



Maximum likelihood estimation of normal-gamma and normal-Nakagami stochastic frontier models

Alexander D. Stead ¹

Accepted: 6 November 2024
© The Author(s) 2024

Abstract

The gamma and Nakagami distributions have an advantage over other proposed flexible inefficiency distributions in that they can accommodate not only non-zero modes, but also cases in which many firms lie arbitrarily close to the frontier. We propose a normal-Nakagami stochastic frontier model, which provides a generalisation of the normal-half normal that is more flexible than the familiar normal-truncated normal. The normal-gamma model has already attracted much attention, but estimation and efficiency prediction have relied on approximation methods. We derive exact expressions for likelihoods and efficiency predictors, and demonstrate direct maximum likelihood estimation of both models. Across three empirical applications, we show that the models avoid a convergence issue that affects the normal-truncated normal model, and can accommodate a concentration of observations near the frontier similar to zero-inefficiency stochastic frontier models. We provide Python implementations via the `FronPy` package.

Keywords Stochastic frontier analysis · Gamma distribution · Nakagami distribution · Maximum likelihood estimation

JEL codes C21 · C46 · D24

1 Introduction

Since the mid- to late 2000s, there has been a marked slowdown in productivity growth in advanced economies (Fernald et al. 2023; Gordon and Sayed 2019). How are increases in productivity in a given industry to be realised? Do all firms have comparable scope for improvement? Or is there a ‘long tail’ of inefficient firms that need to catch up to the frontier or exit the market? These questions have been the subject of several studies (Bersch et al. 2019; Bloom and Van Reenen 2007; Oliveira-Cunha et al. 2021), and are of great relevance to policymakers, for whom increasing productivity is one of the most pressing current challenges.

The original normal-half normal (N-HN) and normal-exponential (N-EXP) stochastic frontier (SF) specifications introduced by Aigner et al. (1977) and Meeusen and van Den Broeck (1977) – still the workhorses of the applied SF literature – are ill-equipped to answer such questions. While

they may provide useful estimates of the average firm-level efficiency, and give some insight into the ranking of firms, they impose inefficiency distributions that have fixed shapes and zero modes. This study explores estimation of – and efficiency prediction from – two specifications that allow for far more flexibility in the shape of the inefficiency distribution: a normal-gamma (N-G) model and a normal-Nakagami (N-NAK) model, which generalise the N-EXP and N-HN specifications, respectively.

The N-G model has attracted considerable attention over the years, and several approaches to approximating the integrals in the N-G model have been proposed. Greene (1990) uses Laguerre and Newton-Coates quadrature methods. Ritter and Simar (1997) use the trapezoidal rule. Greene (2003) simulates the integral. Tsionas (2012) uses a fast Fourier transformation of the characteristic function. An exact solution is available, however – Beckers and Hammond (1987) derive expressions for the N-G likelihood in terms of the confluent hypergeometric function ${}_1F_1$ (Kummer 1837), but to our knowledge this has only ever been used by Hammond (1992).

We extend the approach of Beckers and Hammond (1987) to derive exact expressions for the N-G and N-NAK

✉ Alexander D. Stead
a.d.stead@leeds.ac.uk

¹ University of Leeds, Leeds LS2 9LJ, UK

likelihoods and efficiency predictors. The use of the Nakagami distribution has not, to our knowledge, been explored in the SF literature, and our results relating to the N-NAK model are new. Our expressions for the efficiency predictors for the N-G model are also new, previous literature having used various approximations.

Further motivation is needed. After all, beginning with Stevenson (1980) many alternative inefficiency distributions have been proposed – for recent reviews see Kumbhakar et al. (2020) and Stead et al. (2019). We now have plenty of options to choose from which allow for non-zero modes, or flexible shapes, or both. And estimation of the N-G model has already attracted much discussion in the literature. So what is the rationale for revisiting the N-G model, and why is the N-NAK model worthy of attention? There are both economic and statistical motivations: we show that the N-G and N-NAK models offer greater flexibility in allowing an arbitrary concentration of firms near the frontier, whereas other flexible distributions contain the N-HN and N-EXP models as limiting cases in this regard; we also show that the use of our exact expressions leads to significant improvements in accuracy and optimisation times.

These motivations are discussed in detail in Section 2. The remainder of the paper is structured as follows. In Section 3 we derive the likelihood function and efficiency predictors for N-G and N-NAK SF models and discuss estimation. Section 4 demonstrates implementation of the models using a purpose-built Python package, `FRONPY` (Stead 2023), with application to real and simulated data. Section 5 summarises and concludes.

2 Motivation

In this section, we discuss the motivation for further discussion of maximum likelihood estimation of the N-G SF model, and for proposing the new N-NAK specification, in which U_i follows a Nakagami distribution (Nakagami 1960). To be specific, we are concerned with standard cross-sectional SF models of the form outlined by Assumption 1.

Assumption 1

$$\begin{aligned} y_i &= \mathbf{x}'_i \boldsymbol{\beta} + E_i, & E_i &= V_i - sU_i, & V_i &\sim N(0, \sigma_V), \\ \mathbf{x}_i &\perp\!\!\!\perp V_j \forall i, j, & \mathbf{x}_i &\perp\!\!\!\perp U_j \forall i, j, & V_i &\perp\!\!\!\perp U_j \forall i, j. \end{aligned} \quad (1)$$

for all observations $i = 1, \dots, I$, where y_i is the dependent variable, \mathbf{x}_i is a coefficient vector, $\boldsymbol{\beta}$ is a vector of parameters, $s = 1$ for a production frontier or $s = -1$ for a cost frontier, and V_i and $U_i \geq 0$ are random variables representing noise and inefficiency, respectively.

The N-G and N-NAK models are then obtained by assuming that U_i follows gamma and Nakagami distributions, respectively. Assumptions 2 and 3 give the corresponding densities of U_i .

Assumption 2

$$\begin{aligned} U_i &\sim \text{Gamma}(\mu, \sigma_U) \implies \\ f_U(U_i = u_i) &= \begin{cases} \frac{1}{\Gamma(\mu)\sigma_U^\mu} u_i^{\mu-1} \exp\left(-\frac{u_i}{\sigma_U}\right), & u_i \geq 0. \\ 0, & u_i < 0. \end{cases} \end{aligned} \quad (2)$$

where f_U denotes the density function of U_i .

Assumption 3

$$\begin{aligned} U_i &\sim \text{Nakagami}(\mu, \sigma_U) \implies \\ f_U(U_i = u_i) &= \begin{cases} \frac{2\mu^\mu}{\Gamma(\mu)\sigma_U^{2\mu}} u_i^{2\mu-1} \exp\left(-\mu\left(\frac{u_i}{\sigma_U}\right)^2\right), & u_i \geq 0. \\ 0, & u_i < 0. \end{cases} \end{aligned} \quad (3)$$

where f_U denotes the density function of U_i .

Note that the latter differs from the usual parameterisation introduced by Nakagami (1960), which uses $\Omega = \sigma_U^2$. The gamma and Nakagami distributions are closely related, both belonging to the generalised gamma family (Stacey 1962). Note that $\text{Gamma}(1, \sigma_U) = \text{Exponential}(1/\sigma_U)$, and $\text{Nakagami}(0.5, \sigma_U) = N^+(0, \sigma_U^2)$ so that the N-G model nests the N-EXP model and the N-NAK nests the N-HN when $\mu = 1$ and $\mu = 0.5$, respectively.¹ The economic motivation for these specifications is discussed in Section 2.1, while the statistical motivations are discussed in Section 2.2.

2.1 Economic motivation

In terms of economic motivation, most of the literature proposing alternative inefficiency distributions has focused on allowing for the possibility of a non-zero mode. The opposite problem of allowing for a *greater* concentration of firm-level efficiencies near zero has received little attention. Most alternative inefficiency distributions only allow for as much concentration near zero as the exponential distribution, or less. For instance, the one-parameter Rayleigh and generalised exponential distributions, proposed by Hajar-gasht (2021) and Papadopoulos (2021), respectively, allow *only* for non-zero modes – or no mode, as with the uniform distribution proposed by Li (1996). In common with the

¹ We used the following notation: If $W_i \sim N(\mu, \sigma_V^2)$, then $W_i | W_i > 0 \sim N^+(\mu, \sigma_V^2)$ is a truncated normal random variable. If $W_i \sim N(0, \sigma_V^2)$, it then follows a half normal distribution.

half normal and exponential cases, they have fixed shapes, and the concentration of firms on or near the frontier increases only as $\text{Var}(U_i) \rightarrow 0$.

Flexible distributions with two or more parameters are not necessarily better in this regard. Stevenson (1980) proposed the use of a truncated normal distribution, generalising the half normal and allowing not only for non-zero modes, but also for zero mode distributions with a variety of different shapes. However, Meesters (2014) finds that the truncated normal distribution contains the exponential as a limiting case, so that in terms of the concentration of firms near the frontier, we can do no better under the truncated normal specification than under the exponential. The truncated Laplace distribution used by Horrace and Parmeter (2018) can also accommodate no more concentration near the frontier than the exponential, by virtue of the memorylessness property of the latter.

Table 1 Supra-percentile unconditional mean efficiency estimates when $\mathbb{E}[\exp(-U_i)] = 0.9$ for various values of μ , when $U \sim \text{Gamma}(\mu, \sigma_U)$

p	$\mu = 0.1$	$\mu = 0.2$	$\mu = 0.5$	$\mu = 1.0$	$\mu = 2.0$	$\mu = 5.0$
0.10	1.0000	1.0000	0.9994	0.9943	0.9819	0.9607
0.20	1.0000	1.0000	0.9975	0.9882	0.9729	0.9517
0.30	1.0000	0.9998	0.9943	0.9816	0.9650	0.9447
0.40	1.0000	0.9992	0.9897	0.9745	0.9576	0.9387
0.50	0.9999	0.9976	0.9835	0.9667	0.9502	0.9330
0.60	0.9994	0.9941	0.9754	0.9581	0.9427	0.9275
0.70	0.9971	0.9869	0.9649	0.9483	0.9347	0.9219
0.80	0.9889	0.9737	0.9512	0.9368	0.9259	0.9160
0.90	0.9642	0.9495	0.9324	0.9226	0.9155	0.9094
0.95	0.9389	0.9301	0.9195	0.9134	0.9091	0.9054
0.99	0.9089	0.9077	0.9053	0.9036	0.9024	0.9014
1.00	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000

To see why this is a problem, consider that in terms of economic theory, we might ordinarily expect a large proportion of firms – perhaps the majority, in a context of effective competition or regulation – to be on, or at least very close to, the frontier. However, the exponential and half normal distributions have most of their mass in their tails and shoulders, implying that the majority of firms must have some appreciable level of inefficiency, with very few being close to the frontier. This limitation motivated the zero inefficiency stochastic frontier (ZISF) model (Kumbhakar et al. 2013; Rho and Schmidt 2015), which takes a latent class approach in which some proportion of firms is assumed to be fully efficient.

In contrast to other distributions we have discussed, the gamma – originally proposed by Stevenson (1980) as a means of generalising the exponential specification – can allow for an arbitrary concentration of firms near the frontier by allowing its shape parameter μ to approach zero. The Nakagami distribution behaves in a similar fashion. As inefficiency distributions, they both allow for behaviour very similar to that of a ZISF model. While they do not imply that a certain number of firms lie exactly on the frontier as do ZISF models, the gamma and Nakagami densities can become so concentrated near zero that a non-negligible number of firms end up, for all practical purposes, on the frontier.

The gamma and Nakagami models therefore have the virtue of being able to behave like ZISF models without having to resort to latent class approaches, with all the challenges around estimation and hypothesis testing that that entails. They also have the advantage of greater flexibility – being also able to accommodate non-zero modes.

To demonstrate the value of having such flexibility with respect to shape, imagine that we know mean firm-level efficiency is 0.9, i.e. $\mathbb{E}[\exp(-U_i)] = 0.9$. Then consider $\mathbb{E}[\exp(-U_i) | U_i < F_U^{-1}(p)]$ – in simple terms, the mean efficiency among the top $100 \times p\%$ most efficient firms.

Table 2 Supra-percentile unconditional mean efficiency estimates when $\mathbb{E}[\exp(-U_i)] = 0.9$ for various values of μ , when $U \sim \text{Nakagami}(\mu, \sigma_U)$

p	$\mu = 0.05$	$\mu = 0.10$	$\mu = 0.20$	$\mu = 0.50$	$\mu = 1.00$	$\mu = 2.00$
0.10	1.0000	1.0000	0.9997	0.9915	0.9745	0.9557
0.20	1.0000	1.0000	0.9983	0.9830	0.9636	0.9460
0.30	1.0000	0.9997	0.9954	0.9745	0.9549	0.9390
0.40	1.0000	0.9989	0.9906	0.9659	0.9471	0.9330
0.50	0.9999	0.9968	0.9837	0.9571	0.9399	0.9276
0.60	0.9992	0.9920	0.9744	0.9479	0.9329	0.9226
0.70	0.9964	0.9829	0.9625	0.9383	0.9259	0.9176
0.80	0.9866	0.9673	0.9474	0.9278	0.9185	0.9126
0.90	0.9595	0.9422	0.9282	0.9160	0.9105	0.9071
0.95	0.9350	0.9245	0.9161	0.9090	0.9059	0.9040
0.99	0.9082	0.9060	0.9040	0.9023	0.9015	0.9010
1.00	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000

This is the supra-percentile unconditional mean efficiency for percentile p of the U_i distribution.² If we divide this by $\mathbb{E}[\exp(-U_i)]$, it gives us the potential productivity gains to be realised if the bottom $(1-p) \times 100\%$ of firms catch up to the rest or exit the market. Formulae for $\mathbb{E}[\exp(-U_i)]$ and $\mathbb{E}[\exp(-U_i)|U_i < F_U^{-1}(p)]$ for various distributions of U_i are given in the Supplementary Appendices.

Tables 1 and 2 show the variation in $\mathbb{E}[\exp(-U_i)|U_i < F_U^{-1}(p)]$ when $\mathbb{E}[\exp(-U_i)] = 0.9$ but μ is allowed to vary – σ_U is adjusted accordingly in each case. Table 1 relates to the gamma distribution, in which case when $\mu = 1$ we have the exponential, while Table 2 relates to the Nakagami distribution, in which case when $\mu = 0.5$ we have the half normal. The first thing to note is the similarity of the columns relating to these special cases – for each value of p , the picture that emerges does not differ radically if one switches between the exponential and half normal cases.

Note that the relationship between supra-percentile unconditional mean efficiency and p depends fundamentally on the curvature of $f_U(U_i = u_i)$. In general, it can be shown that

$$\frac{\partial \mathbb{E}[\exp(-U_i)|U_i < F_U^{-1}(p)]}{\partial \ln p} = \exp(-F_U^{-1}(p)) - \mathbb{E}[\exp(-U_i)|U_i < F_U^{-1}(p)], \quad (4)$$

i.e. the derivative of $\mathbb{E}[\exp(-U_i)|U_i < F_U^{-1}(p)]$ with respect to $\ln p$ is simply the difference between percentile p in efficiency terms and the supra-percentile unconditional mean efficiency itself.³ As $F_U^{-1}(p)$ moves further to the right of the peak, this derivative therefore increases more slowly where the density is steeper, and more rapidly where the density is flatter. In Tables 1 and 2, in cases where the mode is zero we therefore see the values changing more slowly the lower the value of μ .

The columns to the right show how the picture changes in each case as we move the mode of U_i away from zero – the difference between the more efficient firms and the frontier increases, while the differences between the lower deciles diminishes. Note that the truncated normal model will offer similar flexibility in this direction due to its ability to accommodate non-zero modes. On the other hand we see from the columns to the left that the gamma and Nakagami distributions, for small μ , can allow for cases in which the first several deciles of firms have practically no inefficiency at all – similar to a ZISF model – whereas the truncated normal does not allow for this situation; it follows from the Meesters (2014) result that when $\mu < 0$, the truncated normal will be intermediate between the half normal and exponential cases.

We therefore motivate the Nakagami distribution for U_i in terms of its flexibility. In particular we argue that, as a generalisation of the familiar half normal case, it is preferable to the established truncated normal distribution on account of its superior ability to allow for a large concentration of firms near the frontier. The gamma distribution has a similar ability – previously overlooked. We argue that this feature allows for more theoretically plausible inefficiency distributions than other specifications. It is clear that the various columns of Tables 1 and 2 could lead to substantially different policy conclusions regarding the ways in which to increase productivity, e.g. whether the need is for broad-based strategies to improve performance in an industry, or more tailored measures encouraging less efficient firms to either catch up to the frontier or exit the market.

2.2 Statistical motivation

In this section we discuss motivations besides the flexibility of the gamma and Nakagami distributions. We first discuss an estimation problem affecting the N-TN model, which provides additional motivation for the N-NAK model in particular. We then discuss improvements in accuracy and speed of estimation and efficiency prediction, which are key motivators for revisiting implementation of the N-G model. Finally, we discuss potential pitfalls.

2.2.1 Estimation issues with the normal-truncated normal model

Although the likelihood function of the N-TN model has a convenient expression, and the model is well-established in the SF literature as a generalisation of the N-HN specification, it appears to be quite common to encounter problems with convergence when attempting to estimate the model by maximum likelihood. Rather than converging, the estimation routine keeps iterating as σ_U and μ grow increasingly large in magnitude.

This phenomenon may be explained with reference to the Meesters (2014) result, discussed in Section 2.1, that the truncated normal distribution contains the exponential as a limiting case. More specifically, Meesters (2014) notes that when $\sigma_U \rightarrow \infty$, $\mu \rightarrow -\infty$ such that $-\mu/\sigma_U^2 > 0$ remains constant, the distribution of U_i converges to Exponential($-\mu/\sigma_U$), so that the N-TN model contains the N-EXP on a boundary of its parameter space. If the true distribution of U_i is exponential – or the exponential is the best possible approximation – then the estimator may well fail to converge as described. In Section 4 we demonstrate that this occurs in all three of our empirical applications,⁴ and that the parameter estimates and efficiency predictions

² For example, if $p = 0.1$, we are considering the 10% most efficient firms, if $p = 0.2$ the 20% most efficient, and so on.

³ Derivation is shown in the Supplementary Appendices.

⁴ This was by chance rather than design.

from the N-TN models become practically identical to those from the N-EXP models.

If this convergence issue with the N-TN model occurs frequently in applications to real data, this provides a powerful motivation for considering alternative flexible specifications, such as the N-G and N-NAK. It strengthens the case for the N-NAK model, as an alternative generalisation of the N-HN, in particular.

2.2.2 Accuracy and speed of estimation

In order to estimate the N-G and N-NAK models via maximum likelihood, we need to derive the marginal densities of E_i in each case, which we can use to derive the log-likelihood functions. In contrast to the N-TN, N-HN, and N-EXP cases, the likelihood and efficiency predictors of the N-G model lack convenient expressions in terms of the kinds of functions routinely available in standard statistical and econometric software packages. Implementation of the N-G model has received considerable attention over the years, given the difficulty of the integral that appears in the marginal density of E_i which, under Assumptions 1 and 2, is given by

$$f_E(E_i = \varepsilon_i) = \frac{1}{\sqrt{2\pi}\Gamma(\mu)\sigma_V\sigma_U^\mu} \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma_V}\right)^2\right) \times \int_0^\infty u_i^{\mu-1} \exp\left(-\frac{u_i^2}{2\sigma_V^2} - \frac{(s\sigma_U\varepsilon_i + \sigma_V^2)u_i}{\sigma_V^2\sigma_U}\right) du_i. \quad (5)$$

Stevenson (1980) derived relatively simple expressions for Eq. (5) for a few cases of integer μ . However the lack of a convenient general solution has, until now, been a barrier to direct maximum likelihood estimation of the model. Various approaches to approximation of the likelihood have therefore been explored: Greene (1990) used Gauss-Laguerre quadrature, Ritter and Simar (1997) used the trapezoid rule, and Greene (2003) used simulation.

In terms of speed and accuracy, de Andrade and Souza (2018) compare various methods of approximating the N-G density and show that the most accurate is inverse fast Fourier transformation (FFT) of the characteristic function, as proposed by Tsionas (2012). This takes advantage of two facts: first, that the characteristic function of the sum of two independent random variables is simply the product of their individual characteristic functions; second, that the density function is the inverse Fourier transformation of the characteristic function. The approach taken by Tsionas (2012) is therefore to apply inverse FFT – a fast, discrete approximation of the inverse Fourier transform – to the N-G characteristic function, in order to approximate the log-likelihood.

Results from Tsionas (2012) and de Andrade and Souza (2018) suggest that this inverse FFT approximation is highly accurate and, for the N-EXP model, yields densities and

maximum likelihood estimates that are practically identical to those obtained using exact expressions within a matter of seconds for moderate sample sizes. Improvements in accuracy and speed from using exact expressions, such as we derive for the N-G and N-NAK likelihoods, may therefore be marginal.

Based on application of the inverse FFT approximation to both real and simulated data it appears, however, that it sometimes leads to issues with respect to accuracy or numerical stability, or both. For the sake of brevity, we do not report these results in full, though a discussion of accuracy and comparison of optimisation times for the N-G models in our empirical applications is included in Section 4.2. The accuracy issue can be remedied by increasing sequence length, at the cost of longer optimisation times – though without also maximising the exact log-likelihood for comparison, it may be hard to diagnose. The issue of instability is more concerning.

A comprehensive investigation of these issues, and possible remedies, is beyond the scope of this paper. We conjecture that they are a result of the periodicity assumption underlying FFT algorithms, i.e. the assumption that the input is one period of a continuous, periodically repeating signal. This should be no issue if we are approximating a symmetric density – the periodic extension of a (symmetrically) truncated symmetric density yields a continuous, repeating signal. But in SF modelling, the composed error density is, by construction, asymmetric; the same process of symmetric truncation and periodic extension will result in jump discontinuities. The presence of such jump discontinuities causes a phenomenon known as *spectral leakage*, which introduces unwanted artefacts into the output – for a discussion see e.g. Brigham (1988).

Figure 1, which compares exact and approximated log-densities for a N-EXP model in which $\sigma_V = \sigma_U = 0.1$, appears to support this conjecture. The density was approximated for the interval $\varepsilon_i \in [-0.8, 0.8]$ and, comparing the approximated and exact log-densities over the same interval, we see that the approximation is practically exact around the mode and right tail of the distribution, but differs significantly at the left tail. In this instance, moving right to left, the inverse FFT approximation diverges from the exact log-density at around $\varepsilon_i = -0.4$ and increases such that both the left and right tails intersect the vertical axis at a value of around -5 . As a result of the FFT algorithm forcing the output to conform to its assumption of periodicity, we therefore have a significant degree of inaccuracy in the left tail of the distribution. The fact that the density is increasing as we move further into the left tail is particularly concerning, since it suggests that outlying observations could cause significant inaccuracy and numerical instability.⁵

⁵ This issue is hard to detect when visually examining the densities as in Tsionas (2012) and de Andrade and Souza (2018) due to the fact it occurs in the tails of the distribution, but is more apparent when comparing log-densities or log-likelihoods.

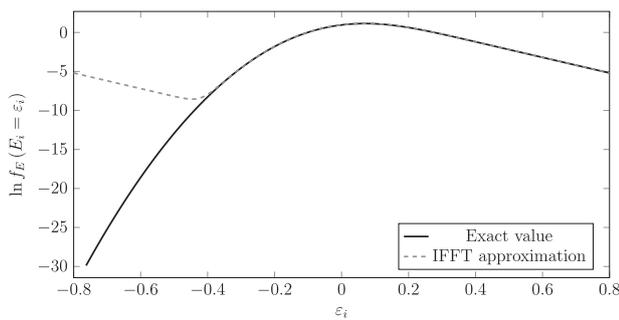


Fig. 1 Comparison of exact and approximated log-densities for $E_i = V_i - U_i$, $V_i \sim N(0, \sigma_V^2)$, $U_i \sim \text{Exponential}(1/\sigma_U)$, $V_i \perp U_i$, $\sigma_V = \sigma_U = 0.1$

As noted by Beckers and Hammond (1987), a solution to Eq. (5) is given by expression 3.462.1 of Gradshteyn and Ryzhik (2014). To our knowledge, however, the only study to have implemented the N-G model using this result is Hammond (1992) – Yuengert (1993) estimates an N-G model, but with the restriction $\alpha = 2$, taking advantage of the results derived by Stevenson (1980) for this special case. By deriving similar results for the N-G and N-NAK likelihoods and efficiency predictors, we avoid potential issues of inaccuracy and instability in estimation, and additionally avoid the need to use approximations for efficiency prediction.

Of course, the evaluation of any special function – even the exponential function, for example – involves numerical approximation methods, e.g. series expansions, asymptotic expansions, recurrence relations, and continued fractions. But these approximations can be optimised for precision and stability such that they can be computed to arbitrary precision, whereas the issue of spectral leakage affecting the inverse FFT approximation is a more fundamental problem arising from limitations of FFT algorithms in handling non-periodic inputs. The use of special function representations of likelihoods will therefore generally offer greater precision than inverse FFT approximations.

2.2.3 Other considerations

A further motivation for the use of the gamma and Nakagami distributions is they are more flexible with respect to skewness than other commonly-used distributions for U_i . For example, as discussed by Papadopoulos and Parmeter (2021),

$$\begin{aligned} U_i \sim N^+(0, \sigma_V^2) &\implies \text{Skewness}(U_i) = \frac{\sqrt{2}(4-\pi)}{(\pi-2)^{3/2}}, \\ U_i \sim N^+(\mu, \sigma_U^2) &\implies \text{Skewness}(U_i) \in (0, 2), \\ U_i \sim \text{Exponential}(0, \sigma_U) &\implies \text{Skewness}(U_i) = 2, \end{aligned}$$

so that the skewness of commonly-used distributions for U_i are either fixed or bound within a relatively narrow range.

By contrast, the gamma and Nakagami distributions are very flexible with respect to their skewnesses, such that

$$\begin{aligned} U_i \sim \text{Gamma}(\mu, \sigma_U) &\implies \text{Skewness}(U_i) = \frac{2}{\sqrt{\mu}} \in [0, \infty), \\ U_i \sim \text{Nakagami}(\mu, \sigma_U) &\implies \text{Skewness}(U_i) = \frac{\Gamma(\mu+3/2)}{\mu^{3/2}\Gamma(\mu)} \in (1, \infty). \end{aligned}$$

We should also address some potential issues. Ritter and Simar (1997) note that the N-G model suffers from potential identification problems owing to the way the composed error becomes Gaussian in three different cases. These cases and their analogues in the N-NAK model are outlined in Propositions 1–3 in the Supplementary Appendices. These issues are not unique to the N-G and N-NAK models, but apply to SF models with flexible distributions for U_i more generally.

It is worth noting that the N-G and N-NAK models appear to suffer from the ‘wrong skewness’ problem – see Papadopoulos and Parmeter (2024) for a recent review – as with other common specifications. They offer additional flexibility with respect to the magnitude of population skewness, but not its sign. In addition, both satisfy Assumptions 1–5 of Horrace and Wright (2020), so we can conclude that the ordinary least squares (OLS) estimator is a stationary point in the N-G and N-NAK likelihoods. Establishing the stability of these stationary points and connecting this to the skewness of the OLS residuals would be challenging, given the complexity of their scores and Hessians. However, our experience using simulated data suggests that maximum likelihood estimation breaks down when the OLS residuals are skewed in the wrong direction – this makes intuitive sense given the identification issues we have just referred to – or when the excess kurtosis of the OLS residuals is negative, pointing to an analogous ‘wrong kurtosis’ problem.

We may also note that the gamma and Nakagami densities, when $\mu < 1$ and $\mu < 0.5$, respectively, possess vertical asymptotes – and thus non-differentiable points – at zero. In Section 3, however, we will see that this is not true of the densities of E_i , so this does not cause any problems with respect to maximum likelihood estimation. Of greater relevance is the fact that the densities of U_i become log-convex everywhere in these cases, with implications for ML estimation and efficiency prediction, discussed further in Section 3.2 and the Supplementary Appendices, respectively.

3 Implementation

In this section, we discuss implementation of the N-G and N-NAK models. In Section 3.1 we derive of key expressions such as log-likelihoods and efficiency predictors. In Section 3.2 we discuss our approach to estimation.

3.1 Derivation

Since one of the main objectives of SF analysis is the prediction of firm-level efficiencies, we must also derive formulae for efficiency predictors in each case. Following Jondrow et al. (1982) and Battese and Coelli (1988), the most commonly used predictors are the conditional means $\mathbb{E}[U_i|E_i = \varepsilon_i]$ and $\mathbb{E}[\exp(-U_i)|E_i = \varepsilon_i]$, and the conditional mode $\text{Mode}(U_i|E_i = \varepsilon_i)$.

Theorems 1 and 2 give the marginal densities of E_i in the N-G and N-NAK cases, respectively, while their corollaries give the relevant conditional densities and efficiency predictors. Eq. (6) was derived previously by Beckers and Hammond (1987), whereas Eqs. (7–15) appear to be new results. In these expressions, D_ν denotes Whittaker’s parabolic cylinder function (Whittaker 1902). The proofs are straightforward, and are shown in the Supplementary Appendices.

These theorems also give us alternative expressions for the key results for the N-EXP and N-HN models, when $\mu = 1$ and when $\mu = 0.5$, respectively. The expressions derived by Aigner et al. (1977) can be recovered given the following special cases of the parabolic cylinder function:

$$D_{-1}(z) = \sqrt{\frac{\pi}{2}} \exp\left(\frac{z^2}{4}\right) \left(1 - \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right),$$

$$D_{-2}(z) = \exp\left(-\frac{z^2}{4}\right) - \sqrt{\frac{\pi}{2}} z \exp\left(\frac{z^2}{4}\right) \left(1 - \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right).$$

Further special cases exist for other integer and rational values of ν .

Theorem 1 Under Assumptions 1 and 2, the marginal density of E_i is given by

$$f_E(E_i = \varepsilon_i) = \frac{\sigma_V^{\mu-1} D_{-\mu}(z_i)}{\sqrt{2\pi\sigma_U^2}} \exp\left(\left(\frac{z_i}{2}\right)^2 - \frac{1}{2}\left(z_i - \frac{\sigma_V}{\sigma_U}\right)^2\right), \quad z_i = \frac{s\varepsilon_i + \sigma_V}{\sigma_U}. \tag{6}$$

Corollary 1.1 Under Assumptions 1 and 2, the conditional density of $E_i = \varepsilon_i|U_i = u_i$ is given by

$$f_{U|E}(U_i = u_i|E_i = \varepsilon_i) = \begin{cases} \frac{u_i^{\mu-1} \exp\left(-\frac{1}{2}\left(\frac{u_i}{\sigma_V}\right)^2 - \frac{u_i z_i}{\sigma_V}\right)}{\sigma_V^\mu \Gamma(\mu) \exp\left(\left(\frac{z_i}{2}\right)^2\right) D_{-\mu}(z_i)}, & u_i \geq 0. \\ 0, & u_i < 0. \end{cases} \tag{7}$$

Corollary 1.2 Under Assumptions 1 and 2,

$$\mathbb{E}[U_i|E_i = \varepsilon_i] = \mu\sigma_V \frac{D_{-\mu-1}(z_i)}{D_{-\mu}(z_i)}. \tag{8}$$

Corollary 1.3 Under Assumptions 1 and 2,

$$\mathbb{E}[\exp(-U_i)|E_i = \varepsilon_i] = \frac{\exp\left(\left(\frac{z_i + \sigma_V}{2}\right)^2\right) D_{-\mu}(z_i + \sigma_V)}{\exp\left(\left(\frac{z_i}{2}\right)^2\right) D_{-\mu}(z_i)}. \tag{9}$$

Corollary 1.4 Under Assumptions 1 and 2,

$$\text{Mode}(U_i|E_i = \varepsilon_i) = \begin{cases} 0, & z_i^2 < 4(1 - \mu). \\ \frac{\sigma_V}{2} \max\left(0, -z_i + \sqrt{z_i^2 + 4(\mu - 1)}\right), & z_i^2 \geq 4(1 - \mu). \end{cases} \tag{10}$$

Theorem 2 Under Assumptions 1 and 3, the marginal density of E_i is given by

$$f_E(E_i = \varepsilon_i) = \frac{\Gamma(2\mu)}{\Gamma(\mu)} \sqrt{\frac{2}{\pi}} \frac{\mu^\mu \sigma_V^{2\mu-1}}{\sigma^{2\mu}} D_{-2\mu}(z_i) \exp\left(\left(\frac{z_i}{2}\right)^2 - \frac{1}{2}\left(\frac{\sigma z_i}{\sigma_U}\right)^2\right),$$

$$\sigma = \sqrt{2\mu\sigma_V^2 + \sigma_U^2}, \quad \lambda = \sigma_U/\sigma_V, \quad z_i = \frac{s\varepsilon_i\lambda}{\sigma}. \tag{11}$$

Corollary 2.1 Under Assumptions 1 and 3, the conditional density of $E_i = \varepsilon_i|U_i = u_i$ is given by

$$f_{U|E}(U_i = u_i|E_i = \varepsilon_i) = \begin{cases} \frac{u_i^{2\mu-1} \exp\left(-\frac{1}{2}\left(\frac{\sigma u_i}{\sigma_V \sigma_U}\right)^2 - \frac{\sigma u_i z_i}{\sigma_V \sigma_U}\right)}{(\frac{\sigma_V \sigma_U}{\sigma})^{2\mu} \Gamma(2\mu) \exp\left(\left(\frac{z_i}{2}\right)^2\right) D_{-2\mu}(z_i)}, & u_i \geq 0. \\ 0, & u_i < 0. \end{cases} \tag{12}$$

Corollary 2.2 Under Assumptions 1 and 3,

$$\mathbb{E}[U_i|E_i = \varepsilon_i] = \frac{2\mu\sigma_V\sigma_U}{\sigma} \frac{D_{-2\mu-1}(z_i)}{D_{-2\mu}(z_i)}. \tag{13}$$

Corollary 2.3 Under Assumptions 1 and 3,

$$\mathbb{E}[\exp(-U_i)|E_i = \varepsilon_i] = \frac{\exp\left(\left(\frac{z_i}{2} + \frac{\sigma_V \sigma_U}{2\sigma}\right)^2\right) D_{-2\mu}(z_i + \frac{\sigma_V \sigma_U}{\sigma})}{\exp\left(\left(\frac{z_i}{2}\right)^2\right) D_{-2\mu}(z_i)}. \tag{14}$$

Corollary 2.4 Under Assumptions 1 and 3,

$$\text{Mode}(U_i|E_i = \varepsilon_i) = \begin{cases} 0, & z_i^2 < 4(1 - 2\mu). \\ \frac{\sigma_V \sigma_U}{2\sigma} \max\left(0, -z_i + \sqrt{z_i^2 + 4(2\mu - 1)}\right), & z_i^2 \geq 4(1 - 2\mu). \end{cases} \tag{15}$$

3.2 Estimation

Beckers and Hammond (1987) proposed direct maximum likelihood estimation of the N-G model using the parabolic cylinder function representation of the N-G density shown in Eq. (6). Specifically, the authors suggested using the definition of D_ν in terms of Kummer’s confluent hypergeometric function ${}_1F_1$ (Kummer 1837):

$$D_\nu(z) = 2^{\nu/2} \sqrt{\pi} e^{-z^2/4} \left(\frac{1}{\Gamma(\frac{1-\nu}{2})} {}_1F_1\left(-\frac{\nu}{2}; \frac{1}{2}; \frac{z^2}{2}\right) - \frac{1}{\Gamma(\frac{\nu}{2})} {}_1F_1\left(\frac{1-\nu}{2}; \frac{3}{2}; \frac{z^2}{2}\right) \right),$$

$${}_1F_1(a; b; z) = \sum_{k=0}^{\infty} \frac{(a)_k z^k}{(b)_k k!}, \tag{16}$$

where $(z)_n = \Gamma(z + n)/\Gamma(z)$ denotes the Pochhammer symbol, and calculating the function ${}_1F_1$, via this series representation, to machine precision. Efforts were made to implement this approach, but these were ultimately unsuccessful (Hammond 2023). The issue appears to be that, since the numerator and denominator of Eq. (16) may grow large very quickly for certain parameter values, we encounter the limits of floating-point arithmetic before the series has converged to machine precision, resulting in problems of arithmetic underflow or overflow. These are familiar issues in the field of computational mathematics, and implementations of special functions typically involve choosing, for a given set of argument values, the best available approximation.

The lack of such implementations of D_ν or ${}_1F_1$ in commonly-used econometric and statistical software packages still presents a hurdle to use of Eqs. (6–9) and (11–14). A later application of the N-G model to UK interwar electricity generating plant costs by Hammond (1992) used `nag_specfun_1f1_real`, an implementation of ${}_1F_1$ from the NAG Fortran library (Numerical Algorithms Group, n.d.). This is proprietary software to which many analysts may lack access, and to our knowledge, there have been no further applications. However, two relatively recent open-source implementations of D_ν are available in Python (Van Rossum and Drake 2009): the `scipy.special.pbdv` function, from the SciPy package (Virtanen et al. 2020), and the `mpmath.pcf` function from the `mpmath` package (Johansson et al. 2023). Implementations of ${}_1F_1$ are also available from these packages, and also from the `hyperg_1F1` function from the R library `gsl` (Hankin 2006), but for the sake of convenience and concision, we will prefer the parabolic cylinder representations.

Figures 2 and 3 show Eqs. (6) and (11), respectively, plotted for several instances where $s = 1$ and $\text{Var}(E_i) = 1$, but μ is allowed to vary. The parabolic cylinder function has been calculated using the `scipy.special.pbdv` function. In Fig. 2, i.e. the N-G case, $\mu = 1$ corresponds to the N-EXP case, in which case E_i follows an exponentially-modified Gaussian (EMG) distribution (Grushka 1972). By

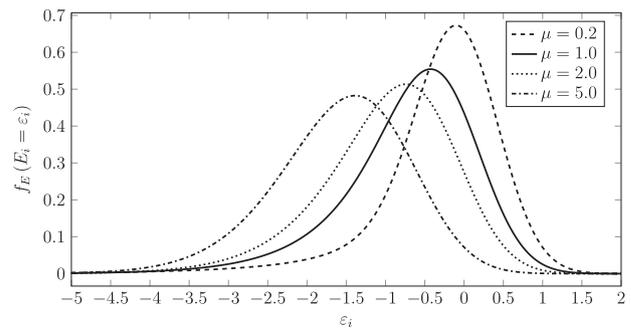


Fig. 2 Density plots for $E_i = V_i - U_i$, $V_i \sim N(0, \sigma_V^2)$, $U_i \sim \text{Gamma}(\mu, \sigma_U)$, $V_i \perp U_i$, $\text{Var}(V_i) = \text{Var}(U_i) = 0.5$

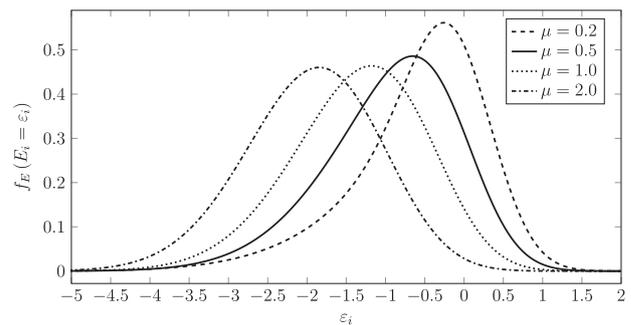


Fig. 3 Density plots for $E_i = V_i - U_i$, $V_i \sim N(0, \sigma_V^2)$, $U_i \sim \text{Nakagami}(\mu, \sigma_U)$, $V_i \perp U_i$, $\text{Var}(V_i) = \text{Var}(U_i) = 0.5$

analogy, in the more general case we will call the distribution of E_i the gamma-modified Gaussian (GMG) distribution. In Fig. 3, i.e. the N-NAK case, $\mu = 0.5$ corresponds to the N-HN case, in which case E_i follows a skew-normal distribution (Azzalini 1985). The distribution of E_i in the N-NAK model is therefore a generalised skew normal distribution. Checking against the EMG and skew normal special cases, we find identical results, as expected.

Under Assumptions 1 and 2 – i.e. in the N-G case – given Eq. (6), we derive the log-likelihood function

$$\ln L = I(\mu - 1) \ln \sigma_V - \frac{I}{2} \ln 2 - \frac{I}{2} \ln \pi - I\mu \ln \sigma_U - \frac{1}{2} \sum_{i=1}^I \left(\frac{\varepsilon_i}{\sigma_V} \right)^2 + \frac{1}{4} \sum_{i=1}^I \left(\frac{s\varepsilon_i}{\sigma_V} + \frac{\sigma_V}{\sigma_U} \right)^2 + \sum_{i=1}^I \ln D_{-\mu} \left(\frac{s\varepsilon_i}{\sigma_V} + \frac{\sigma_V}{\sigma_U} \right). \tag{17}$$

Under Assumptions 1 and 3 – i.e. in the N-NAK case – given Eq. (11), we derive the log-likelihood function

$$\ln L = I \ln \Gamma(2\mu) - I \ln \Gamma(\mu) + \frac{I}{2} \ln 2 - \frac{I}{2} \ln \pi + I\mu \ln \mu + I(2\mu - 1) \ln \sigma_V - I\mu \ln(\sigma_V^2 + 2\mu\sigma_V^2) - \frac{1}{2} \sum_{i=1}^I \left(\frac{\varepsilon_i}{\sigma_V} \right)^2 + \frac{1}{4} \sum_{i=1}^I \left(\frac{s\varepsilon_i \sigma_U / \sigma_V}{\sqrt{\sigma_V^2 + 2\mu\sigma_V^2}} \right)^2 + \sum_{i=1}^I \ln D_{-2\mu} \left(\frac{s\varepsilon_i \sigma_U / \sigma_V}{\sqrt{\sigma_V^2 + 2\mu\sigma_V^2}} \right). \tag{18}$$

Beckers and Hammond (1987) discuss derivation of the Jacobian and Hessian in the N-G case. However, the terms

involving differentiation of D_ν with respect to ν are very cumbersome and involve Kampé de Fériet functions (Kampé de Fériet 1937) and their derivatives – see results from Ancarani and Gasaneo (2008) on the derivatives of ${}_1F_1$. As such, supplying the analytical Jacobian and Hessian could complicate matters considerably, since it would require implementations of additional special functions. For our purposes, it will be much more convenient to approximate the score vector and Hessian matrix numerically.

Suitable optimisation routines are available from SciPy and other Python packages, while the NumPy package (Harris et al. 2020) includes functions for working with arrays and matrices, giving us everything we need for a Python implementation of the N-G model. Using these, we have developed a Python package called FronPy (Stead 2023) which enables estimation of the N-G and other SF models, and has been used to generate all the results shown in the present study.⁶

As noted in Section 2, the gamma and Nakagami densities are log-convex everywhere when $\mu < 1$ and $\mu < 0.5$, respectively. We typically make use of the result that strong log-concavity of f_E implies that the likelihood has only one stationary point, and that this is a maximum. In turn, the fact that log-concavity is preserved under convolution means that, if both f_V and f_U are strongly log-concave, so too is f_E . When one or both of these densities are not strongly log-concave, as in our case, this raises the possibility of multiple stationary points in the N-G and N-NAK likelihood functions. This suggests a need for careful consideration of optimisation algorithms. The use of quasi-Newton methods – which make use of the curvature information from the Hessian to set the direction of travel and ensure we converge to a maximum – offers an advantage over simpler gradient descent methods in this setting.

Fronpy uses, as a default option, a modified Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with a Wolfe-type line search to minimise the negative log-likelihood function. Such an approach has been shown to perform well in converging to global minima even when the objective function is non-convex (Li and Fukushima 2001). The Jacobian and inverse Hessian are calculated numerically, and the latter is used to derive the covariance matrix. The reparameterisation

$$\theta' = (\beta' \quad \ln \sigma_V \quad \ln \sigma_U \quad \ln \mu)'$$

is used to ensure positive values of σ_V , σ_U , and μ . Approaches taken to selecting starting values are discussed in the Supplementary Appendices. Following estimation of

the model, FronPy also calculates the efficiency predictors given in Eqs. (8–10) and (13–15).

4 Application

In this section, we demonstrate three applications of the N-G and N-NAK models to real data, and compare the results to those from N-HN, N-EXP, and N-TN models. Application 1 is to the Christensen and Greene (1976) dataset on the costs, outputs, and input prices of a 1970 cross-section of 123 US electricity generating firms. Application 2 is to the 1990-1997 panel dataset on the outputs and inputs of 43 smallholder rice producers in the Philippines used by Coelli et al. (2005) and others. Application 3 is to the dataset on interwar electricity generating plants in Great Britain used by Foreman-Peck and Waterson (1985) and Hammond (1992), consisting of a cross-section of 184 steam generating plants supplying the Central Electricity Board in 1937 – of which 129 were municipally operated, and the remainder were privately operated.

In these applications, we adopt the functional forms used by Greene (2003), Coelli et al. (2005), and Hammond (1992), respectively. In the latter case, we have a particular interest in comparing our results from our N-G model to those obtained by Hammond (1992) who, as discussed previously, provides – to our knowledge – the sole example of a direct maximum likelihood estimation of the N-G model prior to our own.

In each case, we estimate models as described by Assumption 1 under several different assumptions – Assumptions 2 and 3, i.e. N-G and N-NAK models, and their N-EXP and N-HN special cases, along with a N-TN model for comparison. All models are estimated using FronPy. In Application 1,

$$y_i = \ln\left(\frac{c_i}{e_i}\right), \quad s = -1,$$

$$x'_i \beta = \beta_0 + \beta_1 \ln q_i + \beta_2 \ln^2 q_i + \beta_3 \ln\left(\frac{w_i}{e_i}\right) + \beta_4 \ln\left(\frac{r_i}{e_i}\right),$$

where c_i is total cost, q_i is output in millions of kilowatt-hours generated, w_i is the price of labour, r_i is the price of capital, and e_i is the price of fuel. In Application 2,

$$y_{it} = \ln q_{it}, \quad s = 1,$$

$$x'_{it} \beta = \beta_0 + \beta_1 \ln x_{1it} + \beta_2 \ln x_{2it} + \beta_3 \ln x_{3it}$$

$$+ \beta_{11} \ln^2 x_{1it} + \beta_{12} \ln x_{1it} \ln x_{2it} + \beta_{13} \ln x_{1it} \ln x_{3it}$$

$$+ \beta_{22} \ln^2 x_{2it} + \beta_{23} \ln x_{2it} \ln x_{3it} + \beta_t t,$$

where q_{it} is tonnes of freshly threshed rice, x_{1it} is hectares of land planted, x_{2it} is days of hired and family labour, and x_{3it} is kilograms of nitrogen, phosphorous, and potassium

⁶ see <https://github.com/AlexStead/FronPy> for instructions on installation and usage.

(NPK) fertiliser used. In Application 3,

$$\begin{aligned}
 y_i &= \ln\left(\frac{vc_i}{q_i}\right), \quad s = -1, \\
 \mathbf{x}'_i \boldsymbol{\beta} &= \beta_0 + \beta_q \ln q_i + \beta_{qq} (\ln q_i)^2 + \beta_r \ln r_i \\
 &+ \beta_{rr} (\ln r_i)^2 + \beta_o o_i + \beta_{qr} \ln q_i \ln r_i \\
 &+ \beta_{ro} o_i \ln r_i + \beta_{qo} o_i \ln q_i + \beta_c \ln c_i + \beta_s \ln s_i \\
 &+ \beta_w w_i + \beta_{cc} (\ln c_i)^2 + \beta_{ss} (\ln s_i)^2 \\
 &+ \beta_{cs} \ln c_i \ln s_i + \beta_{cw} w_i \ln c_i + \beta_{sw} w_i \ln s_i \\
 &+ \beta_{cq} \ln c_i \ln q_i + \beta_{sq} \ln s_i \ln q_i \\
 &+ \beta_{wq} w_i \ln q_i + \beta_{cr} \ln c_i \ln r_i + \beta_{sr} \ln s_i \ln r_i \\
 &+ \beta_{wr} w_i \ln r_i + \beta_{co} c_i o_i \\
 &+ \beta_{so} o_i \ln s_i + \beta_{wo} w_i o_i + \beta_k \ln k_i + \beta_{kk} (\ln k_i)^2 \\
 &+ \beta_{kq} \ln k_i \ln q_i + \beta_{kr} \ln k_i \ln r_i \\
 &+ \beta_{ok} o_i \ln k_i + \beta_{ck} \ln c_i \ln k_i \\
 &+ \beta_{sk} \ln s_i \ln k_i + \beta_{wk} w_i \ln k_i,
 \end{aligned}$$

where vc_i is variable costs, q_i is kilowatt-hours generated, r_i is peak capacity utilisation rate, s_i is average salary (of salaried staff), c_i is a coal price, k_i is the maximum generating capacity, o_i and w_i are binary dummies identifying, respectively, continuously-operated plants and plants in London – the latter a proxy for wages.

In all three applications, all independent variables except binary dummies are divided by their sample means before taking logarithms, so that their first-order coefficients may be interpreted as elasticities at the sample means. Note that this means that the first-order coefficients in Application 3 will not be comparable to those presented by Hammond (1992), who did not rescale the data in this way.

4.1 Parameter estimates

Tables 3–5 show selected parameter estimates, standard errors, and significance stars, along with log-likelihoods, for each model, for each of our three applications. Note that, for the sake of brevity, Tables 4 and 5 report only first-order coefficient estimates, along with the distributional parameters and log-likelihoods. Full sets of parameter estimates, including second-order frontier coefficients, are included for replication purposes in the Supplementary Appendices.

Across all three applications, we can see that the parameter estimates and log-likelihoods produced by `FRONPY` for the N-HN and N-EXP models match those reported in Greene (2003), Coelli et al. (2005), and Hammond (1992), respectively.⁷ Strikingly, however, none of our N-TN models converged successfully, and it is immediately

Table 3 Parameter estimates – Application 1

	N-EXP	N-G	N-HN	N-NAK	N-TN
β_0	3.7636 (0.0205)	3.8231 (0.0282)	3.7349 (0.0369)	3.8302 (0.0209)	3.7632 (0.0202)
β_1	0.9664 (0.0129)	0.9639 (0.0120)	0.9659 (0.0131)	0.9638 (0.0124)	0.9664 (0.0119)
β_2	0.0287 (0.0026)	0.0273 (0.0026)	0.0303 (0.0025)	0.0272 (0.0027)	0.0288 (0.0023)
β_3	0.2701 (0.0651)	0.2787 (0.0670)	0.2606 (0.0627)	0.2786 (0.0607)	0.2700 (0.0635)
β_4	0.0332 (0.0599)	0.0216 (0.0585)	0.0553 (0.0583)	0.0206 (0.0565)	0.0334 (0.0556)
$\ln \sigma_V$	-2.2599 (0.1351)	-2.1896 (0.1041)	-2.2179 (0.2098)	-2.1808 (0.0937)	-2.2600 (0.1346)
$\ln \sigma_U$	-2.3285 (0.2403)	-1.4297 (0.7344)	-1.9008 (0.3547)	-2.2989 (0.2809)	0.4953 (0.3616)
$\ln \mu$	0.00000 –	-1.7442 (1.2238)	-0.69315 –	-3.1774 (0.8795)	– –
σ_V	0.1044 (0.0141)	0.1120 (0.0117)	0.1088 (0.0228)	0.1130 (0.0106)	0.1044 (0.0140)
σ_U	0.0974 (0.0234)	0.2394 (0.1758)	0.1494 (0.0530)	0.1004 (0.0282)	1.6410 (0.5934)
μ	1.0000 –	0.1748 (0.2139)	0.5000 –	0.0417 (0.0367)	-27.3703 (18.9008)
$\ln L$	67.9609	68.7326	66.8649	69.0722	67.9504

Standard errors in parentheses

noticeable that the β values and log-likelihoods are practically identical to those from the N-EXP models.⁸ The values of μ and σ_U are negative and positive, respectively, and large in magnitude across all three applications. This suggests that the issue identified by Meesters (2014) in which the N-TN model approaches the N-EXP at this boundary of its parameter space is a real, practical problem, and one that may occur very frequently in applications to real data.

In contrast, our N-G and N-NAK models converged successfully in all three applications, and yield results different from those of the N-HN and N-EXP models. Of immediate interest here are the estimates of the shape parameters from these models. In each of our N-NAK models, the estimated μ parameters are less than the N-HN special case of $\mu = 0.5$. Similarly, two of our three N-G models have estimates of μ below the N-EXP special case of $\mu = 1$. These suggests that the N-G and N-NAK models may be useful primarily in allowing for distributions of U_i with essentially exponential or Gaussian tails, respectively, but greater concentrations around zero than their special

⁷ In the case of Hammond (1992), note however that we are comparing with a reported N-TN model in which $\mu = 0$, making it effectively an N-HN model.

⁸ Note that the N-TN results presented by Coelli et al. (2005) and Hammond (1992) differ. The reasons for this are unclear, but note that our log-likelihoods are higher. We were also able to replicate the same phenomenon reported here using other software packages.

Table 4 Selected parameter estimates – Application 2

	N-EXP	N-G	N-HN	N-NAK	N-TN
β_0	2.0608 (0.0446)	2.0814 (0.1025)	2.1523 (0.0394)	2.0002 (0.0584)	2.0656 (0.0406)
β_1	0.5250 (0.0882)	0.5248 (0.0784)	0.5314 (0.0795)	0.5186 (0.0780)	0.5242 (0.0787)
β_2	0.2477 (0.0847)	0.2469 (0.0738)	0.2309 (0.0744)	0.2482 (0.0728)	0.2474 (0.0743)
β_3	0.2020 (0.0551)	0.2021 (0.0441)	0.2033 (0.0450)	0.2064 (0.0436)	0.2025 (0.0443)
β_t	0.0142 (0.0077)	0.0143 (0.0065)	0.0151 (0.0065)	0.0140 (0.0066)	0.0142 (0.0066)
$\ln \sigma_V$	-1.6743 (0.0973)	-1.6985 (0.1459)	-1.8284 (0.1218)	-1.6350 (0.1164)	-1.6844 (0.1011)
$\ln \sigma_U$	-1.3546 (0.1045)	-1.4200 (0.3073)	-0.8151 (0.0718)	-1.1515 (0.1500)	0.6422 (0.0583)
$\ln \mu$	0.0000 –	0.1440 (0.6373)	-0.6932 –	-1.7094 (0.4117)	– –
σ_V	0.1874 (0.0182)	0.1830 (0.0267)	0.1607 (0.0196)	0.1949 (0.0227)	0.1855 (0.0188)
σ_U	0.2581 (0.0270)	0.2417 (0.0743)	0.4426 (0.0318)	0.3162 (0.0474)	1.9006 (0.1108)
μ	1.0000 –	1.1548 (0.7359)	0.5000 –	0.1810 (0.0745)	-13.2289 (0.0404)
$\ln L$	-71.3256	-71.2989	-74.4099	-71.6658	-71.3133

Standard errors in parentheses

cases – or the N-TN model – can accommodate. In Application 2, however, our N-G model has $\mu = 1.1548$ and implies $\text{Mode}(U_i) = 0.1820$, $\text{Mode}(\exp(-U_i)) = 0.8336$, indicating that it can also be useful to allow for $\text{Mode}(U_i) > 0$.

Note that, in Application 3, our results differ from those of Hammond (1992), who reports an estimate of $\mu = 0.1165$ – our estimate of $\mu = 0.0472$ is even lower. Given that our log-likelihood is higher – 63.0581 compared to 60.7078 – it appears that Hammond (1992) may have encountered some issue in convergence, though his estimates certainly appear to have been heading in the same direction. We were unfortunately unable to obtain the code used by Hammond (1992). However, we found that our `FronPy` package performed well in Monte Carlo simulations for a variety of N-G data generating processes – results are reported in the Supplementary Appendices.

4.2 Optimisation times

As discussed in Section 3, several different approaches have previously been used to approximate the N-G likelihood, which vary with respect to accuracy and computational intensity. One

Table 5 Selected parameter estimates – Application 3

	N-EXP	N-G	N-HN	N-NAK	N-TN
β_0	-1.9075 (0.1446)	-1.7486 (0.1165)	-1.9171 (0.1338)	-1.9531 (0.1308)	-1.9081 (0.3396)
β_q	-0.5150 (0.0818)	-0.5332 (0.0817)	-0.4844 (0.0863)	-0.5567 (0.0766)	-0.5148 (0.0822)
β_r	-0.2967 (0.4329)	0.0221 (0.3999)	-0.3247 (0.4328)	-0.4440 (0.3749)	-0.2988 (1.4358)
β_o	0.0007 (0.2433)	0.1577 (0.2337)	-0.0277 (0.2338)	0.0363 (0.2067)	-0.0002 (0.5485)
β_c	0.5734 (0.2928)	0.3507 (0.2852)	0.6038 (0.3105)	0.5883 (0.2571)	0.5756 (1.5211)
β_s	1.4863 (1.7235)	0.9246 (1.5486)	1.1317 (1.6804)	2.4034 (1.5533)	1.4899 (4.3322)
β_w	0.2803 (0.3286)	0.3011 (0.3097)	0.2658 (0.3500)	0.3611 (0.3060)	0.2799 (0.6383)
β_k	0.2899 (0.2195)	0.4005 (0.1969)	0.3189 (0.2134)	0.1450 (0.2028)	0.2894 (0.5207)
$\ln \sigma_V$	-2.1696 (0.1786)	-1.8901 (0.0771)	-2.2868 (0.2460)	-2.1154 (0.0958)	-2.1721 (0.8572)
$\ln \sigma_U$	-1.9348 (0.2023)	-0.5848 (0.7717)	-1.3543 (0.1477)	-1.7206 (0.1719)	1.0556 (0.7025)
$\ln \mu$	0.00000 –	-3.0527 (1.1932)	-0.69315 –	-2.8671 (0.4696)	– –
σ_V	0.1142 (0.0204)	0.1511 (0.0117)	0.1016 (0.0250)	0.1206 (0.0115)	0.1139 (0.0977)
σ_U	0.1445 (0.0292)	0.5572 (0.4300)	0.2581 (0.0381)	0.1790 (0.0308)	2.8736 (2.0187)
μ	1.0000 –	0.0472 (0.0564)	0.5000 –	0.0569 (0.0267)	-56.6402 (45.4836)
$\ln L$	60.4193	63.0581	54.7594	62.2609	60.3847

Standard errors in parentheses

advantage of the direct implementation in `FronPy`, aside from accuracy, is speed of computation. Table 6 compares, for all three applications, the time taken to maximise the likelihood for each model. Note that optimisation times will vary depending on hardware used, and may be expected to fluctuate whenever the model is re-run depending on, e.g., CPU usage, so the times given here are merely indicative.

As expected, given the additional parameter, optimisation times for the N-G and N-NAK models are slightly longer than for the N-HN and N-EXP models, but broadly in line with those for the N-TN models. All the models converge quickly, in a few seconds or fractions of a second. In contrast, methods such as maximum simulated likelihood may take anywhere from tens of seconds to several minutes, depending on sample size and numbers of pseudorandom or quasirandom draws used.

Table 6 Model optimisation times

Application	Model	Time (seconds)
1	N-EXP	0.0796
	N-G	0.1831
	N-HN	0.1505
	N-NAK	0.1739
	N-TN	0.5467
2	N-EXP	0.1947
	N-G	0.5641
	N-HN	0.3629
	N-NAK	0.9188
	N-TN	0.8495
3	N-EXP	0.5330
	N-G	3.1678
	N-HN	0.8227
	N-NAK	3.0312
	N-TN	3.5505

Table 7 Normal-gamma model optimisation times – maximisation of the exact likelihood (ML) versus the inverse fast Fourier transform approximation (ML-FFT)

Application	Observations	Parameters	Time (seconds)	
			ML	ML-FFT
1	123	8	0.1831	31.0762
2	43	14	0.5641	27.3343
3	184	37	3.1678	189.6788

As discussed in Section 2.2, we compare the time taken to optimise the N-G model by maximising the exact likelihood to the time taken when using the inverse FFT approximation of the likelihood proposed by Tsionas (2012). This depends upon the number of grid points used to approximate the density. When using 2^{12} , as suggested by Tsionas (2012), we find that optimisation times are comparable to those using the exact expression, but that the accuracy of the estimated log-likelihoods could be improved; in the case of Application 3, we found that the accuracy was not sufficient for the estimation algorithm to declare convergence. Increasing the number of points to 2^{18} dealt with this issue and resulted in estimated log-likelihoods and parameter estimates accurate to four decimal places, at the cost of slowing optimisation down considerably.

Table 7 compares optimisation times for our N-G models when using exact expressions and when using the inverse FFT approximation with 2^{18} grid points. With the preceding discussion in mind, there is a clear trade-off between accuracy and optimisation time when using the inverse FFT approximation. By using the exact expressions, we avoid

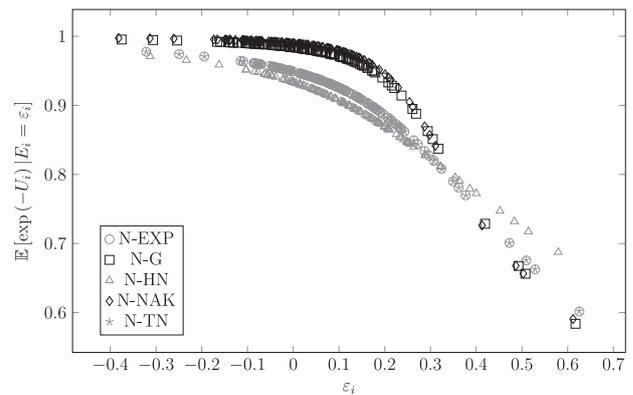


Fig. 4 Plots of efficiency predictions against residuals – Application 1

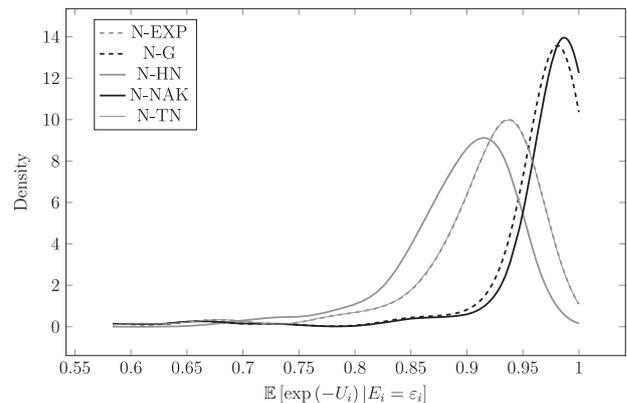


Fig. 5 Kernel density estimates for efficiency predictions – Application 1

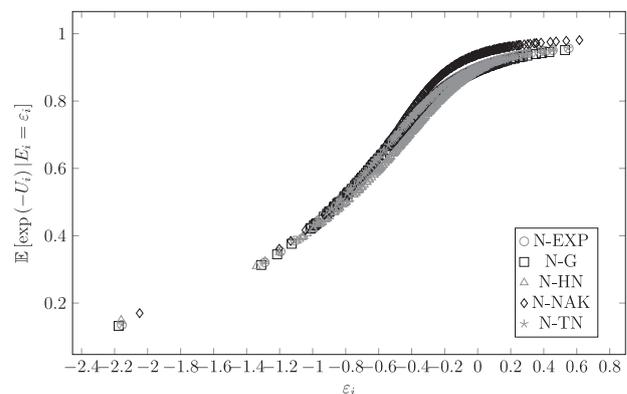


Fig. 6 Plots of efficiency predictions against residuals – Application 2

this trade-off and benefit from a potentially significant reduction in optimisation time.

4.3 Efficiency predictions

Figures 4–9 show scatter plots of $\mathbb{E}[\exp(-U_i)|E_i = \epsilon_i]$ against estimated residuals, and Gaussian kernel density estimates of the distributions of the $\mathbb{E}[\exp(-U_i)|E_i = \epsilon_i]$, across Applications 1–3. Following from the discussion in Section 4.1, we can see that across all three applications, the

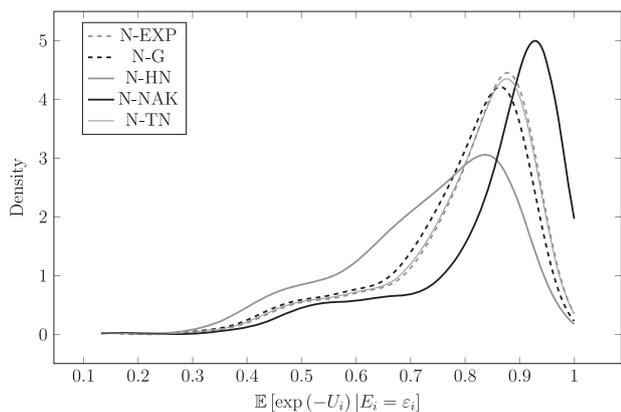


Fig. 7 Kernel density estimates for efficiency predictions – Application 2

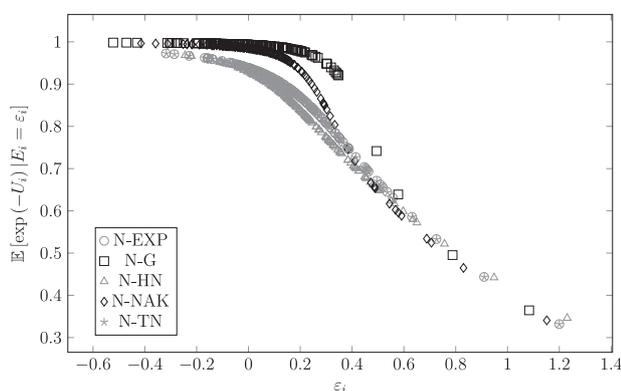


Fig. 8 Plots of efficiency predictions against residuals – Application 3

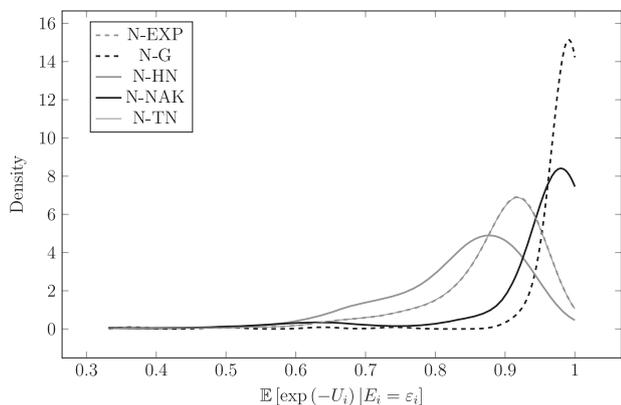


Fig. 9 Kernel density estimates for efficiency predictions – Application 3

N-TN and N-EXP models yield practically identical efficiency scores. Again, this highlights the Meesters (2014) result as a real practical issue which appears to mean that, in many cases, the N-TN model offers little advantage over simpler specifications.

By contrast, with the exception of the N-G model in Application 2, which is not far from the N-EXP special case of $\mu = 1$, the N-G and N-NAK models yield efficiency

predictions noticeably different to those from the N-EXP and N-HN models, with a much greater concentration near 1. This is a reflection of the fact that the estimated values of μ in these cases are far below the values of $\mu = 1$ and $\mu = 0.5$ that give the N-EXP and N-HN special cases. As can be seen in Tables 1 and 2, this means that the densities of U_i are even more concentrated at zero than in the N-HN or N-EXP cases.

The motivation for the N-G and N-NAK models is nicely illustrated by the fact that they give us a significantly different picture than the N-HN and N-EXP models regarding the distribution of efficiency scores, while the N-TN mimics the N-EXP in these cases. We see, in Applications 1 and 3, the concentration of a significant number of firms near the frontier, similar to what a ZISF model might yield. Regarding the differences between the N-G and N-NAK models, we see that in Application 1 they yield highly similar efficiency predictions, while there are significant differences in Application 2; the latter likely due to differences in the sensitivity of the two models to the outlying observations that can be seen in that application, given that the tails of the Nakagami and gamma distributions are essentially Gaussian and exponential, respectively.

It is worth remarking on the economic significance of some of these results. Results from the N-G and N-NAK models in Applications 1 and 3 – relating to electricity generation plants in the US and UK, respectively – suggest a large proportion of plants on or near the frontier. In the UK case, the mean efficiency prediction was 0.9869 among municipally operated plants, and 0.9754 among privately operated plants according to the N-G model (0.9596 and 0.9339 respectively according to the N-NAK model). Different ownership models do not appear to have been decisive. This makes sense given that both types of plant had an incentive to operate efficiently to reduce their unit costs and improve their standing in the ‘merit order’ operated by the Central Electricity Board in order to minimise the cost of supply (Hammond 1992). In contrast, results for Application 2 – relating to rice farms in the Philippines – suggest much lower efficiency scores on average and, in the N-G case, a non-zero mode for inefficiency – results that appear to make sense in the context of significant subsidies, widespread family ownership, and limited technological diffusion. The estimated shapes and scales of the efficiency distributions therefore appear intuitive in economic terms.

4.4 Likelihood ratio tests

If we are concerned with testing down from more flexible models to more restrictive special cases, are three likelihood ratio (LR) tests of interest in each of Applications 1–3. These are testing down from the N-G to the N-EXP, from

Table 8 Likelihood ratio tests

Application	Alternative model	Null model	Likelihood ratio
1	N-G	N-EXP	1.5434
	N-NAK	N-HN	4.4146*
	N-TN	N-HN	2.1709
2	N-G	N-EXP	0.0533
	N-NAK	N-HN	5.4882*
	N-TN	N-HN	6.1933*
3	N-G	N-EXP	5.2776*
	N-NAK	N-HN	15.0030***
	N-TN	N-HN	11.2508***

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

the N-NAK to the N-HN, and from the N-TN. The corresponding null hypotheses are $H_0: \mu = 1$, $H_0: \mu = 0.5$, and $H_0: \mu = 0$, respectively. These each involve only one restriction, and therefore the LR statistics follow a χ_1^2 distribution (Wilks 1938). Table 8 shows the likelihood ratio in each case, along with significance stars.

The results of these tests vary significantly between applications. Although the differences between the N-G and N-NAK models and their special cases are economically significant – in terms of the different pictures they give of efficiency levels – the improvements in the log-likelihoods are modest. From the N-NAK models we only weakly reject the N-HN null models in Applications 1 and 2, but strongly reject it in Application 3, and from the N-G models we fail to reject the null N-EXP models in Applications 1 and 2, but weakly reject it in Application 3. In terms of comparing the N-G and N-NAK models, note that both the gamma and Nakagami distributions belong to the generalised gamma family (Stacey 1962), so one approach is to simply choose the higher log-likelihood. On this basis, the N-NAK model would be preferred in Application 1, while the N-G model would be preferred in Applications 2 and 3.

5 Conclusion

Much prior motivation for the N-G model has focused on its ability to accommodate non-zero modes of inefficiency. Our empirical applications suggest that the N-G and N-NAK models may in fact be more useful for their ability to accommodate a *greater* concentration of inefficiencies near zero than their N-EXP and N-HN special cases permit. The resulting efficiency predictions place most firms very close to the frontier, and thus may tell a more plausible story in the context of competitive markets or effective regulation.

We have derived closed-form solutions for the log-likelihoods, efficiency predictors, and other key results for the normal-gamma (N-G) and normal-Nakagami (N-NAK)

stochastic frontier (SF) models. The relevant functions are expressed in terms of the parabolic cylinder function, mirroring a similar derivation of the N-G likelihood – in terms of the confluent hypergeometric function ${}_1F_1$ – by Beckers and Hammond (1987).

The fact that the N-G likelihood involves special functions not typically found in standard statistical and econometric software packages has hitherto been a barrier to direct maximum likelihood estimation – bar a lone example in Hammond (1992) – and for this reason several approaches to approximating the density have been proposed. Our direct implementation offers improvements in stability, accuracy and speed. In particular, use of exact expressions avoids issues of spectral leakage that arise when applying the inverse fast Fourier transform approximation proposed by Tsionas (2012). Applications to real and simulated data suggest that our implementation – included as part of the Python package `FronPy` – performs well.

The value of this work lies not only in the fact that estimation of the N-G model has been a longstanding challenge in the SF literature, but also because the N-G and N-NAK models appear to perform better than the normal-truncated normal (N-TN) model in applications to real data. In all three of our applications, we note that the N-TN model fails to converge due to an issue noted by Meesters (2014): that the N-TN model can approach the normal-exponential at a boundary of its parameter space. In contrast the N-G and N-NAK models work well in all of our applications. The N-NAK model therefore emerges as a potentially more attractive generalisation of the normal-half normal (N-HN) model.

This paper has focused on alternative inefficiency distributions. In recent years, the importance of the assumed noise distribution has also gained attention – see Papadopoulos (2023) for a recent review. The pairing of gamma and Nakagami inefficiency distributions with alternative noise distributions may be an interesting direction for future research.

Data availability

Guidance on accessing the data used in this study can be found in the supplementary material files.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1007/s11123-024-00742-2>.

Acknowledgements The author thanks Christopher Hammond, Phill Wheat, the editor, two anonymous referees, and participants at the North American Productivity Workshop XI for their helpful comments and suggestions. Any remaining errors are the author's own. The author gratefully acknowledges funding from the CQC Efficiency Network (see https://nhtnetwork.org/nht_product/cqc-efficiency/).

Author contributions The author confirms sole responsibility for the following: study conception and design, analysis and interpretation of results, and manuscript preparation.

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aigner D, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models *J Econ* 6:21–37. [https://doi.org/10.1016/0304-4076\(77\)90052-5](https://doi.org/10.1016/0304-4076(77)90052-5)
- Ancarani LU, Gasaneo G (2008) Derivatives of any order of the confluent hypergeometric function ${}_1F_1(a, b, z)$ with respect to the parameter a or b . *J Math Phys* 49:063508. <https://doi.org/10.1063/1.2939395>
- Azzalini A (1985) A class of distributions which includes the normal ones. *Scand J Stat* 12:171–8. <http://www.jstor.org/stable/4615982>
- Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *J Econ* 38:387–99. [https://doi.org/10.1016/0304-4076\(88\)90053-X](https://doi.org/10.1016/0304-4076(88)90053-X)
- Beckers DE, Hammond CJ (1987) A tractable likelihood function for the normal-gamma stochastic frontier model. *Econ Lett* 24:33–8. [https://doi.org/10.1016/0165-1765\(87\)90177-7](https://doi.org/10.1016/0165-1765(87)90177-7)
- Bersch J, Diekhof J, Krieger B, Licht G, Murmann S (2019) Productivity slowdown, innovation and industry dynamics. In Lehmann EE, Keilbach M (eds) *From Industrial Organization to Entrepreneurship: A Tribute to David B. Audretsch*, 229–41. Springer, Cham, Switzerland. <https://doi.org/10.1007/978-3-030-25237-3>
- Bloom N, Van Reenen J (2007) Measuring and explaining management practices across firms and countries. *Q J Econ* 122:1351–408. <https://doi.org/10.1162/qjec.2007.122.4.1351>
- Brigham EO (1988) *The Fast Fourier Transform and Its Applications*. Prentice-Hall Signal Processing Series, 1st ed. Prentice-Hall, Englewood Cliffs, NJ
- Christensen LR, Greene WH (1976) Economies of scale in U.S. electric power generation. *J Political Econ* 84:655–76. <https://doi.org/10.1086/260470>
- Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE (2005) *An Introduction to Efficiency and Productivity Analysis*, 2nd ed. Springer, New York, NY. <https://doi.org/10.1007/b136381>
- de Andrade BB, Souza GS (2018) Likelihood computation in the normal-gamma stochastic frontier model. *Computational Stat* 33:967–82. <https://doi.org/10.1007/s00180-017-0768-5>
- Fernald J, Inklaar R, Ruzic D (2023) The productivity slowdown in advanced economies: Common shocks or common trends? *Rev Income Wealth* ([Advance online publication]). <https://doi.org/10.1111/roiw.12690>
- Foreman-Peck J, Waterson M (1985) The comparative efficiency of public and private enterprise in Britain: Electricity generation between the world wars. *Economic J* 95:83–95. <https://doi.org/10.2307/2232872>
- Gordon RJ, Sayed H (2019) The industry anatomy of the transatlantic productivity growth slowdown. Working Paper w25703, National Bureau of Economic Research, Cambridge, MA. <https://doi.org/10.3386/w25703>
- Gradshteyn IS, Ryzhik IM (2014) *Table of integrals, series, and products*, 8th ed. Academic Press, Amsterdam, Netherlands. <https://doi.org/10.1016/C2010-0-64839-5>
- Greene W (1990) A gamma-distributed stochastic frontier model. *J Econ* 46:141–63. [https://doi.org/10.1016/0304-4076\(90\)90052-U](https://doi.org/10.1016/0304-4076(90)90052-U)
- Greene W (2003) Simulated likelihood estimation of the normal-gamma stochastic frontier function. *J Product Anal* 19:179–90. <https://doi.org/10.1023/A:1022853416499>
- Grushka E (1972) Characterization of exponentially modified Gaussian peaks in chromatography. *Anal Chem* 44:1733–8. <https://doi.org/10.1021/ac60319a011>
- Hajargasht G (2021) Stochastic frontiers with a Rayleigh distribution. *J Product Anal* 44:199–208. <https://doi.org/10.1007/s11123-014-0417-8>
- Hammond CJ (2023) Personal communication
- Hammond CJ (1992) Privatisation and the efficiency of decentralised electricity generation: Some evidence from inter-war Britain. *Economic J* 102:538–553. <https://doi.org/10.2307/2234290>
- Hankin RKS (2006) Special functions in R: introducing the gsl package. *R News* 6. <https://cran.r-project.org/web/packages/gsl/vignettes/gsl.pdf>
- Harris R, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, Fernández del Río J, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE (2020) Array programming with NumPy. *Nature* 585:357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Horrace WC, Parmeter CF (2018) A Laplace stochastic frontier model. *Econom Rev* 37:260–80. <https://doi.org/10.1080/07474938.2015.1059715>
- Horrace WC, Wright IA (2020) Stationary points for parametric stochastic frontier models. *J Bus Economic Stat* 38:516–26. <https://doi.org/10.1080/07350015.2018.1526088>
- Johansson F, Peterson P, Pernici M, Certik O, Steinberg V, Telang N, Taschuk M, Van Harsen C, Baayen J, Smith C, Arias de Reyna J, Tziakos I, Meurer A, Krastanov S, Allen K, Hartmann T, Kirpichev SB, Kuhlman K, Masson P, Kagalenko M, Warner J, Gaukler M, Navas-Palencia G, Dattani N (2023) mpmath: A Python library for arbitrary-precision floating-point arithmetic (version 1.3.0). <http://mpmath.org/>
- Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. *J Econ* 19:233–8. [https://doi.org/10.1016/0304-4076\(82\)90004-5](https://doi.org/10.1016/0304-4076(82)90004-5)
- Kampé de Fériet MJ (1937) La fonction hypergéométrique. *Mémorial des Sci Math* 85. http://www.numdam.org/item/MSM_1937__85__1_0.pdf
- Kumbhakar SC, Parmeter CF, Tsionas EG (2013) A zero inefficiency stochastic frontier model. *J Econ* 172:66–76. <https://doi.org/10.1016/j.jeconom.2012.08.021>
- Kumbhakar SC, Parmeter CF, Zelenyuk V (2020) Stochastic frontier analysis: Foundations and advances I. In Ray SC, Chambers RG, Kumbhakar SC (eds) *Handbook of Production Economics*, 1–40. Springer, Singapore. https://doi.org/10.1007/978-981-10-3450-3_9-2
- Kummer EE (1837) De integralibus quibusdam definitis et seriebus infinitis". *J für die reine und angew Mathematik* 17:228–42. <https://doi.org/10.1515/crll.1837.17.228>
- Li D-H, Fukushima M (2001) A modified BFGS method and its global convergence in nonconvex minimization. *J Computational Appl Math* 129:15–35. [https://doi.org/10.1016/S0377-0427\(00\)00540-9](https://doi.org/10.1016/S0377-0427(00)00540-9)

- Li Q (1996) Estimating a stochastic production frontier when the adjusted error is symmetric. *Econ Lett* 52:221–8. [https://doi.org/10.1016/S0165-1765\(96\)00857-9](https://doi.org/10.1016/S0165-1765(96)00857-9)
- Meesters A (2014) A note on the assumed distributions in stochastic frontier models. *J Product Anal* 42:171–3. <https://doi.org/10.1007/s11123-014-0387-x>
- Meusen W, van Den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. *Int Economic Rev* 18:435–44. <https://doi.org/10.2307/2525757>
- Nakagami M (1960) The m -distribution—a general formula of intensity distribution of rapid fading. In *Statistical Methods in Radio Wave Propagation: Proceedings of a Symposium Held at the University of California, Los Angeles, June 18–20, 1958* 3–36. The University of California, Los Angeles. <https://doi.org/10.1016/B978-0-08-009306-2.50005-4>
- Numerical Algorithms Group (n.d.). The NAG Library for Fortran. www.nag.com. Oxford, United Kingdom, n.d.
- Oliveira-Cunha J, Kozler J, Shah P, Thwaites G, Valero A (2021) Business time: How ready are UK firms for the decisive decade? *The Economy 2030 Inquiry*, London, UK. <http://eprints.lse.ac.uk/id/eprint/117345>
- Papadopoulos A (2021) Stochastic frontier models using the generalized exponential distribution. *J Product Anal* 55:15–29. <https://doi.org/10.1007/s11123-020-00591-9>
- Papadopoulos A (2023) The noise error component in stochastic frontier analysis. *Empir Econ* 64:2795–829. <https://doi.org/10.1007/s00181-022-02339-w>
- Papadopoulos A, Parmeter CF (2021) Type II failure and specification testing in the stochastic frontier model. *Eur J Operational Res* 293:990–1001. <https://doi.org/10.1016/j.ejor.2020.12.065>
- Papadopoulos A, Parmeter CF (2024) The wrong skewness problem in stochastic frontier analysis: A review. *J Product Anal* 61:121–34. <https://doi.org/10.1007/s11123-023-00708-w>
- Rho S, Schmidt P (2015) Are all firms inefficient? *J Product Anal* 43:327–49. <https://doi.org/10.1007/s11123-013-0374-7>
- Ritter C, Simar L (1997) Pitfalls of normal-gamma stochastic frontier models. *J Product Anal* 8:167–82. <https://doi.org/10.1023/A:1007751524050>
- Stacey EW (1962) A generalization of the gamma distribution. *Ann Math Stat* 33:1187–92. <https://doi.org/10.1214/aoms/1177704481>
- Stead AD (2023) FronPy: A Python package for frontier analysis, <https://github.com/AlexStead/FronPy>. Python package version 0.0.0
- Stead AD, Wheat P, Greene WH (2019) Distributional forms in stochastic frontier analysis. In ten Raa, T, Greene W H (eds) *Palgrave Handbook of Economic Performance Analysis*, chap. 8, 225–274. Palgrave Macmillan, Cham, Switzerland. https://doi.org/10.1007/978-3-030-23727-1_8
- Stevenson RE (1980) Likelihood functions for generalized stochastic frontier estimation. *J Econ* 13:57–66. [https://doi.org/10.1016/0304-4076\(80\)90042-1](https://doi.org/10.1016/0304-4076(80)90042-1)
- Tsionas EG (2012) Maximum likelihood estimation of stochastic frontier models by the Fourier transform. *J Econ* 170:234–48. <https://doi.org/10.1016/j.jeconom.2012.04.001>
- Van Rossum G, Drake F L (2009) *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA
- Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat Methods* 17:261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Whittaker ET (1902) On the functions associated with the parabolic cylinder in harmonic analysis. *Proc Lond Math Soc* s1-35:417–27. <https://doi.org/10.1112/plms/s1-35.1.417>
- Wilks SS (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann Math Stat* 9:60–2. <https://doi.org/10.1214/aoms/1177732360>
- Yuengert AM (1993) The measurement of efficiency in life insurance: Estimates of a mixed normal-gamma error model. *J Bank Financ* 17:483–96. [https://doi.org/10.1016/0378-4266\(93\)90047-H](https://doi.org/10.1016/0378-4266(93)90047-H)