right)^2, \tag{6}$$

while the density is

$$f_{q_u}(q_u) = \frac{2}{\theta_u} \left(q_u^{-1+1/\theta_u} - q_u^{-1+2/\theta_u}\right). \tag{7}$$

This is the distribution of the *minimum* of two i.i.d. Beta random variables, with parameters $\alpha = 1/\theta_u, \beta = 1$. Unlike what happens to the Beta distribution when the parameter β is equal to unity, here, the distribution does not have always a monotonic density. This will depend on the actual value of θ_u . Specifically we have

$$\text{mode}[\exp\{-u\}] = \begin{cases} [(1 - \theta_u)/(2 - \theta_u)]^{\theta_u} & \theta_u < 1 \\ \sim 0 & \theta_u \geq 1. \end{cases}$$

We note that except when $\theta_u = 1$ where the distribution becomes the minimum of two i.i.d $U(0, 1)$ random variables, for $\theta_u > 1$ the density has an asymptote at zero. Also, note that the zero-mode result for $\theta_u \geq 1$ is for the variable $\exp\{-u\}$, so with stronger inefficiency parameter (higher θ_u) comes lower technical efficiency.

Having the distribution function available allows us also to compute a full representative spectrum of quantiles (not just the median), through the quantile function

$$Q_q(p) = \left(1 - \sqrt{1 - p}\right)^{\theta_u}, \quad p \in [0, 1].$$

Turning to moments, by using the additive expression for the $f_u(u)$ density, the unconditional expected value can be calculated through the moment generating function of the Exponential distribution and it is

$$E(\exp\{-u\}) = \frac{2}{(1 + \theta_u)(2 + \theta_u)}.$$

We can obtain its variance from

$$\text{Var}(\exp\{-u\}) = E[\exp\{-2u\}] - [E(\exp\{-u\})]^2.$$

The mirror property of the Beta distribution is inherited here, so if one wants to study the relative technical inefficiency $1 - \exp\{-u\}$, the distribution will be that of the minimum of two i.i.d. Betas with parameters $\alpha = 1, \beta = 1/\theta_u$.

Closing this section, we mention that the ratio of maximum output over actual, $\exp\{u\}$, is the “output-oriented Farrell measure of technical efficiency”, and for businesses it is a meaningful measure and perhaps more meaningful to

their management, although not as a measure of efficiency but as a guidance for efficiency improvement: a value, say, of $E(\exp\{u\}) = 1.15$ would tell us that on average, firms in the sample need to increase their output by 15%, in order to be fully efficient, holding inputs constant. At the level of an individual firm, such a measure would be more relatable to the business mindset, and more directly translatable to specific actions in the pursuit of efficiency (“with given inputs, find new ways to combine and coordinate them better so as to increase output—and you have a 15% increase to chase”). To our knowledge it has not been explored in the efficiency literature and we will not examine it further here. We just note that the results on the marginal distribution of the variable “ k ” in Section 5.1 apply also for the $\exp\{u\}$ variable (but not the individual measures of Section 5.2).

4.2 Individual measures

We can obtain information on $q_{u,i} = \exp\{-u_i\}$, $i = 1, \dots, n$ through ε_i . By standard techniques, the conditional density $f_{u|\varepsilon}(u_i|\varepsilon_i)$ is

$$f_{u|\varepsilon}(u_i|\varepsilon_i) = \frac{f_{v,u}(\varepsilon_i + u_i, u_i)}{f_\varepsilon(\varepsilon_i)} = \frac{\phi((\varepsilon_i + u_i)/\sigma_v)f_u(u_i)}{\sigma_v f_\varepsilon(\varepsilon_i)},$$

where ϕ is the standard Normal density. We show in the Appendix that

$$E(\exp\{-u\}|\varepsilon) = \frac{2}{\theta_u f_\varepsilon(\varepsilon)} \exp\left\{\frac{1+\theta_u}{\theta_u} \varepsilon + \frac{\sigma_v^2}{2} \left(\frac{1+\theta_u}{\theta_u}\right)^2\right\} \times \left[\Phi\left(-\frac{\varepsilon}{\sigma_v} - \sigma_v \left(\frac{1+\theta_u}{\theta_u}\right)\right) - \exp\left\{\frac{\varepsilon}{\theta_u} + \frac{\sigma_v^2}{\theta_u^2} \left(\theta_u + \frac{3}{2}\right)\right\} \Phi\left(-\frac{\varepsilon}{\sigma_v} - \sigma_v \left(\frac{2+\theta_u}{\theta_u}\right)\right)\right]. \tag{8}$$

We will use this expected value in our simulations alongside the conditional mode as predictors of inefficiency (we won’t present formulas for conditional expectations for the other models that we deal with here).

Turning to the conditional mode, in order to compute it we need the conditional density of $f_{q_u|\varepsilon}(q_u|\varepsilon_i)$ which we can readily obtain by applying a change-of-variables in $f_{u|\varepsilon}(u_i|\varepsilon_i)$ to arrive at

$$f_{q_u|\varepsilon}(q_u|\varepsilon_i) = \frac{\phi((\varepsilon_i - \ln q_u)/\sigma_v)f_u(q_u)}{\sigma_v f_\varepsilon(\varepsilon_i)}. \tag{9}$$

This can be maximized numerically with respect to q_u (for each estimated value $\hat{\varepsilon}_i$), to get the conditional mode as an alternative predictor of individual efficiency. Note that this requires maximizing only the numerator in eq. (9), since the denominator does not include the q_u variable.

4.2.1 Monotonicity of the conditional mode in the conditioning variable

We show in the Appendix that the conditional mode $q_{u,*}$ in the NGE production frontier satisfies the following implicit equation:

$$q_{u,*} : \left(q_{u,*}^{-1/\theta_u} - 1\right) \cdot \left[\frac{\varepsilon_i - \ln q_{u,*}}{\sigma_v^2} - 1 + 1/\theta_u\right] - 1/\theta_u = 0.$$

The variable q_u ranges in (0, 1]. This implies that the term in parenthesis is always positive, meaning that the LHS expression is increasing in ε_i . Also, it is evident that both the term in parenthesis and the term in brackets is decreasing in $q_{u,*}$. So if we increase the value of ε_i the LHS will tend to increase, and in order to restore equality with zero we must decrease the LHS and so *increase* $q_{u,*}$. But this proves that the conditional mode is monotonically increasing in the conditioning variable. This shows that the result of Ondrich and Ruggiero (2001) mentioned earlier related to conditional expectations extends to the conditional mode of the NGE production specification.

5 The cost frontier

In a cost frontier setting, the original equation is of the form

$$y_i = f(\mathbf{x}_i) \exp\{v_i + w_i\}, \quad i = 1, \dots, n.$$

Here the dependent variable represents production costs, and the regressors are typically input prices and output, and again we focus on its estimation in logarithms. So the composite error term is $\varepsilon = v + w$ and we assume $v \sim N(0, \sigma_v)$, $w \sim GE(2, \theta_w, 0)$.

The density and distribution function of the composite error term are respectively

$$f_\varepsilon(\varepsilon) = \frac{2}{\theta_w} [\exp\{a_w\} \Phi(b_w) - \exp\{2a_w + (\sigma_v/\theta_w)^2\} \Phi(b_w - \sigma_v/\theta_w)], \tag{10}$$

and

$$F_\varepsilon(\varepsilon) = \Phi\left(\frac{\varepsilon}{\sigma_v}\right) - 2 \exp\{a_w\} \Phi(b_w) + \exp\left\{2a_w + \frac{\sigma_v^2}{\theta_w^2}\right\} \Phi\left(b_w - \frac{\sigma_v}{\theta_w}\right), \tag{11}$$

with

$$a_w = \frac{\sigma_v^2}{2\theta_w^2} - \frac{\varepsilon}{\theta_w}, \quad b_w = \frac{\varepsilon}{\sigma_v} - \frac{\sigma_v}{\theta_w}.$$

As before these can be used in maximum likelihood estimation, while we can also implement the COLS estimator (only, $\kappa_3(\varepsilon)$ would be positive here).

5.1 Assessing and measuring cost (in)efficiency

The standard measure of cost efficiency, is the ratio of minimum cost to actual, in our setting $CE \equiv q_w = \exp\{-w\} \in (0, 1]$ (see Sickles and Zelenyuk 2019 pp. 80–81). The measure enjoys certain theoretical desirable properties, not least being confined in the $(0, 1]$ interval.⁶ For CE all the formulas from the *marginal* analysis of the section on the production frontier applies as is (Section 4.1), inserting everywhere w in place of u .

Somewhat more intuitive but less well-behaved would be the ratio of *actual cost to minimum*, which we could think of as a measure of cost *inefficiency*. This would give us a “gross inefficiency mark up”: a value, say, 1.25 of this ratio would tell us that actual costs are 25% higher than minimum. In our setting this measure is represented by $k = \exp\{w\}$, $k \in [1, \infty)$. Here we have

$$F_k(k) = \Pr(\exp\{w\} \leq k) = \Pr(w \leq \ln k) = F_w(\ln k)$$

$$\Rightarrow F_k(k) = (1 - k^{-1/\theta_w})^2.$$

This is the distribution of the maximum of two i.i.d Pareto random variables (of “Type I”), with minimum value/scale parameter equal to 1, and shape parameter $1/\theta_w$. Its density is

$$f_k(k) = \frac{2}{\theta_w} (k^{-1-1/\theta_w} - k^{-1-2/\theta_w}).$$

The mode of this distribution is

$$\text{mode}(\exp\{w\}) = \left(\frac{2 + \theta_w}{1 + \theta_w}\right)^{\theta_w},$$

while its quantile function is

$$Q_k(p) = (1 - \sqrt{p})^{-\theta_w}, \quad p \in [0, 1].$$

Here, the existence of moments depends on the value of θ_w : as it is the case for the Pareto distribution itself, for the r -th moment of k to exist and be finite, we must have $\theta_w < 1/r$. So if $\theta_w \geq 1$ not even the mean will exist, while to have a variance we would need $\theta_w < 1/2$.

⁶ But it is not immediate to translate it in intuitive terms, because, here, improving cost efficiency means reducing the denominator of the ratio. In fact the value $1 - CE = 1 - \exp\{-w\}$ is more naturally communicated, being the proportional reduction in costs required to attain full cost efficiency. So if, say, $CE = 0.8$, we have to reduce costs by $1 - 0.8 = 20\%$ to reach the frontier, holding output constant.

In cases where $\theta_w < 1$, the mean of this distribution is given by

$$E(\exp\{w\}) = \frac{2}{(1 - \theta_w)(2 - \theta_w)},$$

and if moreover $\theta_w < 1/2$, the variance can be obtained from

$$\text{Var}(\exp\{w\}) = E[\exp\{2w\}] - [E(\exp\{w\})]^2.$$

The *net* mark-up on costs due to inefficiency is $\exp\{w\} - 1$, and this is the maximum of two i.i.d random variables following the Lomax distribution with scale parameter equal to unity (the Lomax distribution is essentially a Pareto law shifted so as its support starts at zero).

5.2 Individual measures

For the cost efficiency measure $CE = \exp\{-w\}$ the conditional density is

$$f_{q_w|\varepsilon}(q_w|\varepsilon_i) = \frac{\phi((\varepsilon_i + \ln q_w)/\sigma_v)f_{q_w}(q_w)}{\sigma_v f_\varepsilon(\varepsilon_i)}. \tag{12}$$

Turning to the cost inefficiency measure $k = \exp\{w\}$, we note that if $\theta_w > 1$ and the marginal distribution does not possess moments, we are left only with the conditional mode as an individual measure. This is because the conditional expectation is fundamentally defined only for variables that are absolutely integrable and have finite expected value (see e.g. Williams 1991, Theorem 9.2, p. 84). The conditional density here is

$$f_{k|\varepsilon}(k|\varepsilon_i) = \frac{\phi((\varepsilon_i - \ln k)/\sigma_v)f_k(k)}{\sigma_v f_\varepsilon(\varepsilon_i)}. \tag{13}$$

Numerical maximization of the numerator in both expressions with respect to k gives the conditional mode, for given ε_i .

6 The two-tier frontier⁷

The two-tier stochastic frontier (2TSF) model was introduced by Polachek and Yoon (1987) in order to measure informational inefficiency of both employers and employees related to the determination of the wage. Since then it has been applied to many other markets apart from the labor market, notably the housing market and the health services market, but also as a method to measure bargaining power in a bilateral bargaining setting (see Papadopoulos 2020c, for a comprehensive survey).

⁷ The mathematical derivations of this section can be found in the Technical Appendix of Papadopoulos (2018, pp. 409–444).

The model is represented by the equation

$$y_i = f(\mathbf{x}_i) \exp\{\varepsilon_i\}, \quad \varepsilon_i = v_i + w_i - u_i, \quad i = 1, \dots, n,$$

and as before we assume that it will be estimated in logarithmic form. The error components follow $v \sim N(0, \sigma_v)$, $w \sim GE(2, \theta_w, 0)$, $u \sim GE(2, \theta_u, 0)$, and are jointly independent. Then we have,

$$f_\varepsilon(\varepsilon) = \frac{2}{\theta_w + \theta_u} \left[\frac{2\theta_u \exp\{a_u\} \Phi(b_u)}{\theta_w + 2\theta_u} - \frac{\theta_u \exp\{2a_u + (\sigma_v/\theta_u)^2\} \Phi(b_u - \sigma_v/\theta_u)}{2\theta_w + \theta_u} + \frac{2\theta_w \exp\{a_w\} \Phi(b_w)}{2\theta_w + \theta_u} - \frac{\theta_w \exp\{2a_w + (\sigma_v/\theta_w)^2\} \Phi(b_w - \sigma_v/\theta_w)}{\theta_w + 2\theta_u} \right]. \tag{14}$$

The distribution function is

$$F_\varepsilon(\varepsilon) = \frac{2}{\theta_w + \theta_u} \left[\frac{2(\theta_w + \theta_u)^3 + \theta_w \theta_u (\theta_w + \theta_u)}{2(\theta_w + 2\theta_u)(2\theta_w + \theta_u)} \Phi\left(\frac{\varepsilon}{\sigma_v}\right) + \frac{2\theta_u^2}{\theta_w + 2\theta_u} \exp\{a_u\} \Phi(b_u) - \frac{\theta_u^2}{2(2\theta_w + \theta_u)} \exp\left\{2a_u + \frac{\sigma_v^2}{\theta_u^2}\right\} \Phi\left(b_u - \frac{\sigma_v}{\theta_u}\right) - \frac{2\theta_w^2}{2\theta_w + \theta_u} \exp\{a_w\} \Phi(b_w) + \frac{\theta_w^2 \exp\{2a_w + \sigma_v^2/\theta_w^2\} \Phi\left(b_w - \frac{\sigma_v}{\theta_w}\right)}{2(\theta_w + 2\theta_u)} \right]. \tag{15}$$

It is straightforward to obtain the moments of ε using the cumulant expressions, since the three components are independent. We have

$$\kappa_2(\varepsilon) = \sigma_v^2 + \frac{5}{4}(\theta_w^2 + \theta_u^2), \quad \kappa_3(\varepsilon) = \frac{9}{4}(\theta_w^3 - \theta_u^3), \\ \kappa_4(\varepsilon) = \frac{51}{8}(\theta_w^4 + \theta_u^4).$$

Implementing the COLS estimator here requires also the use of the sample mean of the 4th power of the OLS residuals that estimates consistently the 4th central moment of the composed error term, and it must be set equal to $\kappa_4(\varepsilon) + 3\kappa_2^2(\varepsilon)$.

The two-tier frontier specification faces no problems with “wrong skew” samples, because its theoretical skewness can be positive or negative. Papadopoulos (2020b) develops a 2TSF model for production in order

to measure the effects that unobservable management has on output (management being represented by the error component w).

6.1 Assessing and measuring the net effect

To analyze each one-sided component individually, we can use the results from the single-tier frontiers as regards the marginal distributions, but not the results on individual measures, since these depend on the composite error term which here is different.

But specific to the 2TSF model is the joint presence of opposing forces w and u on the outcome, and so of special interest is their net effect $z = w - u$, and, for the logarithmic model we are considering, the exponentiated one, $\xi = \exp\{w - u\}$. After deriving the density and distribution function of z , which we do not present here, we can obtain the density and distribution function of ξ with the same techniques as before,

$$F_\xi(\xi) = \begin{cases} \frac{\theta_u^2}{\theta_w + \theta_u} \left[\frac{4\xi^{1/\theta_u}}{\theta_w + 2\theta_u} - \frac{\xi^{2/\theta_u}}{2\theta_w + \theta_u} \right] & \xi \leq 1 \\ 1 - \frac{\theta_w^2}{\theta_w + \theta_u} \left[\frac{4\xi^{-1/\theta_w}}{2\theta_w + \theta_u} - \frac{\xi^{-2/\theta_w}}{\theta_w + 2\theta_u} \right] & \xi > 1. \end{cases} \tag{16}$$

From this, one can obtain probabilistic events of interest like for example,

$$\Pr(\exp\{w - u\} \leq 1) = \frac{\theta_u^2}{\theta_w + \theta_u} \left(\frac{4}{\theta_w + 2\theta_u} - \frac{1}{2\theta_w + \theta_u} \right).$$

The marginal density of ξ is

$$f_\xi(\xi) = \frac{2}{\theta_w + \theta_u} \times \begin{cases} \theta_u \cdot \left[\frac{2\xi^{-1+1/\theta_u}}{\theta_w + 2\theta_u} - \frac{\xi^{-1+2/\theta_u}}{2\theta_w + \theta_u} \right] & 0 < \xi \leq 1 \\ \theta_w \cdot \left[\frac{2\xi^{-1-1/\theta_w}}{2\theta_w + \theta_u} - \frac{\xi^{-1-2/\theta_w}}{\theta_w + 2\theta_u} \right] & \xi > 1. \end{cases} \tag{17}$$

The mode of this density is

$$\text{mode}(\exp\{w - u\}) = \begin{cases} 0 & \theta_u \geq 1 \\ \max\{q_0 I\{q_0 \leq 1\}, q_1 I\{q_1 > 1\}\} & \theta_u < 1, \end{cases} \tag{18}$$

with

$$q_0 = \left(\frac{1 - \theta_u}{2 - \theta_u} \cdot \frac{4\theta_w + 2\theta_u}{\theta_w + 2\theta_u} \right)^{\theta_u}, \quad q_1 = \left(\frac{2 + \theta_w}{1 + \theta_w} \cdot \frac{2\theta_w + \theta_u}{2\theta_w + 4\theta_u} \right)^{\theta_w},$$

and where $I\{\cdot\}$ is the indicator function. One can determine that

$$\text{mode}(\exp\{w - u\}) \neq \text{mode}(\exp\{w\}) \text{mode}(\exp\{-u\}).$$

The proper predictor of the net effect of the two opposing one-sided components is the mode of the difference, not the difference of the marginal modes, since, even if they are independent, the two error components happen concurrently and so we must consider their variation jointly. But the marginal modes are also useful since they predict the value of each component, *if it operated without the presence of the other*.

As regards expected values, due to independence we have,

$$E[\exp\{w - u\}] = E(\exp\{w\})E(\exp\{-u\}),$$

and it will exist if $\theta_w < 1$.

6.2 Individual measures

In the 2TSF setting, of main interest are the variables

$$k = \exp\{w\}, \quad q_w = \exp\{-w\},$$

$$q_u = \exp\{-u\}, \quad \xi = \exp\{w - u\}.$$

Closed-form expressions for the one-sided conditional expectations for the 2TSF NE and NHN specifications have been collected in Papadopoulos (2018, ch.3). Up to now no attention has been given to the conditional mode as a predictor at the individual level. For the NGE specification, closed-formed expressions for conditional expectations can be derived, but they are much more involved and lengthy than for the single-tier models. We also note that even though w and u are assumed independent, they stop being so conditional on ε , so

$$E[\exp\{w - u\}|\varepsilon] \neq E[\exp\{w\}|\varepsilon]E[\exp\{-u\}|\varepsilon].$$

To obtain predictors based on the mode we need four conditional densities. These are

$$f_{k|\varepsilon}(k|\varepsilon) = \frac{2f_k(k)}{\theta_w f_\varepsilon(\varepsilon)} \times \left[k^{-1/\theta_u} \exp\{a_u\} \Phi\left(\frac{\ln k}{\sigma_v} + b_u\right) - k^{-2/\theta_u} \exp\left\{2a_u + \frac{\sigma_v^2}{\theta_u^2}\right\} \Phi\left(\frac{\ln k}{\sigma_v} + b_u - \frac{\sigma_v}{\theta_u}\right) \right]. \tag{19}$$

$$f_{q_w|\varepsilon}(q_w|\varepsilon) = \frac{2f_{q_w}(q_w)}{\theta_w f_\varepsilon(\varepsilon)} \times \left[q_w^{1/\theta_u} \exp\{a_u\} \Phi\left(\frac{-\ln q_w}{\sigma_v} + b_u\right) - q_w^{2/\theta_u} \exp\left\{2a_u + \frac{\sigma_v^2}{\theta_u^2}\right\} \Phi\left(\frac{-\ln q_w}{\sigma_v} + b_u - \frac{\sigma_v}{\theta_u}\right) \right], \tag{20}$$

where $f_{q_w}(q_w)$ is $f_{q_u}(q_u)$ (eq. (7)) with the symbol w in place of the symbol u . Also,

$$f_{q_u|\varepsilon}(q_u|\varepsilon) = \frac{2f_{q_u}(q_u)}{\theta_w f_\varepsilon(\varepsilon)} \times \left[q_u^{1/\theta_w} \exp\{a_w\} \Phi\left(\frac{-\ln q_u}{\sigma_v} + b_w\right) - q_u^{2/\theta_w} \exp\left\{2a_w + \frac{\sigma_v^2}{\theta_w^2}\right\} \Phi\left(\frac{-\ln q_u}{\sigma_v} + b_w - \frac{\sigma_v}{\theta_w}\right) \right], \tag{21}$$

and finally,

$$f_{\xi|\varepsilon}(\xi|\varepsilon) = \frac{\phi((\varepsilon - \ln \xi)/\sigma_v)}{\sigma_w f_\varepsilon(\varepsilon)} \times f_\xi(\xi). \tag{22}$$

As before, obtaining the conditional mode series requires numerical maximization of these expressions with respect to their variable, for each estimated value of ε_i .

7 Simulation studies

In this section we provide simulation results related to the NGE production frontier model, using maximum likelihood estimation.

In all simulations the regression equation is

$$y = 1 + x_1 + x_2 + \varepsilon, \quad \varepsilon = v - u,$$

with $x_1 \sim \chi^2_1$, $x_2 \sim \text{Bern}(0.65)$, $v \sim N(0, 1)$.

In each simulation we run 1000 replications and we report sample averages over them. We consider two sample sizes $n = 200, 1000$ and two different values of $\theta_u = 0.5, 1.5$. For each value we report also the “signal-to-noise” ratio (SNR), the ratio of the standard deviation of the inefficiency error component over that of the random component. This is a model-free measure that summarizes well how strong is the signal that interests us (the inefficiency), relative to noise. Note that it equals the familiar $\lambda = \sigma_u/\sigma_v$ only for the NE specification.

We observed that when the data generating process was NGE, the correctly specified MLE failed to converge 2–5% of the times, while when the sample size was small and the SNR large, this percentage rose to $\approx 15\%$. We noted earlier that this is not related to the “wrong skew” issue. Rather, it indicates a sensitivity of the NGE likelihood to starting values, and in empirical studies many different sets of initial values for the parameters should be tried.

7.1 Performance when the DGP is NGE

In the first part of our simulation study, the DGP has an NGE composite error term. We report OLS, NHN, and NE estimates also, to see how these fare when the true specification is NGE. Essentially we attempt to answer the

Table 1 Production frontier—Data generating process is NGE

$\theta = 0.5$		$n = 200$				$n = 1000$			
SNR = 0.559	TRUE	OLS	NHN	NE	NGE	OLS	NHN	NE	NGE
Const.	1	0.240	0.960	0.683	0.859	0.251	1.052	0.744	0.933
x_1 (cont.)	1	1.001	1.000	1.000	1.000	1.000	1.000	1.000	1.000
x_2 (discrete)	1	1.009	1.010	1.010	1.011	0.999	0.999	0.999	1.000
σ_v	1	–	0.943	1.018	0.999	–	0.954	1.024	1.008
$E(\epsilon)$	0.75	–	0.720	0.443	0.624	–	0.801	0.492	0.684
$Var(\epsilon)$	1.313	1.306	1.186	1.233	1.215	1.307	1.277	1.291	1.276
$\gamma_1(\epsilon)$	-0.187	-0.180	-0.124	-0.127	-0.121	-0.182	-0.153	-0.163	-0.148
$\gamma_2(\epsilon)$	0.231	0.164	0.054	0.152	0.129	0.204	0.071	0.212	0.169
$E(\exp\{-u\})$	0.533	–	0.551	0.693	0.585	–	0.522	0.670	0.559

$\theta = 1.5$		$n = 200$				$n = 1000$			
SNR = 1.677	TRUE	OLS	NHN	NE	NGE	OLS	NHN	NE	NGE
Const.	1	-1.245	0.942	0.346	0.972	-1.253	0.971	0.361	0.999
x_1 (cont.)	1	0.999	0.993	0.992	0.992	1.001	1.000	1.000	1.000
x_2 (discrete)	1	0.995	1.002	0.996	1.000	1.004	1.003	1.003	1.003
σ_v	1	–	0.881	1.112	0.979	–	0.894	1.125	0.995
$E(\epsilon)$	2.25	–	2.208	1.586	2.214	–	2.248	1.612	2.250
$Var(\epsilon)$	3.813	3.759	3.558	3.752	3.681	3.817	3.683	3.864	3.803
$\gamma_1(\epsilon)$	-1.020	-0.937	-0.688	-1.098	-1.024	-1.007	-0.689	-1.103	-1.024
$\gamma_2(\epsilon)$	2.220	1.763	0.531	2.697	2.232	2.154	0.533	2.713	2.231
$E(\exp\{-u\})$	0.229	–	0.260	0.387	0.232	–	0.256	0.383	0.229

True values are indicated in bold

question “suppose that the DGP is indeed NGE. Does it matter for inference?” We certainly expect that the correctly specified MLE will fare better compared to a misspecified likelihood, but how much better?

In order to have comparable results between specifications, we present not estimates of the scale parameters of the inefficiency component, but calculated moments based on these estimates and on the assumed specification in each case.

The results are presented in Table 1. They validate the robustness/consistency property for the Skew Normal quasi-MLE (the NHN specification) as regards the regression slope coefficients, proven in Papadopoulos and Parmeter (2020). The same robustness property appears to characterize the NE specification. Regarding the performance of the misspecified likelihoods, we see that the MLE under the NHN specification does remarkably well, perhaps against expectations, with rather small bias in the estimation of average Technical Efficiency. On the other hand, the NE specification leads to high overestimation of TE.

Probing deeper we computed also JLMS conditional expectations for the NHN and NGE models. In each replication we obtained the conditional expectation series, and computed the 0.2, 0.4, 0.6, 0.8 quintiles.

Table 2 Production frontier—Data generating process is NGE. Empirical medians of Quintiles of JLMS measures $E[\exp\{-u\}]$

cum. Prob	$\theta = 0.5/\text{SNR} = 0.559$				$\theta = 1.5/\text{SNR} = 1.677$			
	$n = 200$		$n = 1000$		$n = 200$		$n = 1000$	
	NHN	NGE	NHN	NGE	NHN	NGE	NHN	NGE
0.2	0.40	0.48	0.40	0.46	0.05	0.07	0.05	0.07
0.4	0.49	0.54	0.49	0.53	0.15	0.17	0.15	0.17
0.6	0.56	0.59	0.56	0.58	0.29	0.27	0.29	0.27
0.8	0.63	0.63	0.63	0.62	0.46	0.38	0.46	0.38

Table 2 presents the median estimate of each quintile over all replications. We observe that sample size does not appear to affect behavior. When the SNR is low, the predictor coming from the NHN misspecified model tends to underestimate Technical Efficiency: for example, it places 40% of firms at a Technical Efficiency level below 0.48–0.49 (1st column 2nd row), while the correctly specified NGE predictor does that for only 20% of firms (2nd column 1st row). When the SNR is large, the two predictors allocate probability mass essentially in the same way up to cumulative probability 0.6, but then the NHN tends to overestimate Technical Efficiency since its 0.8-quantile

Table 3 Production frontier—Data generating process is NHN

$\sigma_u = 0.927$		$n = 200$				$n = 1000$			
SNR = 0.559	TRUE	OLS	NHN	NE	NGE	OLS	NHN	NE	NGE
Const.	1	0.260	0.871	0.620	0.759	0.265	0.917	0.654	0.794
x_1 (cont.)	1	1.002	1.001	1.001	1.001	0.999	0.999	0.999	0.999
x_2 (discrete)	1	0.994	0.996	0.995	0.994	0.994	0.994	0.994	0.995
σ_v	1	–	0.979	1.048	1.034	–	1.009	1.066	1.056
$E(\epsilon)$	0.740	–	0.611	0.359	0.504	–	0.652	0.389	0.532
$Var(\epsilon)$	1.313	1.303	1.172	1.227	1.210	1.314	1.261	1.288	1.273
$\gamma_1(\epsilon)$	0.116	–0.119	–0.077	–0.068	–0.064	–0.113	–0.084	–0.081	–0.070
$\gamma_2(\epsilon)$	0.049	0.013	0.029	0.066	0.055	0.047	0.032	0.083	0.062
$E(\exp\{-u\})$	0.544	–	0.595	0.736	0.641	–	0.578	0.720	0.627

$\sigma_u = 2.782$		$n = 200$				$n = 1000$			
SNR = 1.677	TRUE	OLS	NHN	NE	NGE	OLS	NHN	NE	NGE
Const.	1	–1.202	0.971	0.267	0.901	–1.221	0.983	0.289	0.942
x_1 (cont.)	1	0.997	0.990	0.993	0.993	1.000	0.999	0.999	0.999
x_2 (discrete)	1	0.986	1.005	0.995	0.996	1.003	1.007	1.006	1.006
σ_v	1	–	0.974	1.276	1.137	–	0.997	1.291	1.153
$E(\epsilon)$	2.220	–	2.177	1.467	2.108	–	2.205	1.512	2.168
$Var(\epsilon)$	3.813	3.745	3.654	3.780	3.760	3.799	3.770	3.953	3.939
$\gamma_1(\epsilon)$	–0.631	–0.606	–0.634	–0.859	–0.856	–0.624	–0.629	–0.880	–0.868
$\gamma_2(\epsilon)$	0.473	0.422	0.477	1.945	1.757	0.449	0.471	2.007	1.791
$E(\exp\{-u\})$	0.259	–	0.263	0.405	0.244	–	0.260	0.398	0.237

True values are indicated in bold

value is 0.46 while the corresponding quantile value of the NGE predictor is only 0.38.

7.2 Performance when the DGP is NHN

Here the DGP includes a NHN composite error term, and we are interested to see how misleading could estimation be if it is carried based on a misspecified NGE likelihood.

The results from the NHN DGP are in Table 3. When the SNR is low, the now misspecified NGE model leads to an overestimation of Technical Efficiency, but not as much as the misspecified NE model. With high SNR, the NGE model performs comparably with the NHN model in the previous simulation, while the NE model continues to be highly biased.

7.3 Performance when the DGP is NE

Here the DGP includes a NE error term, and the goal is the same as in the previous subsection.

From Table 4 we see that the now misspecified NHN and NGE models perform comparably, and both underestimate the average Technical Efficiency measure.

Overall, there appears to be a certain degree of congruence and “mutual robustness” between the NGE and the NHN specifications, and exploring whether this has some deeper theoretical justification could be an interesting topic of study. When we misspecify the one against the other, results are not that different, although of course it is always preferable to use the correct specification, especially when one is interested not just in sample averages but individual estimates of inefficiency. On the other hand the misspecified NE likelihood visibly overestimates Technical Efficiency against both an NGE and a NHN data generating process, and these two return the favor by underestimating TE when the true model is NE.

7.4 Rankings and correlation

In this subsection we examine how close are the firm rankings obtained by the NGE predictor compared to the rankings based on OLS residuals. We included an additional sample size, $n = 5000$, and two additional values for $\theta_u = 1.0, 2.0$. We computed individual efficiency measures for the NGE specification, both conditional expectations and conditional modes (for the exponentiated variable). In Table 5, we compare the observation rankings obtained

Table 4 Production frontier—Data generating process is NE

$\sigma_u = 0.559$		$n = 200$				$n = 1000$			
SNR = 0.559	TRUE	OLS	NHN	NE	NGE	OLS	NHN	NE	NGE
Const.	1	0.446	1.220	0.938	1.139	0.440	1.316	0.990	1.217
x_1 (cont.)	1	1.002	1.000	1.000	1.000	1.001	1.000	1.000	1.000
x_2 (discrete)	1	0.993	0.996	0.995	0.996	0.999	0.999	0.999	0.999
σ_v	1	–	0.916	0.992	0.969	–	0.922	0.999	0.978
$E(\epsilon)$	0.559	–	0.775	0.492	0.695	–	0.877	0.550	0.777
$Var(\epsilon)$	1.313	1.297	1.181	1.226	1.208	1.311	1.290	1.300	1.292
$\gamma_1(\epsilon)$	–0.232	–0.230	–0.155	–0.175	–0.169	–0.235	–0.198	–0.224	–0.213
$\gamma_2(\epsilon)$	0.340	0.281	0.073	0.233	0.202	0.332	0.101	0.324	0.275
$E(\exp\{-u\})$	0.641	–	0.531	0.670	0.555	–	0.497	0.645	0.523

$\sigma = 1.677$		$n = 200$				$n = 1000$			
SNR = 1.677	TRUE	OLS	NHN	NE	NGE	OLS	NHN	NE	NGE
Const.	1	–0.673	1.553	0.993	1.625	–0.675	1.556	0.997	1.626
x_1 (cont.)	1	1.000	0.992	0.994	0.993	0.999	0.998	0.998	0.998
x_2 (discrete)	1	1.000	1.002	0.999	1.000	0.999	0.999	1.000	1.001
σ_v	1	–	0.763	0.978	0.837	–	0.792	0.996	0.865
$E(\epsilon)$	1.677	–	2.265	1.658	2.289	–	2.274	1.672	2.301
$Var(\epsilon)$	3.813	3.769	3.511	3.705	3.612	3.799	3.579	3.788	3.690
$\gamma_1(\epsilon)$	–1.267	–1.168	–0.758	–1.278	–1.165	–1.249	–0.745	–1.268	–1.146
$\gamma_2(\epsilon)$	3.265	2.644	0.605	3.303	2.650	3.155	0.591	3.267	2.593
$E(\exp\{-u\})$	0.374	–	0.255	0.376	0.225	–	0.254	0.374	0.223

True values are indicated in bold

Table 5 Production frontier—NGE ranking changes compared to OLS ranking

	$E(\exp\{-u\})$	mode($\exp\{-u\})$	$E(\exp\{-u\})$	mode($\exp\{-u\})$
$n = 200$	$\theta = 0.5$	SNR = 0.559	$\theta = 1$	SNR = 1.118
% of rank changes	48.4%	31.2%	64.3%	64.2%
% of ventile moves	8.7%	0.1%	1.7%	1.4%
	$\theta = 1.5$	SNR = 1.667	$\theta = 2$	SNR = 2.236
% of rank changes	73.4%	73.2%	77.0%	77.0%
% of ventile moves	4.9%	4.3%	6.5%	6.3%
$n = 1000$	$\theta = 0.5$	SNR = 0.559	$\theta = 1$	SNR = 1.118
% of rank changes	56.1%	54.6%	80.6%	80.6%
% of ventile moves	0.2%	0.0%	0.4%	0.1%
	$\theta = 1.5$	SNR = 1.667	$\theta = 2$	SNR = 2.236
% of rank changes	86.5%	86.5%	88.1%	88.2%
% of ventile moves	0.2%	0.2%	0.4%	0.5%
$n = 5000$	$\theta = 0.5$	SNR = 0.559	$\theta = 1$	SNR = 1.118
% of rank changes	73.7%	73.6%	90.3%	90.3%
% of ventile moves	0.0%	0.0%	0.1%	0.0%
	$\theta = 1.5$	SNR = 1.667	$\theta = 2$	SNR = 2.236
% of rank changes	93.4%	93.4%	94.4%	94.4%
% of ventile moves	0.0%	0.0%	0.0%	0.1%

Table 6 Production frontier—NGE specification. Individual efficiency measures

	$E(\exp\{-u\})$	mode($\exp\{-u\}$)	$E(\exp\{-u\})$	mode($\exp\{-u\}$)
$n = 200$	$\theta = 0.5$	SNR = 0.559	$\theta = 1$	SNR = 1.118
% of rank changes		17.4%		0.3%
% of ventile moves		8.6%		0.3%
abs. diff. in effic. perc. points		6.3		9.3
Pearson's corr with $\exp\{-u\}$	0.35	0.45	0.63	0.61
Kendall's tau with $\exp\{-u\}$	0.28	0.29	0.48	0.48
	$\theta = 1.5$	SNR = 1.667	$\theta = 2$	SNR = 2.236
% of rank changes		1.1%		1.0%
% of ventile moves		1.1%		0.6%
abs. diff. in effic. perc. points		9.2		7.7
Pearson's corr. with $\exp\{-u\}$	0.68	0.66	0.72	0.70
Kendall's tau with $\exp\{-u\}$	0.59	0.60	0.67	0.67
$n = 1000$	$\theta = 0.5$	SNR = 0.559	$\theta = 1$	SNR = 1.118
% of rank changes		1.8%		0.5%
% of ventile moves		0.1%		0.4%
abs. diff. in effic. perc. points		5.9		10.0
Pearson's corr. with $\exp\{-u\}$	0.45	0.46	0.63	0.61
Kendall's tau with $\exp\{-u\}$	0.29	0.29	0.48	0.49
	$\theta = 1.5$	SNR = 1.667	$\theta = 2$	SNR = 2.236
% of rank changes		0.0%		0.2%
% of ventile moves		0.0%		0.1%
abs. diff. in effic. perc. points		9.3		7.8
Pearson's corr. with $\exp\{-u\}$	0.71	0.67	0.74	0.70
Kendall's tau with $\exp\{-u\}$	0.60	0.60	0.68	0.68
$n = 5000$	$\theta = 0.5$	SNR = 0.559	$\theta = 1$	SNR = 1.118
% of rank changes		0.1%		0.1%
% of ventile moves		0.0%		0.1%
abs. diff. in effic. perc. points		5.7		10.2
Pearson's corr. with $\exp\{-u\}$	0.46	0.46	0.64	0.61
Kendall's tau with $\exp\{-u\}$	0.30	0.30	0.48	0.48
	$\theta = 1.5$	SNR = 1.667	$\theta = 2$	SNR = 2.236
% of rank changes		0.0%		0.3%
% of ventile moves		0.0%		0.1%
abs. diff. in effic. perc. points		9.4		7.8
Pearson's corr. with $\exp\{-u\}$	0.71	0.67	0.74	0.70
Kendall's tau with $\exp\{-u\}$	0.60	0.60	0.68	0.68

using them, to those based on OLS residuals. We report the proportion of observations that changed rank, but also the proportion of observations that moved by a *ventile*, namely those that moved at least five percentiles in the ranking distribution.⁸ We consider this a realistic criterion in order to conclude that the rankings differ in a substantial manner from an economic point of view, since, moving a few positions among hundreds of observations does not signal that the two measures rank really differently an observation.

What we observe in Table 5 is that while the large majority, and even almost all of observations change rank under the JLMS measures from the NGE specification compared to the OLS ranking, very few or even none move by a ventile. This does not invalidate the Ondrich and

Ruggiero (2001) result, it just clarifies it appropriately: rankings coming from log-concave densities in SF models don't differ from OLS rankings in an *economically* important way. We also see that the conditional mode behaves much like the conditional expectation in that respect, as anticipated from the related theoretical result obtained earlier.

Finally, we examined how close are the two JLMS measures of inefficiency, but also how they correlate with the true value $\exp\{-u_i\}$. The results are presented in Table 6.

The first two rows of Table 6 in each block are analogous to the first two rows of Table 5: they tell us how many observations in percentage terms changed rank and how many moved at least by a ventile, when ranked by the conditional expectation and when ranked by the conditional mode. Except when the sample size is small and the inefficiency parameter is the lowest simulated, in all other cases

⁸ This is not the same as “changing ventile” which could happen if a firm moves from, say, the 5th percentile to the 6th.

these proportions are both negligible or even zero, indicating that the two JLMS measures result in almost identical rankings, both from a statistical but also from an economic point of view.

The 3rd row gives the average absolute difference in efficiency percentage points, as measured by the two measures. This difference is not negligible: it tells us that on average, the two measures differ in the assessment of individual efficiency by 5 to 10 points (in the 1–100 scale that $100 \times \exp\{-u\}$ lives), which in relative terms may translate to an efficiency score that differs by more than 10–15%, depending on which measure is used.

The 4th and 5th row provide association measures of the JLMS scores with the true variable they predict. The differences are minor to none, with the Conditional Expectation showing a slightly higher linear correlation with $\exp\{-u\}$. We also note that the sample size appears not to matter for how well these measures correlate with the true variable, but the strength of inefficiency (the values of θ_u) does. Overall, the behavior of the two measures is very close, which does not help us to choose among them, in light of their previous difference as regards the actual quantitative estimation of the efficiency score.

But perhaps this is as it should be, since each measure may be appropriate for different situations: mid- and long-term analysis of repeated actions may be more secure using the conditional expectation and the approach of “averaging” that it represents, while short-term one-off decisions may be served better by the “most likely” prediction, and hence by the conditional mode.

Compliance with ethical standards

Conflict of interest The author declares that he has no conflict of interest.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

8 Appendix

8.1 The conditional expectation in the NGE production frontier

In the production SF model we want to compute

$$\begin{aligned} E(\exp\{-u\}|\varepsilon) &= \int_0^\infty e^{-u} f_{u|\varepsilon}(u|\varepsilon) du \\ &= \frac{2}{\sigma_v f_\varepsilon(\varepsilon)} \int_0^\infty e^{-u} \phi((\varepsilon + u)/\sigma_v) h_E(u; \theta_u) du \\ &\quad - \frac{1}{\sigma_v f_\varepsilon(\varepsilon)} \int_0^\infty e^{-u} \phi((\varepsilon + u)/\sigma_v) h_E(u; \theta_u/2) du, \end{aligned}$$

where we have used the additive form of the GE density, eq. (2). Decomposing the integrands,

$$\begin{aligned} E(\exp\{-u\}|\varepsilon) &= \frac{2 \exp\{-\varepsilon^2/2\sigma_v^2\}}{\theta_u \sigma_v f_\varepsilon(\varepsilon) \sqrt{2\pi}} \times (I_1 - I_2), \\ I_1 &= \int_0^\infty \exp\{-u^2/2\sigma_v^2 - (\varepsilon/\sigma_v^2 + 1 + 1/\theta_u)u\} du \\ I_2 &= \int_0^\infty \exp\{-u^2/2\sigma_v^2 - (\varepsilon/\sigma_v^2 + 1 + 2/\theta_u)u\} du. \end{aligned}$$

Combining formulas from Gradshteyn and Ryzhik (2007) and Erdelyi et al. (1954), we have (for $a > 0$),

$$\int_0^\infty \exp\{-az^2 - bz\} dz = \frac{\sqrt{\pi}}{\sqrt{a}} \exp\left\{\frac{b^2}{4a}\right\} \Phi\left(-\frac{b}{\sqrt{2a}}\right).$$

For both integrals, we have the common mapping

$$a = 1/2\sigma_v^2 \Rightarrow \frac{1}{\sqrt{a}} = \sigma_v \sqrt{2}, \quad \frac{1}{\sqrt{2a}} = \sigma_v, \quad \frac{1}{4a} = \sigma_v^2/2.$$

Moreover we have, using indices 1 and 2 for the two integrands,

$$\begin{aligned} b_1 &= \varepsilon/\sigma_v^2 + 1 + 1/\theta_u, \quad b_2 = b_1 + 1/\theta_u \\ \Rightarrow b_2^2 &= b_1^2 + 2b_1/\theta_u + 1/\theta_u^2. \end{aligned}$$

Exploiting these results to make some initial simplifications we can write

$$\begin{aligned} E(\exp\{-u\}|\varepsilon) &= \frac{2 \exp\{-\varepsilon^2/2\sigma_v^2\}}{\theta_u f_\varepsilon(\varepsilon)} \exp\left\{\frac{\sigma_v^2 b_1^2}{2}\right\} \times (A_1 - A_2), \\ A_1 &= \Phi(-\sigma_v b_1), \\ A_2 &= \exp\left\{\frac{\sigma_v^2 b_1}{\theta_u} + \frac{\sigma_v^2}{2\theta_u^2}\right\} \Phi\left(-\sigma_v b_1 - \frac{\sigma_v}{\theta_u}\right) \end{aligned}$$

Now,

$$\begin{aligned} b_1^2 &= \frac{\varepsilon^2}{\sigma_v^4} + 2 \frac{1 + \theta_u}{\theta_u} \frac{\varepsilon}{\sigma_v^2} + \left(\frac{1 + \theta_u}{\theta_u}\right)^2 \\ \Rightarrow \frac{\sigma_v^2 b_1^2}{2} &= \frac{\varepsilon^2}{2\sigma_v^2} + \frac{1 + \theta_u}{\theta_u} \varepsilon + \frac{\sigma_v^2}{2} \left(\frac{1 + \theta_u}{\theta_u}\right)^2. \end{aligned}$$

Also,

$$\sigma_v b_1 = \frac{\varepsilon}{\sigma_v} + \sigma_v \left(\frac{1 + \theta_u}{\theta_u}\right), \quad \frac{\sigma_v^2 b_1}{\theta_u} = \frac{\varepsilon}{\theta_u} + \frac{\sigma_v^2}{\theta_u} + \frac{\sigma_v^2}{\theta_u^2}.$$

Combining all, we arrive at

$$\begin{aligned} E(\exp\{-u\}|\varepsilon) &= \frac{2}{\theta_u f_\varepsilon(\varepsilon)} \exp\left\{\frac{1 + \theta_u}{\theta_u} \varepsilon + \frac{\sigma_v^2}{2} \left(\frac{1 + \theta_u}{\theta_u}\right)^2\right\} \\ &\quad \times \left[\Phi\left(-\frac{\varepsilon}{\sigma_v} - \sigma_v \left(\frac{1 + \theta_u}{\theta_u}\right)\right) \right. \\ &\quad \left. - \exp\left\{\frac{\varepsilon}{\theta_u} + \frac{\sigma_v^2}{\theta_u^2} \left(\theta_u + \frac{3}{2}\right)\right\} \Phi\left(-\frac{\varepsilon}{\sigma_v} - \sigma_v \left(\frac{2 + \theta_u}{\theta_u}\right)\right) \right]. \end{aligned}$$

8.2 The conditional mode in the NGE production frontier

We want to maximize the conditional density in eq. (9) with respect to q . Ignoring the terms that do not include q , the derivative we want to set equal to zero is

$$\begin{aligned} & \frac{\partial}{\partial q} \left[\phi \left(\frac{\varepsilon_i - \ln q}{\sigma_v} \right) \cdot \left(q^{-1+1/\theta_u} - q^{-1+2/\theta_u} \right) \right] \\ &= \phi' \left(\frac{\varepsilon_i - \ln q}{\sigma_v} \right) (-1/q\sigma_v) \cdot \left(q^{-1+1/\theta_u} - q^{-1+2/\theta_u} \right) \\ & \quad + \frac{1}{q} \phi \left(\frac{\varepsilon_i - \ln q}{\sigma_v} \right) \times \left[\left(-1 + \frac{1}{\theta_u} \right) q^{-1+1/\theta_u} - \left(-1 + \frac{2}{\theta_u} \right) q^{-1+2/\theta_u} \right] \end{aligned}$$

Using the property of the standard Normal density $\phi'(z) = -z\phi(z)$ we obtain

$$\begin{aligned} \dots &= \frac{1}{q} \phi \left(\frac{\varepsilon_i - \ln q}{\sigma_v} \right) \times \left\{ \frac{\varepsilon_i - \ln q}{\sigma_v} \cdot \left(q^{-1+1/\theta_u} - q^{-1+2/\theta_u} \right) \right. \\ & \quad \left. + \left(-1 + \frac{1}{\theta_u} \right) q^{-1+1/\theta_u} - \left(-1 + \frac{2}{\theta_u} \right) q^{-1+2/\theta_u} \right\}. \end{aligned}$$

Setting this expression equal to zero is equivalent to setting the expression in curly brackets equal to zero since the term outside is always positive. Manipulating further,

$$\begin{aligned} \dots \Rightarrow & \left(q^{-1+1/\theta_u} - q^{-1+2/\theta_u} \right) \cdot \left[\frac{\varepsilon_i - \ln q}{\sigma_v} - 1 + \frac{1}{\theta_u} \right] \\ & - \frac{1}{\theta_u} q^{-1+2/\theta_u} = 0. \end{aligned}$$

Taking q^{-1+2/θ_u} out as common factor we arrive at the expression used in the main text.

References

- Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. *J Econom* 6:21–37
- Azzalini A (1985) A class of distributions which includes the normal ones. *Scand J Stat* 12:171–178
- Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *J Econom* 38:387–399
- Cai J, Feng Q, Horrace WC, Wu GL (2020) Wrong skewness and finite sample correction in parametric stochastic frontier models. *Empir Econ* Forthcom
- Egozcue M (2015) Some covariance inequalities for non-monotonic functions with applications to mean-variance indifference curves and bank hedging. *Cogent Math* 2:991082
- Erdelyi A, Magnus W, Oberhettinger F, Tricomi FG (eds) (1954) *Tables of Integral Transforms: Vol.1*. New York, McGraw-Hill
- Gradshteyn IS, Ryzhik IM (2007) *Tables of integrals, series and functions* (7th ed). Burlington MA USA, Academic Press
- Greene W (1980) On the estimation of a flexible frontier production model. *J Econom* 13:101–115
- Greene W (1990) A gamma-distributed stochastic frontier model. *J Econom* 46:141–163
- Grushka E (1972) Characterization of exponentially modified gaussian peaks in chromatography. *Anal Chem* 44:1733–1738
- Gupta RD, Kundu D (1999) Generalized Exponential distributions. *Austral N. Z. J Stat* 41:173–188
- Hafner CM, Manner H, Simar L (2018) The wrong skewness problem in stochastic frontier models: a new approach. *Econom Rev* 37:380–400
- Jondrow J, Lovell CK, Materov I, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. *J Econom* 19:233–238
- Kumbhakar SC, Lovell C (2000) *Stochastic Frontier Analysis*. Cambridge, Cambridge University Press
- Materov I (1981) On full identification of the stochastic production frontier model. *Ekonomika i matematische metode* 17:784–788
- Meeusen W, van-den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. *Int Econ Rev* 18:435–444
- Ondrich J, Ruggiero J (2001) Efficiency measurement in the stochastic frontier model. *Eur J Oper Res* 129:434–442
- Papadopoulos A (2018) *The two-tier stochastic frontier framework: Theory and applications, models and tools*. PhD Thesis. Athens Greece, Athens University of Economics and Business
- Papadopoulos A (2020a) Accounting for endogeneity in regression models using Copulas: a step-by-step guide for empirical studies. Manuscript. Athens University of Economics and Business
- Papadopoulos A (2020b) Measuring the effect of management on production: a two-tier stochastic frontier approach. *Empir Econ*. <https://doi.org/10.1007/s00181-020-01946-9>
- Papadopoulos A, Parmeter CF (2020) Quasi-maximum likelihood estimation for the stochastic frontier model. Manuscript. Athens University of Economics and Business
- Papadopoulos A, Parmeter CF (2021) Type II failure and specification testing for the stochastic frontier model. *Eur J Oper Res*. <https://doi.org/10.1016/j.ejor.2020.12.065>
- Papadopoulos A (2020c) The two-tier stochastic frontier framework (2TSF): measuring frontiers wherever they may exist. In: Parmeter CF, Sickles R (eds) *Advances in Efficiency and Productivity Analysis*, Cham Switzerland, Springer
- Parmeter CF, Kumbhakar SC (2014) Efficiency analysis: a primer on recent advances. *Found Trends Econom* 7:191–385
- Parmeter CF (2018) Estimation of the two-tiered stochastic frontier model with the scaling property. *J Product Anal* 49:37–47
- Polachek SW, Yoon BJ (1987) A two-tiered earnings frontier estimation of employer and employee information in the labor market. *Rev Econ Stat* 69:296–302
- Ritter C, Simar L (1997) Pitfalls of normal-gamma stochastic frontier models. *J Product Anal* 8:167–182
- Sickles R, Zelenyuk V (2019) *Measurement of productivity and efficiency*. Cambridge, Cambridge University Press
- Simar L, Wilson PW (2009) Inferences from cross-sectional, stochastic frontier models. *Econom Rev* 29:62–98
- Stead AD, Wheat P, Greene WH (2019) Distributional forms in stochastic frontier analysis. In: Ten Raa T, Greene WH (eds) *The Palgrave Handbook of Economic Performance Analysis*, Cham, Switzerland, Palgrave Macmillan
- Stevenson R (1980) Likelihood functions for generalized stochastic frontier estimation. *J Econom* 13:57–66
- Tran KC, Tsionas EG (2015) Endogeneity in stochastic frontier models: copula approach without external instruments. *Econ Lett* 133:85–88
- Wang JH, Schmidt P (2002) One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *J Product Anal* 18:129–144
- Williams D (1991) *Probability with martingales*. Cambridge, Cambridge University Press