Distributional Forms in Stochastic Frontier Analysis

Alexander D. Stead, Phill Wheat and William H. Greene

1   Introduction

This chapter provides an overview of a particular aspect of stochastic frontier analysis (SFA). The SF model is typically used to estimate best-practice 'frontier' functions that explain production or cost and predict firm efficiency relative to these. Extensive reviews of the broad stochastic frontier (SF) methodology are undertaken by Kumbhakar and Lovell (2000), Murillo-Zamorano (2004), Coelli et al. (2005), Greene (2008), and Parmeter and Kumbhakar (2014). This review will focus on the many different uses of various distributional forms.

Section 2 begins with a brief account of the motivation and development of efficiency analysis and prediction based on the standard SF model. A key feature of SF models is the focus on unobserved disturbance in the econometric model. This entails a deconvolution of the disturbance into a firm inefficiency component— quantification of which is the goal of the analysis—and a statistical noise term. Following this general outline, we discuss approaches to dealing with some key specification issues. Section 3 considers alternative distributional assumptions for inefficiency. Section 4 examines panel data issues.

A. D. Stead · P. Wheat
University of Leeds, Leeds, UK

W. H. Greene (✉)
Stern School of Business, New York University, New York, NY, USA
e-mail: wgreene@stern.nyu.edu

Section 5 considers modelling heteroskedasticity in error terms and its usefulness for policy analysis. Section 6 considers alternative noise distributions within SF models. Section 7 considers amendments to the standard SF model when the data contains efficient firms. Section 8 considers other received proposals relevant to appropriate distributional assumptions in SFA. Section 9 concludes.

2   Departure Points

The standard theory of the firm holds that firms seek to maximise profit. Under certain assumptions, a *profit function* exists that reflects the maximum profit attainable by the firm. The profit function is derived from the firm's *cost function*, which represents the minimum cost given outputs and input prices, and its *production function*, which describes the firm's technology. These are 'frontier' functions in the sense that they represent optimal outcomes that firms cannot improve upon given their existing technology. The duality of the production and cost functions was demonstrated by Shephard (1953). Debreu (1951) introduced the notion of a distance function to describe a multiple output technology and proposed that the radial distance of a producer's outputs from the distance function be used as a measure of technical inefficiency. Koopmans (1951) provided a definition of technical efficiency.

The idea that firms might depart from profit maximisation was first suggested in passing by Hicks (1935), who speculated that firms with market power in particular may choose to enjoy some of their rents not as profit, but as reduced effort to maximise profits, or 'a quiet life'. Later, Leibenstein (1966, 1975) discussed various empirical indications of firm-level 'X-inefficiency' and how it might arise. The debate between Leibenstein (1978) and Stigler (1976) highlighted two alternative characterisations of inefficiency: as a result of selective rationality and non-maximising behaviour, resulting in non-allocative welfare loss, or as the redistribution of rents within the firm, and therefore consistent with the idea of maximising outward behaviour. The latter characterisation essentially posits that firms are maximising an objective function including factors other than profit, and encompasses a wide range of specific hypotheses about firm behaviour. The revenue maximisation hypothesis of Baumol (1967), the balanced growth maximisation hypothesis of Marris (1964) and the expense preference hypothesis of Williamson (1963) are examples of hypotheses within which the firm (or its managers, given informational asymmetry between principal

and agent) pursues other objectives jointly with profit or subject to a profit constraint. We should therefore bear in mind that when we discuss efficiency, it is relative to an objective that we define, and not necessarily that of the firm (or its agents).

The early literature on econometric estimation of cost functions has focal points at Johnston (1960) for the UK coal industry and Nerlove (1963) for US electricity generation. These authors focused primarily on estimation of the shape of the empirical cost or production functions. Typically, ordinary least squares (OLS) was used to estimate a linear model:

$$y_i = x_i\beta + \varepsilon_i, \tag{1}$$

where y_i is cost or output, β is a vector of parameters to be estimated, ε_i is a random error term, $i = 1, 2, \ldots, I$ denotes an observed sample of data and x_i is the vector of independent variables. In the case of a production function, independent variables include input quantities and other factors affecting production, while in the case of a cost frontier, independent variables include output quantities and input prices, along with other factors affecting cost (Shephard 1953). Commonly, the dependent and independent variables are logged, in order to linearise what is assumed to be a multiplicative functional form. Note that the estimation of (1) via least squares, where a symmetric error term is assumed, is only consistent with the idea of a frontier function if we assume that firms are all fully efficient, and that departures from the estimated frontier are explained purely by measurement error and other random factors, such as luck. This fact has motivated many alternative proposals that are consistent with the notion of a frontier. Farrell (1957) proposed the use of linear programming to construct, assuming constant returns to scale, a piecewise linear isoquant and to define technical inefficiency as the radial distance of the firm from this isoquant.

An approach that amends (1) so the error is one-sided, yields a deterministic or 'full' frontier specification, in which the residuals are attributed entirely to inefficiency. Since a firm must be operating on or below its production frontier, and on or above its cost frontier, this means that $s\varepsilon_i \leq 0$, where $s = 1$ for a production frontier and $s = -1$ for a cost frontier. Aigner and Chu (1968) suggested linear or quadratic programming approaches to deterministic frontier estimation. Respectively, these minimise $\sum_{i=1}^{I} \varepsilon_i$ or $\sum_{i=1}^{I} \varepsilon_i^2$, subject to the constraint that $s\varepsilon_i \leq 0$. Schmidt (1976) noted that these are maximum likelihood (ML) estimators under the assumptions that the error term is exponentially or half normally distributed. Omitting the restriction that the residuals be one-sided leads to OLS and least absolute deviations (LAD) estimation, which would be ML estimation under the

assumptions that ε_i follows the normal or Laplace distributions, two-sided counterparts of the half-normal and exponential distributions, respectively. Afriat (1972) proposed a deterministic frontier model in which $\exp(\varepsilon_i)$ follows a two-parameter beta distribution, to be estimated via ML, which as Richmond (1974) noted is equivalent to assuming a gamma distribution for ε_i. The usual regularity conditions for ML estimation do not hold for deterministic frontier functions, since the range of variation of the dependent variable depends upon the parameters. Greene (1980) points out that under certain specific assumptions, this irregularity is actually not the relevant constraint. Specifically, if both the density and first derivative of the density of ε converge to zero at the origin, then the log-likelihood function is regular for ML estimation purposes. Deterministic frontier models with gamma and lognormal error term are examples.

Deterministic frontier models suffer from a serious conceptual weakness. They do not account for noise caused by random factors such as measurement error or luck. A firm whose production is impacted by a natural disaster might by construction appear to be inefficient. In order to account for measurement error, Timmer (1971) suggested amending the method so that the constraint $s\varepsilon_i \leq 0$ holds only with a given probability, thereby allowing a proportion of firms to lie above (below) the production (cost) frontier. However, this probability must be specified in advance in an arbitrary fashion. An alternative proposal made by Aigner et al. (1976) has the error drawn from a normal distribution with variance $\sigma^2\theta$ when $s\varepsilon_i \leq 0$ and $\sigma^2(1-\theta)$ when $s\varepsilon_i > 0$, where $0 < \theta < 1$. Essentially, though this is not made explicit, this allows for normally distributed noise with variance $\sigma^2(1-\theta)$ and inefficiency implicitly following a half-normal distribution with variance $(1 - 2/\pi)\sigma^2(1-\theta)$, under the assumption that where $s\varepsilon_i \leq 0$ firms are fully efficient. The resulting likelihood function is that of a 50:50 mixture of two differently scaled normal distributions truncated at zero from the left and right, respectively. The discontinuity of this specification once again violates the standard regularity conditions for ML estimation.

The issues with the models suggested by Timmer (1971) and Aigner et al. (1976) stem in both cases from their peculiar assumption that firms must be fully efficient when $s\varepsilon_i \leq 0$, which remains rooted in an essentially deterministic view of frontier estimation. The current literature on SFA, which overcomes these issues, begins with Aigner et al. (1977) and Meeusen and van Den Broeck (1977). They proposed a composed error:

$$\varepsilon_i = v_i - su_i \tag{2}$$

where v_i is a normally distributed noise term with zero mean, capturing random factors such as measurement error and luck, and u_i is a non-negative random variable capturing inefficiency and is drawn from a one-sided distribution. Battese and Corra (1977) proposed an alternative parameterisation of the model. Given particular distributional assumptions about the two error components, the marginal distribution of the composed error ε_i may be derived by marginalising u_i out of the joint probability;

$$f_\varepsilon(\varepsilon_i) = \int_0^\infty f_v(\varepsilon_i + su_i)f_u(u_i)\mathrm{d}u_i \tag{3}$$

where f_ε, f_v, and f_u are the density functions for ε_i, v_i, and u_i, respectively. The half-normal and exponential distributions were originally proposed for u_i. Assuming a normal distribution for v_i, the resulting distributions for ε_i are the skew-normal distribution, studied by Weinstein (1964) and Azzalini (1985), and the exponentially modified Gaussian distribution originally derived by Grushka (1972).

The ultimate objective of SFA is deconvolution of estimated residuals into separate predictions for the noise and inefficiency components. The latter is the focus of efficiency analysis. Since the parameters of f_u are outputs of the estimation process, we obtain an estimated distribution of efficiency, and as proposed by Lee and Tyler (1978), the first moment of this estimated distribution may be used to predict overall average efficiency. However, decomposing estimated residuals into observation-specific noise and efficiency estimates was elusive until Jondrow et al. (1982) suggested predicting based on the conditional distribution of $u_i|\varepsilon_i$, which is given by

$$f_{u|\varepsilon}(u_i|\varepsilon_i) = \frac{f_v(\varepsilon_i + su_i)f_u(u_i)}{f_\varepsilon(\varepsilon_i)}. \tag{4}$$

They derived (4) for the normal-half normal and normal-exponential cases. The mean, $E(u_i|\varepsilon_i)$, and mode, $M(u_i|\varepsilon_i)$, of this distribution were proposed as predictors. Waldman (1984) examined the performance of these and other computable predictors. Battese and Coelli (1988) suggest the use of $E\left[\exp\left(-u_i\right)|\varepsilon_i\right]$ when the frontier is log-linear. Kumbhakar and Lovell (2000) suggest that this is more accurate than $\exp\left[-E(u_i|\varepsilon_i)\right]$, especially when u_i is large. In practice, the difference often tends to be very small. It should be noted that the distribution of the efficiency predictions, $E(u_i|\varepsilon_i)$ will not match the unconditional, marginal distribution of

the true, unobserved u_i. Wang and Schmidt (2009) derived the distribution of $E(u_i|\varepsilon_i)$ and show that it is a shrinkage of u_i towards $E(u_i)$, with $E(u_i|\varepsilon_i) - u_i$ approaching zero as $\sigma_v^2 \to 0$.

3   Alternative Inefficiency Distributions

The efficiency predictions of the stochastic frontier model are sensitive to the assumed distribution of u_i. A number of alternatives have been proposed. Several two-parameter generalisations of the half-normal and exponential distributions, respectively, allow for greater flexibility in the shape of the inefficiency distribution, with non-zero modes in particular. The flexible forms generally enable testing against their simpler nested distributions. Stevenson (1980) proposed the truncated normal model[1]; Greene (1990) and Stevenson (1980) proposed gamma distributions. The truncated normal distribution, denoted $N^+\left(\mu, \sigma_u^2\right)$, nests the half normal when its location parameter μ (the pre-truncation mean) is zero, and its mode is μ when $\mu \geq 0$. The similar 'folded normal distribution' denoted $\left|N\left(\mu, \sigma_u^2\right)\right|$, i.e. that of the absolute value of an $N\left(\mu, \sigma_u^2\right)$ normal random variable, also nests the half normal when μ is zero, but has a non-zero mode only when $\mu \geq \sigma_u$ (Tsagris et al. 2014; Hajargasht 2014).

The gamma distribution with shape parameter k and scale parameter σ_u nests the exponential distribution when $k = 1$. A two-parameter lognormal distribution, which resembles the gamma distribution, for u_i is adopted by Migon and Medici (2001). It is possible to adopt even more flexible distributional assumptions; Lee (1983) proposed using a very general four-parameter Pearson distribution for u_i as a means of nesting several simpler distributions. On the other hand, Hajargasht (2015) proposed a one-parameter Rayleigh distribution for u_i which has the attraction of being a parsimonious way of allowing for a non-zero mode. Griffin and Steel (2008) proposed a three-parameter extension of Greene's two-parameter gamma model that nests the gamma, exponential, half-normal and (heretofore never considered) Weibull models. Some of these represent minor extensions of the base case models. In all cases, however, the motivation is a more flexible, perhaps less restrictive characterisation of the variation of efficiency across

---

[1]In the SF literature, 'truncated normal' refers specifically to the left truncation at zero of a normal distribution with mean μ and variance σ_u^2.

firms. In many cases, the more general formulations nest more restrictive, but common distributional forms.

The inefficiency distributions discussed above were proposed to enable more flexible distributional assumptions about u_i. Other proposals have addressed specific practical and theoretical issues. One is the 'wrong skew' problem, which is discussed in more detail below. Broadly, the skewness of su_i should be negative, both in the theory and as estimated using data. In estimation, it often happens that the information extracted from the data suggests skewness in the wrong direction. This would seem to conflict with the central assumption of the stochastic frontier model. The problem for the theoretical specification is that, since $\text{Skew}(\varepsilon_i) = \text{Skew}(v_i) - s\text{Skew}(u_i) = -s\text{Skew}(u_i)$ when v_i is symmetrically distributed, the skewness of the composed error ε_i is determined by that of u_i. Therefore, imposing $\text{Skew}(u_i) > 0$ implies that $-s\text{Skew}(\varepsilon_i) > 0$. Since all of the aforementioned distributions for u_i allow only for positive skewness, this means that the resulting SF models cannot handle skewness in the 'wrong' direction. An estimated model based on sample data will typically give an estimate of zero for $\text{Var}(u_i)$ if the estimated skewness (however obtained) goes in the wrong direction.

'Wrong skew' could be viewed as a finite sample issue, as demonstrated by Simar and Wilson (2010). Even when the assumed distribution of ε_i is correct, samples drawn from this distribution can have skewness in the 'wrong' direction with some probability that decreases with the sample size. Alternatively, it may indeed be the case that, though non-negative, the distribution of u_i has a zero or negative skew, and therefore, our distributional assumptions need to be changed accordingly. To this end, Li (1996) and Lee and Lee (2014)[2] consider a uniform distribution, $u_i \sim U(a,b)$, so that u_i and ε_i are both symmetric, and Carree (2002) and Tsionas (2007) consider the binomial distribution and Weibull distributions, respectively, which both allow for skewness in either direction. Arguably, these 'solutions' are ad hoc remedies to what might be a fundamental conflict between the data and the theory. Notwithstanding the availability of these remedies, negative skewness, defined appropriately is a central feature of the model.

Also relevant here are SF models with 'bounded inefficiency'. These are motivated by the idea that there is an upper bound on inefficiency beyond

---

[2]Lee and Lee (2014) focus on the upper bound on inefficiency in the normal-uniform model and appear to have been unaware of the model's earlier introduction by Li (1996), who was motivated by the skewness issue.

which firms cannot survive. Such a boundary could be due to competitive pressure, as suggested by Qian and Sickles (2008). However, we also consider that it could arise in monopolistic infrastructure industries which are subject to economic regulation, since depending on the strength of the regulatory regime, some inefficiency is likely to be tolerated.[3]

Implementation of bounded inefficiency involves the *right-truncation* of one of the canonical inefficiency distributions found in the SF literature. The upper tail truncation point is a parameter that would be freely estimated and is interpreted as the inefficiency bound. Lee (1996) proposed a tail-truncated half-normal distribution for inefficiency, and Qian and Sickles (2008) and Almanidis and Sickles (2012) propose a more general 'doubly truncated normal' distribution (i.e. the tail truncation of a truncated normal distribution). Almanidis et al. (2014) discuss the tail-truncated half-normal, tail-truncated exponential and doubly truncated normal inefficiency distributions. The latter of these may have positive or negative skewness depending on its parameter values. In fact, it is clear that this may be true of the right-truncation of many other non-negative distributions with non-zero mode.

A difficulty with certain distributional assumptions is that the integral in (3) may not have a closed-form solution, so that there may not be an analytical expression for the log-likelihood function. This issue first arose in the SF literature in the case of the normal-gamma model, in which case the problem was addressed in several different ways. Stevenson (1980) noted that relatively straightforward closed-form expressions exist for integer values of the shape parameter k, of the normal-gamma model and derived the marginal density of ε_i for $k = 0$, $k = 1$, and $k = 2$. Restricting k to integer values gives the Erlang distribution, so this proposal amounts to a restrictive normal-Erlang model. The need to derive distinct formulae for every possible integer value of k makes this approach unattractive. Beckers and Hammond (1987) derived a complete log-likelihood for the normal-gamma model, but due to its complexity their approach has not been implemented. Greene (1990) approximated the integral using quadrature, but this approximation proved rather crude (Ritter and Simar 1997). An alternative approach, proposed by Greene (2003), is to approximate the integral via simulation

---

[3]Such 'tolerance' does not necessarily reflect the technical competence or experience of regulators per se. It could reflect the perceived limitations on the robustness of the analysis (e.g. data quality), which necessitates a risk averse efficiency finding from a regulatory review.

in order to arrive at a maximum simulated likelihood (MSL) solution. For more detail on MSL estimation, see Train (2009). In the context of SFA, Greene and Misra (2003) note that the simulation approach could be used to approximate the integral in (3) for many distributional assumptions as long as the marginal variable u_i can be simulated. Since the integral is the expectation of $f_v(\varepsilon_i + su_i)$ given the assumed distribution for u_i, it can be approximated by averaging over Q draws from the distribution of u_i:

$$f_\varepsilon(\varepsilon_i) = \int\limits_0^\infty f_v(\varepsilon_i + su_i)f_u(u_i)\mathrm{d}u_i \approx \frac{1}{Q}\sum_{q=1}^Q f_v\left[\varepsilon_i + sF_u^{-1}(d_q)\right] \quad (5)$$

where d_q is draw number q from the standard uniform distribution, transformed by the quantile function F_u^{-1} into a draw from the distribution of u_i. In cases in which there is no analytical F_u^{-1}, such as the normal-gamma model, the integral may nevertheless be expressed in terms of an expectation that may be approximated via simulation. Greene (2003) recommends using Halton sequences, which aim for good coverage of the unit interval, rather than random draws from the uniform distribution, in order to reduce the number of draws needed for a reasonable approximation of the integral.

As an alternative to simulation, various numerical quadrature approaches may be used. Numerical quadrature involves approximating an integral by a weighted sum of values of the integrand at various points. In many cases, this involves partitioning the integration interval and approximating the area under the curve within each of the resulting subintervals using some interpolating function. The advantage of quadrature over simulation lies in speed of computation, given that the latter's time-consuming need to obtain potentially large numbers of independent draws for each observation. However, it may be challenging to find appropriate quadrature rules in many cases. Another alternative, proposed by Tsionas (2012), is to approximate f_ε using the (inverse) fast Fourier transform of the characteristic function of f_ε. The characteristic function, φ_ε, is the Fourier transform of f_ε, and as shown by Lévy's inversion theorem (see Theorem 1.5.4 in Lukacs and Laha 1964), the inverse Fourier transform of the characteristic function can be used to obtain f_ε. Since the Fourier transform of a convolution of two functions is simply the product of their Fourier transforms, i.e. $\varphi_\varepsilon = \varphi_v\varphi_u$ (see Bracewell 1978, p. 110), φ_ε may be relatively simple even when f_ε has no closed form, and f_ε may be approximated by the inverse fast Fourier transform of φ_ε. On the basis of Monte Carlo experiments, Tsionas (2012) finds that this is a faster method for approximating f_ε in the normal-gamma

and normal-beta cases than either Gaussian quadrature or Monte Carlo simulation, with the former requiring a large number of quadrature points and the latter an even larger number of draws for comparable accuracy. This approach has not yet been adopted as widely as simulation, perhaps due to its relative complexity.

A natural question would be which, of the many alternatives discussed above, is the most appropriate distribution for inefficiency? Unfortunately, theory provides little guidance on this question. Oikawa (2016) argues that a simple Bayesian learning-by-doing model such as that of Jovanovic and Nyarko (1996), in which a firm (or manager) maximises technical efficiency given prior beliefs about and previous realisations of an unknown technology parameter, supports a gamma distribution for inefficiency. However, Tsionas (2017) shows that this conclusion is sensitive to the sampling of, and assumed prior for, the firm-specific parameter, and that under alternative formulations there is no basis for favouring the gamma distribution (or any known distribution). Furthermore, both authors assume that firms maximise expected profits, whereas alternative behavioural assumptions may yield very different results. Of course, the choice of inefficiency distribution may be driven by practical considerations, such as a need to allow for wrong skewness or to estimate an upper bound on inefficiency. The question of which inefficiency distribution to use is an empirical one and leads us to consider testing in the context of SFA. As noted previously, some of the more flexible inefficiency distributions nest simpler distributions. In these cases, we may test against to simpler nested models. For example, we may test down from the normal-gamma to the normal-exponential model by testing the null hypothesis that $k = 1$. We may test down from the normal-truncated normal (or the normal-folded normal) to the normal-half normal model by testing the null hypothesis that $\mu = 0$. These are standard problems.

There are some remaining complications in the specification search for the SF model. We may wish to test for the presence of the one-sided error, often interpreted as a test for the presence of inefficiency. In this case, the errors are normally distributed under the null hypothesis $H_0 : \sigma_u = 0$. This is a non-standard problem because the scale parameter σ_u is at a boundary of the parameter space under H_0. Case 5 in Self and Liang (1987) shows that where a single parameter of interest lies on the boundary of the parameter space under the null hypothesis, the likelihood ratio (LR) statistic follows a 50:50 mixture of χ_0^2, and χ_1^2 distributions, denoted $\chi_{1:0}^2$, for which the 95% value is 2.706 (Critical values are presented in Kodde and Palm 1986). Lee (1993) finds that this is the case under $H_0 : \sigma_u = 0$ in the normal-half

normal model. A Lagrange multiplier test for this case in the SF model is developed in Lee and Chesher (1986).

This result does not apply when f_u has two or more parameters. Coelli (1995) states that, in the normal-truncated normal model, the LR statistic under $H_0 : \sigma_u = \mu = 0$ follows a 25:50:25 mixture of χ_0^2, χ_1^2 and χ_2^2 distributions, and that this is a special case of the result for two restrictions in Gouriéroux et al. (1982), which deals with inequality restrictions.[4] This result matches Case 7 in Self and Liang (1987), in which two parameters of interest lie on the boundary of the parameter space under the null. The test seems to have been incorrectly applied; under $H_0 : \sigma_u = \mu = 0$, only one parameter lies on the boundary. Equivalently, viewing the test as a one-tailed test of $H_0 : \sigma_u \leq 0, \mu = 0$, we only have one inequality restriction. Case 6 in Self and Liang (1987), in which there are two parameters of interest, one on a boundary, and one not on a boundary, seems to be more applicable, suggesting a 50:50 mixture of χ_1^2 and χ_2^2 distributions, denoted $\chi_{2:1}^2$. More fundamentally, $H_0 : \sigma_u = \mu = 0$ may not be the appropriate null hypothesis: when the scale parameter of the inefficiency distribution is set to zero, all other parameters of the distribution are in fact unidentified. Equivalently, a normal distribution for ε_i can be recovered in the normal-truncated normal case as $\mu \to -\infty$, for any value of σ_u. The general problem of testing when there are unidentified nuisance parameters under the null hypothesis is discussed by Andrews (1993a, b) and Hansen (1996). To our knowledge has not been addressed in the SF literature.

We may wish to choose between two non-nested distributions. In this case, Wang et al. (2011) suggest testing goodness of fit by comparing the distribution of the estimated residuals to the theoretical distribution of the compound error term. This is a simpler method than, for example, comparing the distribution of the efficiency predictions to the theoretical distribution of $E(u|\varepsilon)$ as derived by Wang and Schmidt (2009), since the distribution of the compound error is much simpler. For example, as discussed previously, ε_i follows a skew-normal distribution in the normal-half normal model, and an exponentially modified Gaussian distribution in the normal-exponential model. Under alternative specifications, the distribution of the estimated residuals may become rather complex, however.

---

[4]If we view the normal-half normal model as a skew-normal regression model in which we expect (but do not restrict) the skewness parameter σ_u/σ_v to be positive, then we view the test for the presence of inefficiency as a one-tailed test of the H_0 that $\sigma_u \leq 0$, or equivalently that $\sigma_u/\sigma_v = 0$, rather than as a test involving a boundary issue. Comparing the case of one inequality constraint in Gouriéroux et al. (1982) to Case 5 in Self and Liang (1987), we see the same result.

4   Panel Data

The basic panel data SF model in the contemporary literature is as in (1) with the addition of a t subscript to denote the added time dimension of the data:

$$y_{it} = x_{it}\beta + \varepsilon_{it}, \tag{6}$$

where $t = 1, 2, \ldots, T$. The composite error term is now

$$\varepsilon_{it} = \alpha_i + v_{it} - su_{it}. \tag{7}$$

Along with the usual advantages of panel data, Schmidt and Sickles (1984) identify three benefits specific to the context of SFA. First, under the assumption that inefficiency is either time invariant or that it varies in a deterministic way, efficiency prediction is consistent as $T \rightarrow \infty$. In contrast, this is not the case as $N \rightarrow \infty$. Second, distributional assumptions can be rendered less important, or avoided altogether, in certain panel data specifications. In particular skewness in the residual distribution does not have to be the only defining factor of inefficiency. Instead, time persistence in inefficiency can be exploited to identify it from random noise. Third, it becomes possible, using a fixed-effects approach, to allow for correlation between inefficiency and the variables in the frontier.[5] In addition, the use of panel data allows for the modelling of dynamic effects.

In the context of panel data SF modelling, one of the main issues is the assumption made about the variation (or lack thereof) of inefficiency over time. Another is the way in which we control (or do not control) for firm-specific unobserved heterogeneity and distinguishes this from inefficiency. For the purposes of this discussion, we divide the received panel data SF models into three classes: models in which inefficiency is assumed to be time-invariant, models in which inefficiency is time-varying, and models which control for unobserved heterogeneity with either time-invariant or time-varying inefficiency. To finish this section, we consider briefly multi-level panel datasets and the opportunities that they provide for analysis.

---

[5]However, since Schmidt and Sickles (1984), cross-sectional models have been proposed, such as those of Kumbhakar et al. (1991), Huang and Liu (1994), and Battese and Coelli (1995), that allow for dependence between inefficiency and frontier variables. These are discussed in Sect. 4.

4.1   Time-invariant Efficiency

One approach to panel data SFA is to assume that efficiency varies between firms but does not change over time, as first proposed by Pitt and Lee (1981). Referring to (6) and (7), the basic panel data SF model with time-invariant efficiency assumes that $\alpha_i = 0, u_{it} = u_i$, so that we have:

$$y_{it} = x_{it}\beta + v_{it} - su_i. \tag{8}$$

This specification has the advantage that prediction (or estimation) of u_i is consistent as $T \to \infty$. The appeal of this result is diminished given that the assumption of time-invariance is increasingly hard to justify as the length of the panel increases. In contrast to the cross-sectional case, there is no need to assume that u_i is a random variable with a particular distribution, and therefore, there are several different methods may be used to estimate (8), depending on our assumptions about u_i.

Schmidt and Sickles (1984) proposed four alternative approaches. First, we may assume that u_i is a firm-specific fixed effect, and to estimate the model using either a least squares dummy variable (LSDV) approach, in which u_i is obtained as the estimated parameter on the dummy variable for firm i, or equivalently by applying the within transformation, in which case u_i is obtained as firm i's mean residual. Second, we may assume that u_i is a firm-specific random effect and estimate the model using feasible generalised least squares (FGLS). The difference between the fixed-effects and random-effects approaches is that the latter assumes that the firm-specific effects are uncorrelated with the regressors, while the former does not. Third, Schmidt and Sickles (1984) suggested instrumental variable (IV) estimation of the error components model proposed by Hausman and Taylor (1981) and Amemiya and MaCurdy (1986), which allows for the firm-specific effect to be correlated with some of the regressors and uncorrelated with others, and is thus intermediate between the fixed-effects and random-effects models. Fourth, as in Pitt and Lee (1981), u_i could be regarded as an independent random variable with a given distribution, as in the cross-sectional setting, with the model being estimated via ML.

The first three approaches share the advantage that no specific distributional assumption about u_i is required. As a consequence, the estimated firm-specific effects could be positive. As a result, firm-specific efficiency can only be measured relative to the best in the sample, not to an absolute benchmark. The estimated u_i is given by

$$u_i = \max_j sa_j - sa_i, \tag{9}$$

where a_i is the estimated firm-specific effect for firm i. The fixed-effects specification has the advantage of allowing for correlation between u_i and x_{it}. But the drawback is that time-invariant regressors cannot be included, meaning that efficiency estimates will be contaminated by any differences due to time-invariant variables. The assumption that the factors are uncorrelated with errors (noise or inefficiency) can be examined using the Hausman test (Hausman 1978; Hausman and Taylor 1981). If this assumption appears to hold, a random effects approach such as Pitt and Lee (1981) may be preferred. Another approach is to estimate a correlated random-effects model using Chamberlain-Mundlak variables—see Mundlak (1978) and Chamberlain (1984)—to allow for correlation between the random effects and the regressors. Griffiths and Hajargasht (2016) propose correlated random effects SF models using Chamberlain-Mundlak variables to allow for correlation between regressors and error components, including inefficiency terms.

The ML approach to estimation of (8) was first suggested by Pitt and Lee (1981), who derived an SF model for balanced panel data with a half-normal distribution for u_i and a normal distribution for v_{it}. This model therefore nests the basic cross-sectional model of Aigner et al. (1977) when $T = 1$. As in the cross-sectional setting, alternative distributional assumptions may be made. Battese and Coelli (1988) generalise the Pitt and Lee (1981) model in two ways: first, by allowing for an unbalanced panel and second, by assuming a truncated normal distribution for u_i. Normal-exponential, normal-gamma and normal-Rayleigh variants of the Pitt and Lee (1981) model are implemented in LIMDEP Version 11 (Greene 2016). As in the cross-sectional setting, parameter estimates and efficiency predictions obtained under the ML approach are more efficient than those from semi-parametric models if the distributional assumptions made are valid. If those assumptions are not valid, they may be inconsistent and biased. To be sure, the ability to test distributional assumptions is very limited.

4.2   Time-Varying Efficiency

Allowing for variation in efficiency over time is attractive for a number of reasons. As already noted, the assumption that efficiency is time-invariant is increasingly hard to justify as T increases. We would expect average efficiency to change over time. There may also be changes in the relative positions of firms, in terms of convergence or divergence in efficiency between firms, and potentially also changes in rankings through firms overtaking

each other. A wide variety of time-varying efficiency SF specifications have been proposed, each differing with respect to their flexibility in modelling the time path of efficiency and each having their own advantages and disadvantages.

As Amsler et al. (2014) note, panel data SF specifications can be grouped into four categories with respect to how u_{it} changes over time. One of these, covered in the preceding section, is models with time-invariant efficiency, so that $u_{it} = u_i$. Second, we could assume independence of u_{it} over t. In this case, we may simply estimate a pooled cross-sectional SF model, the possibility of unobserved heterogeneity notwithstanding. The advantages of this approach are the flexibility of u_{it}—and by extension, that of $E(u_{it}|\varepsilon_{it})$—over time, its simplicity and its sparsity, given that it adds no additional parameters to the model. However, the assumption of independence over time is clearly inappropriate.

Third, we may treat u_{it} as varying deterministically over time. One approach is to include time-varying fixed or random effects, a_{it}, with u_{it} being given by

$$u_{it} = \max_j sa_{jt} - sa_{it}. \tag{10}$$

Of course, given that $N \leq IT$ firm- and time-specific parameters[6] cannot be identified, some structure must be imposed. Kumbhakar (1991, 1993) proposed combining firm-specific (but time-invariant) and time-specific (but firm-invariant) effects, such that $a_{it} = \lambda_i + \sum_{t=2}^{T} \lambda_t$. This imposes a common trend in u_{it} among firms, albeit one that may be quite erratic. Lee and Schmidt (1993) proposed a specification, $a_{it} = \lambda_t \alpha_i$, which again imposes a trend over time. This is common for all firms, but complicates estimation due to its non-linearity. An alternative approach is to specify that $a_{it} = g(t)$ as proposed by Cornwell et al. (1990), who specifically suggested a quadratic time trend with firm-specific parameters, such that $a_{it} = \lambda_i + \lambda_{i1}t + \lambda_{i2}t^2$. This specification is flexible, in that it allows for firms to converge, diverge or change rankings in terms of efficiency. Ahn et al. (2007) propose a specification which nests both the Lee and Schmidt (1993) and Cornwell et al. (1990) models, in which $a_{it} = \sum_{j=1}^{p} \lambda_{jt}\alpha_{ji}$, thus allowing for arbitrary, firm-specific time trends. This specification nests the Lee and Schmidt (1993) model when $p = 1$, and the Cornwell et al. (1990) model when $p = 3$, $\lambda_{1t} = 1$, $\lambda_{2t} = t$, $\lambda_{3t} = t^2$. The value of p is estimated along with

---

[6]N being the total number of observations, so that $N = IT$ in the case of a balanced panel.

the model parameters. The authors discuss estimation and identification of model parameters. Ahn et al. (2013) discuss estimation of this model when there are observable variables correlated with the firm-specific effects, but not with v_{it}. An alternative approach based on factor modelling and allowing for arbitrary, smooth, firm-specific efficiency trends is proposed by Kneip et al. (2012).

Because semi-parametric models yield only relative estimates of efficiency, it is not possible to disentangle the effects of technical change (movement of the frontier) and efficiency change. An analogous approach in the context of parametric specifications is to use a 'scaling function', so that

$$u_{it} = g(t)u_i. \tag{11}$$

Here, u_i is a time-invariant random variable following a one-sided distribution—as in the time-invariant specification of Pitt and Lee (1981)—and $g(t)$ is a non-negative function of t. Kumbhakar (1990) proposed $g(t) = 1/\left[1 + \exp\left(\lambda_1 t + \lambda_2 t^2\right)\right]$; Battese and Coelli (1992) proposed $g(t) = \exp\left[\lambda_1(t - T)\right]$ and $g(t) = \exp\left[\lambda_1(t - T) + \lambda_2(t - T)^2\right]$. In each case, u_i is assumed to follow half-normal distribution. In these models, efficiency moves in the same direction for all firms, but there may be convergence of firms over time. In addition, with the exception of the one-parameter Battese and Coelli (1992) scaling function, these allow for non-monotonic trends in u_{it} over time. However, they do not allow for changes in rank over time, which requires firm-specific time trends.

Cuesta (2000) generalised the one-parameter Battese and Coelli (1992) scaling function to allow for firm-specific time trends, so that $g(t) = \exp\left[\lambda_{1i}(t - T)\right]$. An extension to the two-parameter case would be straightforward. This allows for firm-specific time trends, as in the Cornwell et al. (1990) model, but again at the cost of increasing the number of parameters in the model by a factor of I. However, Wheat and Smith (2012) show that the Cuesta (2000) specification, unlike that of Battese and Coelli (1992), can lead to a counterintuitive 'falling off' of firms with high $E(u_{it}|\varepsilon_{it})$ in the final year of the sample They propose a model in which $g(t) = \exp\left[\lambda_{1i}(t - \lambda_{2i})\right]$, that does not have the same feature.[7] More generally, as Wheat and Smith (2012) note, the many different models using that use functions are sensitive to the precise form of $g(t)$ in terms of parameter estimates, fit and efficiency predictions.

---

[7]Clearly, this model is far from parsimonious, since $g(t)$ includes $2I$ parameters. In fact, the authors apply a simpler model, $g(t) = \exp\left[\lambda_{1i}(t - \lambda_2)\right]$ after failing to reject $H_0 : \lambda_{2i} = \lambda_2$.

A fourth approach to time-variation of u_{it} in panel data SF models is to allow for correlation between u_{it} over time by assuming that $(u_{i1}, \ldots, u_{iT})$ are drawn from an appropriate multivariate distribution. Among their various proposals, Pitt and Lee (1981) suggested that $(u_{i1}, \ldots, u_{iT})$ could be drawn from a multivariate truncated normal distribution. They abandoned this approach, after noting that the likelihood function for this model involves intractable T-dimensional integrals.[8] In addition, Horrace (2005) showed that the marginal distribution of u_{it} in this case is not truncated normal. However, as suggested by Amsler et al. (2014), it is possible to specify a multivariate distribution with the desired marginal distributions, and also obviate T-dimensional integration when evaluating $\ln L$, by using a copula function. Sklar's theorem—see Nelsen (2006, pp. 17–14)—states that any multivariate cumulative density function can be expressed in terms of a set of marginal cumulative density functions and a copula. For example, we have

$$H_u(u_{i1}, \ldots, u_{iT}) = C[F_{u1}(u_{i1}), \ldots, F_{uT}(u_{iT})] \tag{12}$$

where H_u is a multivariate cumulative density function for $(u_{i1}, \ldots, u_{iT})$, $C[.]$ is the copula function, and $F_{u1}(u_{i1}), \ldots, F_{uT}(u_{iT})$ are the marginal cumulative density functions for u_{it} for each time period. We would normally assume that $F_{ut} = F_u$ for all t, so that we have

$$H_u(u_{i1}, \ldots, u_{iT}) = C[F_u(u_{i1}), \ldots, F_u(u_{iT})]. \tag{13}$$

From this, it can be seen that the probability density function is given by

$$h_u(u_{i1}, \ldots, u_{iT}) = \prod_{t=1}^{T} \left[f_u(u_{it}) \right] c[F_u(u_{i1}), \ldots, F_u(u_{iT})] \tag{14}$$

where c is the derivative of the copula. It follows from this that a multivariate density $h(u_{i1}, \ldots, u_{iT})$ with the desired marginal densities given by f_u can be obtained by combining f_u and F_u with an appropriate copula density c. Many different copula functions exist—it is beyond the scope of this chapter to review the various candidates—each embodying different dependence structures. Note that $c = 1$ relates to the special case of independence. This allows marginal distributions to be specified, but the

---

[8]The authors instead estimate a system of T equations via the seemingly unrelated regressions (SUR) model proposed by Zellner (1962). However, this approach offers no way of predicting observation-specific efficiencies.

problem of T-dimensional integration to evaluate the log-likelihood persists. For this reason, Amsler et al. (2014) propose and implement an alternative approach whereby instead of specifying a copula for $(u_{i1}, \ldots, u_{iT})$, a copula is specified for the composite errors $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$. In this case, we have

$$h_\varepsilon(\varepsilon_{i1}, \ldots, \varepsilon_{iT}) = \prod_{t=1}^{T} \left[f_\varepsilon(\varepsilon_{it}) \right] c[F_\varepsilon(\varepsilon_{i1}), \ldots, F_\varepsilon(\varepsilon_{iT})] \qquad (15)$$

where h_ε is the multivariate distribution for $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$ and F_ε is the marginal cumulative density function for ε_{it}. In this case, an appropriate marginal distribution for ε_{it} is chosen, such as the skew-normal distribution. In this case, the correlation is between the composite errors, introducing dependency between both error components. Amsler et al. (2014) take both approaches, estimating a model in which $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$ is drawn from a joint distribution as in (15) via ML, and a model in which $(u_{i1}, \ldots, u_{iT})$ is drawn from a joint distribution as in (14), while v_{it} is assumed independent, via MSL. A Gaussian copula function is used in both cases. The authors discuss prediction of efficiency. In this case, it is based on $u_{it}|\varepsilon_{i1}, \ldots, \varepsilon_{iT}$. This results in improved predictions relative to those based on $u_{it}|\varepsilon_{it}$, since the composite errors from all years are informative about u_{it} when there is dependency between them.

The copula approach proposed by Amsler et al. (2014) is attractive, since it can be seen as intermediate between the pooled SF approach and the approach of specifying SF models with deterministically time-varying u_{it}. As such, it retains the advantage of the latter approach in allowing for dependency over time, without specifying a particular functional form for the time trend. It also obviates the large number of additional parameters otherwise needed to allow flexibility with respect to time trends. Rather than some factor of I, the number of new parameters is limited to the correlation coefficients $\rho_{ts} \forall t \neq s$. A number of simplifying assumptions can be made to reduce the number of these while retaining flexibility. Firms may converge or diverge, or change rankings, using a relatively parsimonious specification under this approach.

4.3   Unobserved Heterogeneity

Aside from considerations of the appropriate way to model trends in u_{it} over time, which is peculiar to the panel data SF context, more general panel data issues are also relevant. Primary among these is the need to account

for possible unobserved heterogeneity between firms. In general, this means incorporating firm-specific effects which are time-invariant but not captured by the regressors included in the frontier. These may be either correlated or uncorrelated with the regressors, i.e. they may be fixed or random effects, respectively. In general, failure to account for fixed effects may bias parameter estimates, while failure to account for random effects generally will not.[9] In the SF context, failure to account for fixed or random effects means such effects may be attributed to u_{it}.

A number of models have been proposed which incorporate fixed or random effects. These are interpreted as capturing unobserved heterogeneity rather than as inefficiency effects. Kumbhakar (1991) proposed extending the pooled cross-section model to incorporate firm and time effects uncorrelated with the regressors, so that

$$\varepsilon_{it} = v_{it} + a_i + a_t - su_{it} \tag{16}$$

where a_i and a_t are firm- and time-specific fixed or random effects. In the fixed-effects case, Kumbhakar (1991) suggests estimation via ML with firm dummy variables, under the assumptions that a_i, a_t, and v_{it} are drawn from normal distributions with zero means and constant variances, and u_{it} is drawn from a truncated normal distribution. A simplified version of this model, omitting a_t and treating a_i as a fixed effect, was used by Heshmati and Kumbhakar (1994). This model was also considered by Greene (2004, 2005a, b), who proposed the specification

$$\varepsilon_{it} = v_{it} + a_i - su_{it} \tag{17}$$

where a_i is a time-invariant fixed or random effect, and the specification is referred to as the 'true fixed effects' (TFE) or 'true random effects' (TRE) model, accordingly. In the TFE case, estimation proceeds by simply replacing the constant term in the standard pooled with a full set of firm dummies and estimating the model via ML. However, evidence presented by Greene (2005b) from Monte Carlo experiments suggests that this approach suffers from the incidental parameters problem. As a result, Chen et al. (2014) propose an alternative ML approach based on the within transformation, which is not subject to this problem, and Belotti and Ilardi (2018) extend this approach to allow for heteroscedastic u_{it}.

---

[9]However, in the context of a log-linear model, the estimate of the intercept will be biased in either case.

In the TRE case, Greene (2004, 2005a, b) proposed estimation of the model via MSL, assuming that $a_i \sim N(0, \sigma_a^2)$. Greene (2005b) notes that the TRE approach—and indeed the standard SF model—can be seen as special cases of a random parameters model and proposes a random parameters specification incorporating heterogeneity in β, so that

$$y_{it} - x_i \beta_i = \varepsilon_{it} = v_{it} - s u_{it} \tag{18}$$

where β_i is assumed to follow a multivariate normal distribution with mean vector β and covariance matrix $\sum$. The random intercept is $\beta_{0i} = \beta_0 + a_i$ in terms of the TRE notation. The model is estimated via MSL. The resemblance of this approach to the Bayesian SF specifications considered by Tsionas (2002) is noted. However, the Bayesian approach has the drawback of requiring some prior distribution to be chosen for all parameters, including those of f_u. Greene (2008) notes that in the classical framework, 'randomness' of the parameters reflects technological heterogeneity between firms, whereas in the Bayesian framework, 'randomness' of the parameters is supposed to reflect the uncertainty of the analyst.[10]

A discrete approximation to the random parameters SF model is possible using a latent class approach to capture heterogeneity in some or all of the β parameters, as proposed by Orea and Kumbhakar (2004). In this specification, each firm belongs to one of J classes, each class having a distinct technology, so that for class j, we have technology parameters β_j. Class membership is unknown. Each firm is treated as belonging to class j with unconditional probability p_j, where the unconditional probabilities are estimated as parameters after normalising such that $\sum_{j=1}^{J} p_j = 1$ (leaving $J - 1$ additional parameters to be estimated). The model may be estimated via ML. Conditional probabilities of class membership for each observation obtained by

$$p_{ij} = \frac{p_j f_\varepsilon (y_{it} - x_i \beta_j)}{\sum_{j=1}^{J} \left[p_j f_\varepsilon (y_{it} - x_i \beta_j) \right]} \tag{19}$$

The primary issue with the TFE and TRE and similar models is that any time-invariant effects are attributed to a_i, when it is entirely possible that they should, partly or wholly, be attributed to u_{it}. Several recent proposals therefore extend this modelling approach to allow for u_{it} to be broken down

---

[10]Despite this, Tsionas (2002) does interpret the models as incorporating technological heterogeneity.

into separate time-invariant and time-varying components capturing 'persistent' and 'transient' inefficiency effects, respectively. Thus,

$$u_{it} = w_i + w_{it} \tag{20}$$

where typically both w_i and w_{it} are random variables drawn from some one-sided distribution. A similar decomposition of u_{it} was first suggested by Kumbhakar and Heshmati (1995), who proposed that $u_{it} = a_i + w_{it}$ and Kumbhakar and Hjalmarsson (1995), who proposed $u_{it} = a_i + \alpha_t + w_{it}$, where a_i and α_t are firm- and time-specific fixed or random effects, respectively.[11] Colombi et al. (2014) and Tsionas and Kumbhakar (2014) propose an extension of the TRE model, accordingly referred to as the generalised true random effects (GTRE) model, in which

$$\varepsilon_{it} = v_{it} + a_i - s(w_i + w_{it}) \tag{21}$$

This model therefore includes four error components, allowing for noise, unobserved heterogeneity, and persistent and transient inefficiency. Identification requires specific distributional assumptions to be made about either a_i or w_i, or both. The following distributional assumptions are typically made: $v_{it} \sim N\left(0, \sigma_v^2\right)$, $a_i \sim N\left(0, \sigma_a^2\right)$ and w_i and w_{it} are follow half-normal distributions with constant variances. Each of the error components is assumed to be independent. Various approaches to estimation of the GTRE model have been proposed. Kumbhakar et al. (2014) suggest a multi-step approach. In the first step, a standard random effects panel data model including a noise component $v_{it}^* = v_{it} + w_{it}$ and a time-invariant random effects component $a_i^* = a_i + w_i$. This can be estimated via FGLS, avoiding any explicit distributional assumptions. Subsequently, the estimates of these error components are used as the dependent variables in separate constant-only SF models, which decompose them into their two-sided and one-sided components. This is straightforward to implement using standard software packages.

Alternatively, Colombi et al. (2014) use the result that ε_{it} in the GTRE model is the sum of two random variables, each drawn from an independent closed skew-normal distribution.[12] As its name suggests, the closed

---

[11]Note that these proposals are very similar to those of Kumbhakar (1991) and Heshmati and Kumbhakar (1994), the difference being the interpretation of a_i and α_t as picking up inefficiency effects, rather than unobserved heterogeneity.

[12]The univariate skew normal distribution is a special case of the closed skew-normal distribution. To see that ε_{it} is the sum of two closed skew-normal random variables, therefore, consider that $v_{it}^* = v_{it} + w_{it}$. and $a_i^* = a_i + w_i$ both follow skew-normal distributions. For details on the closed skew-normal distribution, see González-Farías et al. (2004).

skew-normal distribution is closed under summation—see Proposition 2.5.1 of González-Farías et al. (2004b) or Theorem 1 in González-Farías et al. (2004a). Therefore, ε_{it} follows a skew-normal distribution. This enables estimation of the model via ML. However, Filippini and Greene (2016) note that this is extremely challenging, since the log-likelihood involves the probability density function for a T-variate normal distribution and the cumulative density function for a $T + 1$-variate normal distribution. They proposed a simpler approach based on MSL, which exploits the fact that the GTRE model is simply the TRE model in which the time-invariant error component follows a skew-normal distribution. Colombi et al. (2014) show how to obtain predictions for w_i and w_{it}.

The attraction of the GTRE model is that it is quite general, in that it allows for the decomposition of the composite error into noise, random effects, persistent inefficiency, and transient inefficiency components. It also nests various simpler models, such as the TRE model, the standard pooled SF model, the Pitt and Lee (1981) model, and a standard random-effects model. However, Badunenko and Kumbhakar (2016) recently concluded on the basis of Monte Carlo experiments that the model is very limited in its ability to precisely predict the individual error components in practice, and suggest that the model may not outperform simpler models in many cases.

4.4   Multi-level Panel Datasets

This section has outlined how panel data allows for a richer characterisation of efficiency and thus panel data is desirable for undertaking efficiency analysis. Both Smith and Wheat (2012) and Brorsen and Kim (2013) have considered using data on a number of organisations over time, but disaggregated on sub-firm divisions (henceforth: plants) for each organisation. Thus, there are two levels of data

$$y_{its} = x_{ijt}\beta + \tau_{its} + v_{its} \tag{22}$$

which is (6) but with the addition of a plant subscript j. There are two key advantages to considering data of this form. Firstly, such an approach allows for the measurement of internal efficiency variation within an organisation, as well as simultaneously measuring efficiency against comparator organisations (external efficiency). Smith and Wheat (2012) propose a model (ignoring the time dimension for simplicity) in which $u_{ij} = u_i + u_{ij}^*$, where u_i is a one-sided component common to all of firm i's plants, and u_{ij}^* is a plant-specific component assumed to follow a half-normal distribution. The

authors suggest estimating the model using a two-step approach, in which u_i is obtained from a fixed or random effect in the first step. Note that, in the one-period or pooled cross-section cases, this is simply the panel data specification of Kumbhakar and Hjalmarsson (1995) and Kumbhakar and Heshmati (1995).

Lai and Huang (2013) argue that there is likely to be intra-firm correlation between both plant-level efficiency and noise effects. Rather than allow for separate correlations between the v_{ij} and the u_{ij}, the authors propose a model in which the ε_{ij} are correlated such that $\rho(\varepsilon_{ij}, \varepsilon_{il}) = \rho$. The components and v_{ij} and u_{ij} are assumed to be drawn from marginal normal and half-normal distributions, respectively, the authors allow for correlation between the composed errors using a Gaussian copula.

Secondly, both Brorsen and Kim (2013) and Smith and Wheat (2012) demonstrate that there is a need to model costs at the level that management autonomy resides. Failure to do so can result in misleading predictions of efficiency as it mismatches returns to scale properties of the cost function with efficiency. Brorsen and Kim (2013) used data on schools and school districts to show that if the model were estimated using data at district level then returns to scale are found to be decreasing rather than finding that these schools are inefficient. Ultimately, the aggregation bias is resulting in correlation between errors and regressors, since true measures of scale/density (at the disaggregate level) are not included in the model.

5   Heteroscedasticity and Modelling Inefficiency

In many applications of SFA, the analyst is interested in not only in the estimation or prediction of efficiency, but also in its variation in terms of a set of observable variables. However, the standard SF model assumes that u_i is independent of observed variables. Many applications, including Pitt and Lee (1981) as an early example, take a two-step approach to modelling efficiency: first, a standard SF model is estimated and used to generate efficiency predictions, and second, these predictions are regressed on a vector of explanatory variables. However, the second-step regression violates the assumption of independence in the first step, and Wang and Schmidt (2002) show that the two-step approach is severely biased. Given that u_i is a random variable, appropriate approaches involve specifying one or more parameters of the error distributions as a function of a set of covariates.

Deprins and Simar (1989a, b), Reifschneider and Stevenson (1991), Kumbhakar et al. (1991), Huang and Liu (1994), and Battese and Coelli (1995) all propose extensions of the basic SF model whereby

$$u_i = g(z_i, \delta) + w_i \tag{23}$$

where z_i is a vector of 'environmental' variables influencing inefficiency, δ is a vector of coefficients, and w_i is a random error. In the Deprins and Simar (1989a, b) specification, $g(z_i, \delta) = \exp(z_i\delta)$ and $w_i = 0$, and the model may estimated via non-linear least squares or via ML assuming $v_i \sim N(0, \sigma_v^2)$.[13] Reifschneider and Stevenson (1991) propose restricting both components of u_i to be non-negative, i.e. $g(z_i, \delta), w_i \geq 0$, though as Kumbhakar and Lovell (2000) and Greene (2008) note, this is not required for $u_i \geq 0$. An alternative approach was proposed by Kumbhakar et al. (1991), in which $g(z_i, \delta) = 0$ and w_i is the truncation at zero of a normally distributed variable with mean $z_i\delta$ and variance σ_u^2. Huang and Liu (1994) proposed a model in which $g(z_i, \delta) = z_i\delta$ and w_i is the truncation at $-z_i\delta$ of an $N(0, \sigma_u^2)$ random variable. The latter two models are in fact equivalent, as noted by and Battese and Coelli (1995). In simple terms, the model assumes that $v_i \sim N(0, \sigma_v^2)$ and $u \sim N^+(\mu_i, \sigma_u^2)$, where $\mu_i = z_i\delta$. Note that a constant term is included in z_i, so that the model nests the normal-truncated normal model of Stevenson (1980) and the normal-half normal model.

Another set of models, motivated by the desire to allow for heteroskedasticity in u_i, specify the scale parameter, rather than the location parameter, of the distribution of u_i as a function of a set of covariates.[14] Reifschneider and Stevenson (1991) first proposed amending the normal-half normal model so that $\sigma_{ui} = h(z_i), h(z_i) \in (0, \infty)$, but did not make any particular suggestions about $h(z_i)$ other than noting that the function must be constrained to be non-negative. Caudill and Ford (1993) suggested the functional form $\sigma_{ui} = \sigma_u(z_i\gamma)^\alpha$, which nests the standard homoskedastic normal-half normal model when $\alpha = 0$. Caudill et al. (1995) suggested a slightly sparser specification in which $\sigma_{ui} = \sigma_u \exp(z_i\gamma)$, and Hadri (1999) proposed a similar 'doubly heteroskedastic' SF model, $\sigma_{vi} = \exp(z_i\theta), \sigma_{ui} = \exp(z_i\gamma)$.

---

[13]Note that the authors in fact proposed a deterministic frontier model in which $E(u_i|z_i) = \exp(z_i\delta)$, but if we interpret the random error as v_i rather than a component of u_i, we have an SF model with a deterministic u_i.

[14]Note, however, that since the (post-truncation) variance of the truncated normal distribution is a function of the pre-truncation mean, the Kumbhakar et al. (1991), Huang and Liu (1994), and Battese and Coelli (1995) model also implies heteroskedasticity in u_i.

The approaches discussed above can be combined for an encompassing model in which both the location and scale parameters are functions of z_i. Wang (2002) proposed a model in which $u_i \sim N^+\left(\mu_i, \sigma_{ui}^2\right)$, where $\mu_i = z_i\delta$ and $\sigma_{ui}^2 = \exp\left(z_i\gamma\right)$, while Kumbhakar and Sun (2013) took this a step further, estimating a model in which $u_i \sim N^+\left(\mu_i, \sigma_{ui}^2\right)$ and $v_i \sim N\left(0, \sigma_{vi}^2\right)$, where $\mu_i = z_i\delta$, $\sigma_{vi} = \exp\left(z_i\theta\right)$, and $\sigma_{ui} = \exp\left(z_i\gamma\right)$, effectively combining the Hadri (1999) 'doubly heteroskedastic' model with that of Kumbhakar et al. (1991), Huang and Liu (1994), and Battese and Coelli (1995).[15]

Given the motivation of explaining efficiency in terms of z_i, and since z_i enters the model in a non-linear way. It is desirable to calculate the marginal effect of these z_{li}, the lth environmental variable, on efficiency. Of course, given that u_i is a random variable, we can only predict the marginal effect of z_l on *predicted* efficiency, and this means that the marginal effects formula used depends fundamentally on the efficiency predictor adopted. Where $u_i \sim N^+\left(\mu_i, \sigma_{ui}^2\right)$, $\mu_i = z_i\delta$, the parameter δ_l is the marginal effect of z_{li} on the mode of the distribution of u_i, except when $z_i\delta \leq 0$. The derivative of the unconditional mode predictor,

$$\partial M(u_i)/\partial z_{li} = \begin{cases} \delta_l, z_i\delta > 0 \\ 0, z_i\delta \leq 0 \end{cases}. \tag{24}$$

Therefore, the unconditional mode yields a relatively simple marginal effect. Alternatively, Wang (2002) derived a marginal effects formula based on the derivative of the unconditional mean, $\partial E(u_i)/\partial z_{li}$. As the author shows, since $E(u_i)$ depends on the scale parameter, as well as the location parameter, of the distribution, marginal effects calculated using this formula can be non-monotonic even if z_{li} enters both functions in a linear fashion. This lends itself to potentially useful discussion of the 'optimal' (i.e. efficiency maximising) level of z_{li}. As noted by Hadri (1999), the variables entering μ_i, σ_{vi}, and σ_{ui} need not be the same in practice.

The efficiency prediction is usually based on the distribution of $u_i|\varepsilon_i$ (specifically its mean) rather than u_i. Kumbhakar and Sun (2013) argue that marginal effects should be based on $\partial E(u_i|\varepsilon_i)/\partial z_{li}$ rather than $\partial E(u_i)/\partial z_l$. and show that in this case, marginal effects depend upon the parameters not only of f_u but also of f_v and upon ε_i, i.e. all of the model's variables and parameters. Stead (2017) derives a marginal effects formula based on the conditional mode, $\partial M(u_i|\varepsilon_i)/\partial z_{li}$, which is somewhat

---

[15]Note the two similar but subtly different parameterisations, $\sigma_{ui} = \exp\left(z_i\gamma\right)$ and $\sigma_{ui}^2 = \exp\left(z_i\gamma\right)$.

simpler, particularly when both $\sigma_{vi} = \sigma_v$ and $\sigma_{ui} = \sigma_u$ in which case $\partial M(u_i|\varepsilon_i)/\partial z_{li} = \delta_l[\sigma_v^2/(\sigma_v^2 + \sigma_u^2)]$ when $M(u_i|\varepsilon_i) > 0$. Note that the marginal effects formulae discussed so far relate to changes in predicted u_i rather than predicted efficiency: Stead (2017) derives a marginal effect based on the Battese and Coelli (1988) predictor, $\partial E[\exp(-u_i)|\varepsilon_i]/\partial z_{li}$,, and notes that other formulae should be transformed into inefficiency space by multiplying by $-\exp(-\hat{u}_i)$ where $\hat{u}_i$ is the predictor for u_i since $\partial \exp(-\hat{u}_i)/\partial z_{li} = -(\partial \hat{u}_i/\partial z_{li})\exp(-\hat{u}_i)$. The choice between conditional and unconditional marginal effects formulae is between prediction of marginal effects for specific observations, and quantifying the relationship between environmental variables and inefficiency in general.

The idea that marginal effects should be based on a predictor of $u_i|\varepsilon_i$ rather than u_i has the appeal that the marginal effects discussed are consistent with the preferred efficiency predictor, in the sense that they indicate the change in predicted efficiency resulting from a change in z_{li}. On the other hand, such marginal effects are sensitive to changes in the frontier variables and parameters and the parameters of f_v, despite the fact that efficiency is not specified in this way. Another drawback is that while $\partial E(u_i)/\partial z_{li}$ and $\partial M(u_i)/\partial z_{li}$ are parameters for which standard errors and confidence intervals may be estimated, $\partial E(u_i|\varepsilon_i)/\partial z_{li}$ and $\partial M(u_i|\varepsilon_i)/\partial z_{li}$ are random variables for which prediction intervals are the only appropriate estimate of uncertainty, making hypothesis testing impossible. Kumbhakar and Sun (2013) suggest a bootstrapping approach to derive confidence intervals for $\partial E(u_i|\varepsilon_i)/\partial z_{li}$, but this is inappropriate since it treats ε_i as known.[16]

Given the rather complex marginal effects implied by the models discussed above, alternative specifications with simpler marginal effects have been proposed. Simar et al. (1994) propose that z_i should enter as a scaling function, such that $u_i = f(z_i\eta)u_i^*$, where u_i^* is assumed to follow some non-negative distribution that does not depend on z_i, and $f(z_i\eta)$ is a non-negative scaling function similar to those used in Battese and Coelli (1992) type panel data models. Wang and Schmidt (2002) note several features of this formulation: first, the shape of the distribution of u_i is the same for all observations, with $f(z_i\eta)$ simply scaling the distribution; models with this property are described as having the 'scaling property'. Second, it may yield relatively simple marginal effects expressions, e.g. when

---

[16]Note the similarity of the issues here to those around 'confidence intervals' and prediction intervals for $E(u_i|\varepsilon_i)$, discussed by Wheat et al. (2014).

$f(z_i\eta) = \exp(z_i\eta)$ or similar.[17] Third, as suggested by Simar et al. (1994), the β and η may be estimated via non-linear least squares without specifying a particular distribution for u_i^*. The scaling property is discussed further by Alvarez et al. (2006), who suggested testing for the scaling property.

More recently, Amsler et al. (2015) suggested an alternative parameterisation such that z_i enters the model through the post-truncation, rather than the pre-truncation, parameters of f_u. For example, the left truncation at zero of an $N(\mu_i, \sigma_{ui}^2)$ random variable, which we have denoted $N^+(\mu_i, \sigma_{ui}^2)$, may be reparameterised in terms of $E(u_i)$ and $\text{VAR}(u_i)$; that is, f_u may be expressed in terms of these parameters, and as a result, so may f_ε. The authors show that marginal effects are simpler and easier to interpret when environmental variables enter the model such that $E(u_i) = g(z_i, \delta), \text{VAR}(u_i) = h(z_i, \gamma)$ than when $\mu_i, \sigma_{ui}^2 = g(z_i, \delta), \sigma_{ui}^2 = h(z_i, \gamma)$. This is intuitive, given that we predict based on post-truncation parameters of f_u or $f_{u|\varepsilon}$. This approach is complicated somewhat by the requirement that $E(u_i) > \text{VAR}(u_i)$, as shown by Eq. (3) in Barrow and Cohen (1954), Eq. (16) in Bera and Sharma (1999), and Lemma 1 of Horrace (2015). For this reason, the authors suggest a specification in which $\text{VAR}(u_i) = \exp(z_i\gamma)$ and $E(u_i) = \text{VAR}(u_i) + \exp(z_i, \delta)$.

An additional motivation for the models discussed in this section is the analysis of production risk. Bera and Sharma (1999) proposed, in the context of a production frontier model, that $\text{VAR}(u_i|\varepsilon_i)$ be used as a measure of 'production uncertainty or risk. Note however that this is a far more restrictive measure than that used in the wider literature on production risk, which is variability of output, measured, for example, by $\text{VAR}(y_i)$. Nevertheless, these models offer considerable flexibility in modelling production risk according to this definition. Just and Pope (1978) showed that a drawback of log-linear (non-frontier) production function specifications, in which $q_i = \exp(y_i)$, is that the marginal production risk (i.e. the partial derivative of production risk) with respect to a given variable must always be the same as that variable's marginal product. The authors proposed an alternative specification with an additive error term multiplied by a scaling function. The form allows for variables that affect production and production risk in potentially opposite directions for variables that affect one but not the other. Kumbhakar (1993) and Battese et al. (1997) proposed SF variants

---

[17]However, the authors' discussion overstates the simplicity of marginal effects in this case, since it focuses on $\partial \ln \hat{u}_i / \partial z_{li}$, which is η_l regardless of the distribution of u_i^* (or indeed the choice of predictor). However, $\partial \hat{u}_i / \partial z_{li}$ is more complex, and as previously noted, the translation into efficiency space via $\partial \exp(-\hat{u}_i) / \partial z_{li}$ adds additional complexity.

of this model by including an inefficiency term u_i. Note, however, that any SF model in which one or both error terms are heteroskedastic allows for observation-specific production risk.

6   Alternative Noise Distributions

In the standard SF model, the noise term is assumed to follow a normal distribution. In contrast to the many different proposals concerning the distribution of u_i, discussed in Sect. 3, the distribution of v_i has received relatively little attention. This is perhaps natural, given that the main focus of SFA is on estimation or prediction of the former component. Nevertheless, consideration of alternative distributions for v_i is important for at least two main reasons. First, the standard model is not robust to outliers caused by noise, i.e. when the true noise distribution has thick tails. Second, and perhaps more importantly, the distribution of v_i has implications for the deconvolution of ε_i into noise and inefficiency components. Specifically, the distribution of $u_i|\varepsilon_i$, on which efficiency prediction is typically based, is influenced by f_v as well as f_u, as shown in (4).

The latter point in particular is not trivial. A change in distributional assumptions regarding v_i affects the degree of shrinkage of u_i towards $E(u_i)$ using $E(u_i)$.[18] A change in the assumed noise distribution can even be sufficient to change the rankings of firms[19] by altering the monotonicity properties of $E(u_i|\varepsilon_i)$ with respect to ε_i, which are in turn linked to the log-concavity properties of f_v. Ondrich and Ruggiero (2001) prove that $E(u_i|\varepsilon_i)$ is a weakly (strictly) monotonic function of ε_i for any weakly (strictly) log-concave f_v. Since the normal density is strictly log-concave everywhere, $E(u_i|\varepsilon_i)$ is a monotonic function of ε_i in the standard model. Under alternative noise distributions for which f_v is not strictly log-concave everywhere, there may be a weakly monotonic or even non-monotonic relationship between $E(u_i|\varepsilon_i)$ and ε_i. Such relationships have been noted in several studies proposing alternative, heavy tailed, noise distributions, which are discussed below.

Nguyen (2010) proposed SF models with Cauchy and Laplace distributions for v_i, pairing the former with half Cauchy and truncated Cauchy,

---

[18] For an explanation of shrinkage in the context of the predictor $E(u_i|\varepsilon_i)$, see Wang and Schmidt (2009).

[19] Holding β constant.

and the latter with exponential and truncated Laplace distributed for has received v_i terms.[20] Gupta and Nguyen (2010) derive a Cauchy-half Cauchy panel data model with time-invariant inefficiency. Horrace and Parmeter (2018) consider the Laplace-truncated Laplace and Laplace-exponential SF models further, showing that $f_{u|\varepsilon}$ (and therefore also $E(u_i|\varepsilon_i)$, or for that matter any predictor based on $f_{u|\varepsilon}$) is constant for $s\varepsilon_i \geq 0$. The authors conjecture that the assumption of a Laplace distributed v_i may be advantageous in terms of estimation of f_u, and therefore for the deconvolution of the composed error. Fan (1991) showed that optimal rates of convergence in deconvolution problems decrease with the smoothness of the noise distribution and are considerably faster for ordinary smooth distributions, such as the Laplace, than for super smooth distributions, such as the normal distribution. Optimal convergence rates for nonparametric Gaussian deconvolution are discussed by Fan (1992). Horrace and Parmeter (2011) find that consistent estimation of the distribution of u_i in a semparametric SF model, in which $v_i \sim N\left(0, \sigma_v^2\right)$ and f_u is unknown, has a $\ln n$ convergence rate. This implies that convergence rates when $v_i \sim N\left(0, \sigma_v^2\right)$ are rather slow.

In the aforementioned proposals, the distribution of u_i is the left truncation at zero of the distribution of v_i. In many cases, this ensures that f_ε can be expressed analytically. Proposition 9 of Azzalini and Capitanio (2003) shows the density of the sum of a random variable and the absolute value of another random variable following the same elliptical distribution. Stead et al. (2018) propose the use of MSL to pair a thick-tailed distribution for v_i with any given distribution for u_i, and estimate a logistic-half normal SF model. The authors show that the model yields a narrower range of efficiency scores compared to the normal-half normal model.

There are two drawbacks of the above proposals for v_i. First, they have fixed shapes, so there is no flexibility in the heaviness of their tails. Second, they do not nest the normal distribution, which makes testing against the standard SF model difficult. One potential noise distribution with neither of these shortcomings is the Student's t distribution, which has a 'degrees of freedom' parameter α that determines the heaviness of the tails, and which approaches the normal distribution as $\alpha \to \infty$. Tancredi (2002) proposed an SF model in which v_i and u_i follow non-standard Student's t distribution and half t distributions, with scale parameters σ_v and σ_u, respectively,

---

[20]In keeping with previous terminology, 'truncated' (without further qualification) refers specifically to the left truncation at zero of a distribution with mean μ, and 'half' refers to the special case where $\mu = 0$. Note that truncating the Laplace distribution thus yields the exponential distribution whenever $\mu \leq 0$ due to the memorylessness property of the exponential distribution.

and a common degrees of freedom parameter α. The author shows that $f_{u|\varepsilon} \to 0$ as $s\varepsilon_i \to \infty$ and that $E\left[\exp\left(-u_i|\varepsilon_i\right)\right]$ and $\text{VAR}\left[\exp\left(-u_i|\varepsilon_i\right)\right]$ are non-monotonic functions of ε_i. Wheat et al. (2019) estimate a t-half normal model via MSL, similarly finding that $E(u_i|\varepsilon_i)$ is non-monotonic, decreasing with $s\varepsilon_i$ at either tail, and discuss testing against the normal-half normal SF model. Bayesian estimation of the t-half t model, and of t-half normal, t-exponential, and t-gamma SF models are discussed by Tchumtchoua and Dey (2007) and Griffin and Steel (2007), respectively.

Another proposal which nests the standard SF model and allows for flexibility in the kurtosis of v_i, is that of Wheat et al. (2017), in which v_i follows a mixture of two normal distributions with zero means, variances σ_{v1}^2 and σ_{v2}^2, respectively, and mixing parameter p. This is often referred to as the contaminated normal distribution.[21] Alternatively, the model can be interpreted as a latent class model with two regimes having differing noise variances. Efficiency prediction in latent class and mixture SF models is discussed, and $E(u_i|\varepsilon_i)$ is shown to be non-monotonic in the contaminated normal-half normal case, as in the t-half normal. Testing down to the standard SF model is less straightforward in this case, since there is an unidentified parameter under the null hypothesis.

The proposals discussed in this section have all been motivated to one degree or another by the need to accommodate outliers in a satisfactory way. An exception to this general rule is Bonanno et al. (2017), who propose an SF model with correlated error components—for a discussion of such models, see Sect. 8.1—in which the marginal distributions of v_i and u_i are skew logistic and exponential, respectively. The motivation in this case is to allow for non-zero efficiency predictions in the presence of 'wrong skew', which the model ascribes to the skewness of v_i.

7 Presence of Efficient Firms

A number of papers have considered SFA in the case where some significant proportion of firms lie on the frontier—i.e. are fully efficient—and discussed SF specifications and efficiency prediction appropriate for this case, along with methods used to identify subset of efficient firms.

Horrace and Schmidt (2000) discuss multiple comparisons with the best (MCB)—see Hsu (1981, 1984) for background on MCB—in which there

---

[21]Or more specifically, the scale contaminated normal distribution.

are I populations each with their own distinct parameter values, a_i, one of which—e.g. the maximum or the minimum—is the 'best' in some sense, against which we want to compare the remaining $I - 1$ populations. Rather than make individual comparisons, e.g. by testing $H_0 : a_i = a_b$ where $a_b = \max_{j \neq i} sa_j$), MCB constructs joint confidence intervals for a vector of differences $\left(a_b - a_1 \; a_b - a_2 \; \ldots \; a_b - a_{I-1} \right)$. This is motivated by the need to consider the 'multiplicity effect' (Hochberg and Tamhane 1987), i.e. the fact that if a large enough number of comparisons are made, some differences are bound to appear significant. MCB is also concerned with constructing a set of populations which could be the best. Horrace and Schmidt (2000) discuss application of MCB to derive such multivariate intervals in the context of the fixed effects, time-invariant efficiency panel SF model of Schmidt and Sickles (1984), and the selection of a set of efficient (or probably efficient) firms based on these.

An alternative approach proposed by Jung (2017) is to use a least absolute shrinkage and selection operator (LASSO) variant of the Schmidt and Sickles (1984) model. LASSO is a method used for variable selection and to penalise overfitting by shrinking the parameter estimates towards zero and was introduced by Tibshirani (1996) in the context of OLS, such that

$$\hat{\beta}_{LASSO} = \underset{\beta}{\mathrm{argmin}} \left[\frac{1}{I} \sum_{i=1}^{I} \varepsilon_i^2 + \lambda \sum_{k=1}^{K} |\beta_k| \right] \qquad (25)$$

where K is the number of regressors, and λ is a tuning parameter that determines the strength of the penalty (or the degree of shrinkage). The constant term β_0 is excluded from the penalty term. The penalty is such that it forces some of the coefficients to be zero, hence, its usefulness in variable selection. It is straightforward to extend the approach to a fixed-effects panel data model. Jung (2017) proposes extending the approach to the Schmidt and Sickles (1984) fixed effects SF model, in which $\beta_0 = \max_j sa_j$ and $u_i = \max_j sa_j - sa_i$, and introduces an additional penalty term such that the inefficiency parameters are shrunk towards zero, and $u_i = 0$ for a subset of firms. The author discusses the properties of the model, and in applying the model to a dataset used by Horrace and Schmidt (2000), notes that the resulting set of efficient firms is similar to that obtained using the MCB approach.

Kumbhakar et al. (2013) proposed a zero inefficiency stochastic frontier (ZISF) model. The ZISF model adapts the standard parametric SF model to account for the possibility that a proportion, p, of the firms in the sample

are fully efficient using a latent class approach in which $u_i = 0$ with probability p. That is, the ZISF model is a latent class model in which

$$f_\varepsilon(\varepsilon_i) = pf_v(\varepsilon_i) + (1-p)\int_0^\infty f_v(\varepsilon_i + su_i)f_u(u_i)\mathrm{d}u_i \qquad (26)$$

where f_v is the density of v_i and assumed noise distribution, and f_u is the density of u_i in the second regime. In the first regime, u_i can be thought of as belonging to a degenerate distribution at zero. The ZISF model nests the standard SF model when $p = 0$, and testing down to the SF model is a standard problem. On the other hand, testing $H_0 : p = 1$, i.e. that all firms are fully efficient, is more complicated, that the splitting proportion p lies on the boundary of the parameter space in this case. The authors suggest that the LR statistic follows a $\chi^2_{1:0}$ distribution.[22] That is, a 50:50 mixture of χ^2_0 and χ^2_1 distributions. However, Rho and Schmidt (2015) question the applicability of this result, noting an additional complication: under $H_o : p = 1$, σ_u is not identified. Equivalently, p is not identified under $H_o : \sigma_u = 0$. Simulation evidence provided by the authors suggests that estimates of these two parameters are likely to be imprecise when either is small.

Kumbhakar et al. (2013) suggest several approaches to efficiency prediction from the ZISF model. First, the authors suggest weighting regime-specific efficiency predictions by unconditional probabilities of regime membership. Since $\hat{u}_i = 0$ in the first regime regardless of the predictor used, this amounts to using $(1-p)E(u_i|\varepsilon_i)$. This is clearly unsatisfactory, as each firm is assigned the same (unconditional) probabilities for regime membership. A preferable alternative, suggested by both Kumbhakar et al. (2013) and Rho and Schmidt (2015), suggest using $(1-p_i)E(u_i|\varepsilon_i)$, where $p_i = pf_v(\varepsilon_i)/f_\varepsilon(\varepsilon_i)$, which is a firm-specific probability conditional on ε_i. Note that $(1-p)E(u_i|\varepsilon_i)$ and $(1-p_i)E(u_i|\varepsilon_i)$ for all i and any value of ε_i will yield non-zero predictions of u_i under the assumption that $v_i \sim N(0, \sigma_v^2)$ (see the discussion of the monotonicity properties of $E(u_i|\varepsilon_i)$ in Sect. 6), despite the fact we expect pI efficient firms in the sample. Kumbhakar et al. (2013) suggest identifying firms as efficient when p_i is greater than some cut-off point; however, the choice of such a cut-off point is arbitrary.

---

[22]As discussed in Sect. 3, see Case 5 in Self and Liang (1987).

Despite the ZISF model's motivation, efficient firms cannot be identified on the basis of the resulting point predictions of efficiency or conditional probabilities of regime membership. Firms may be predicted as fully efficient if the conditional mode predictor is used, or possibly if an alternative distribution for v_i is assumed (again, refer to Sect. 6), but this is equally true in the standard SF context. An appropriate approach to classifying firms would be to identify those with minimum width prediction intervals, analogous to those derived by Wheat et al. (2014) for $u_i|\varepsilon_i$ in the standard SF model, including zero.

There are trade-offs between each of the three proposed methods. Compared to the ZISF model, the MCB and LASSO approaches have the advantage that no particular distribution for u_i is imposed, and efficient firms can be identified on the basis of hypothesis tests. In contrast, the ZISF model limits us to examining prediction intervals. On the other hand, Horrace and Schmidt (2000) and Jung (2017) assume time-invariant efficiency. While Horrace and Schmidt (2000) state that the MCB approach could be adapted to allow for time-varying efficiency (and the same may be true of the LASSO approach), the ZISF approach is the only one that can be applied to cross-sectional data. In addition, it would be straightforward to extend the ZISF approach to incorporate many features found in the SF literature.

8 Miscellaneous Proposals

In this section, we discuss several of the lesser and relatively tangential strands of the SF literature which have adopted novel distributional forms.

8.1 Correlated Errors

A common assumption across all of the aforementioned SF specifications is that the error components, including all noise, inefficiency and random effects components are distributed independently of one another.[23] Relaxing this assumption seems particularly justified in cases in which there are two or more inefficiency components. Independence between noise and inefficiency terms is usually assumed on the basis that noise represents random

---

[23]Again, as an exception to this, dependency between error components may be introduced via 'environmental' variables influencing the parameters of their distributions as discussed in Sect. 5.

factors unrelated to efficiency. On the other hand, it has been argued that such factors may affect firm decision making and therefore efficiency.

Similar to the panel data case discussed in Sect. 4.1, one approach to modelling dependence between errors has been to specify some multivariate analogue to common distributional assumptions under independence. Schmidt and Lovell (1980), Pal and Sengupta (1999), and Bandyopadhyay and Das (2006) consider a left truncated a bivariate normal distribution at zero with respect to a one-sided inefficiency component.[24] The two-sided component represents noise in the latter two cases and allocative inefficiency in the former. Pal and Sengupta (1999) likewise included allocative inefficiency components, which are assumed to follow a multivariate normal distribution. However, the marginal distributions of the error components are not those commonly used under independence and, more importantly, that they may be inappropriate. Bandyopadhyay and Das (2006) show that while the marginal distribution of u_i in their model is half normal, that of v_i is skew normal, with skewness determined by the correlation between the two error components. An unusual approach was proposed by Pal (2004), in which conditional distributions for the error components are specified directly along with their marginal distributions. Prediction of efficiency is based on $f_{u|\varepsilon}$ as in the case of independence.

The use of a copula function to allow for dependence between v_i and u_i was proposed by Smith (2008) and El Mehdi and Hafner (2014). Various alternatives are considered, including the Ali-Mikhail-Haq, Clayton, Fairlie-Gumbel-Morgenstern, Frank and Gaussian copula. From Sklar's theorem, the joint density $f_{v,u}$ is the product of the marginal densities and the density of the copula. It follows that (3) and (4) must be modified such that

$$f_\varepsilon(\varepsilon_i) = \int_0^\infty f_v(\varepsilon_i + su_i)f_u(u_i)c_{v,u}[F_v(\varepsilon_i + su_i), F_u(u_i)]\mathrm{d}u_i \qquad (27)$$

and

$$f_{u_i|\varepsilon_i}(u_i|\varepsilon_i) = \frac{f_v(\varepsilon_i + su_i)f_u(u_i)c_{v,u}[F_v(\varepsilon_i + su_i), F_u(u_i)]}{f_\varepsilon(\varepsilon_i)} \qquad (28)$$

---

[24]Schmidt and Lovell (1980) fold, rather than truncate.

where $c_{v,u}$ is the copula density. Gómez-Déniz and Pérez-Rodríguez ([2015](#)) specify a bivariate Sarmanov distribution for v_i and u_i with normal and half-normal marginal distributions, respectively. Again, the advantage of the copula approach is that the desired marginal distributions are obtained, with the dependence between the error components captured by $c_{v,u}$.

8.2   Sample Selection and Endogenous Switching

In the context of linear regression, the sample selection model of Heckman ([1976](#), [1979](#)) is such that

$$y_i = \begin{cases} x_i\beta + \varepsilon_i, & d_i = 1 \\ \text{unobserved,} & d_i = 0 \end{cases}, \quad d_i = I\big(d_i^* = z_i\alpha + w_i > 0\big), \quad (29)$$

where symmetric error terms ε_i and w_i are assumed to follow a bivariate normal distribution with zero means, variances σ_ε^2 and 1, and correlation coefficient ρ. Unless $\rho = 0$, least squares will yield biased estimates. Since $E(y_i|x_i, d_i = 1) = x_i\beta + \rho\sigma_\varepsilon f_w(z_i\alpha)/F_w(z_i\alpha)$, Heckman ([1979](#)) proposed a two-step, limited information method in which y_i is regressed on x_i and the inverse Mills' ratio $f_w\big(z_i\hat{\alpha}\big)/F_w\big(z_i\hat{\alpha}\big)$, where $\hat{\alpha}$ is obtained from a single equation probit model estimated by ML. Alternatively, a full information ML approach may be used to estimate the parameters of the model simultaneously, as in Heckman ([1976](#)) and Maddala ([1983](#)).

A similar problem is that of endogenous switching. The endogenous switching model of Heckman ([1978](#)) has two regimes, membership of which is dependent upon a binary switching dummy:

$$y_i = \begin{cases} x_i\beta_1 + \varepsilon_{1i}, \ d_i = 1 \\ x_i\beta_2 + \varepsilon_{2i}, \ d_i = 0 \end{cases}, \quad d_i = I\big(d_i^* = z_i\alpha + w_i > 0\big) \quad (30)$$

where ε_{1i}, ε_{2i} and w_i are assumed to follow a trivariate normal distribution with zero means, and variances $\sigma_{1\varepsilon}^2$, $\sigma_{2\varepsilon}^2$, and σ_w^2. The correlations of ε_{1i} and ε_{2i} with w_i are given by ρ_1 and ρ_2, respectively, while ρ_{12} is the correlation between ε_{1i} and ε_{2i}. Again, both two-step partial information and full information ML approaches may be used to estimate the parameters of the model.

In recent years, SF models incorporating sample selection and endogenous switching have been proposed. Bradford et al. ([2001](#)) and Sipiläinen and Oude Lansink ([2005](#)) use the Heckman ([1979](#)) two-step

approach, including the estimated inverse Mills' ratios from single equation probit selection and switching models, respectively, as independent variables in their SF models. However, this is inappropriate in non-linear settings such as SFA, since it is generally *not* the case that $E\left[g(x_i\beta + \varepsilon_i)|d_i = 1\right] = g\left[x_i\beta + \rho\sigma_\varepsilon f_w(z_i\alpha)/F_w(z_i\alpha)\right]$ where g is some non-linear function. Terza (2009) discusses ML estimation of non-linear models with endogenous switching or sample selection in general.

In the SF context, there are many alternative assumptions that may be made about the relationship between noise, inefficiency, and the stochastic component of the selection (or switching) equation. Perhaps the natural approach, implicit in Bradford et al. (2001) and Sipiläinen and Oude Lansink (2005), is to assume that the symmetric noise terms follow a multivariate normal distribution as in the linear model, while the inefficiency terms are drawn from independent one-sided univariate distributions. This is proposed by Greene (2010), who estimates an SF model with sample selection via MSL, and also by Lai (2015), who uses the result that, in both the sample selection and endogenous switching cases, $f_{\varepsilon|d}$ follows a closed skew-normal distribution when the inefficiency terms are truncated normal. This results in analytical log-likelihoods, and the author proposes to predict efficiency based on the distribution of $u_i|(\varepsilon_i|d_i)$, specifically using $E\left[\exp(-u_i)|(\varepsilon_i|d_i)\right]$.

Note that the distributional assumptions in Greene (2010) and Lai (2015) ensure appropriate marginal distributions for each error component, but do not allow for correlation between the inefficiency terms and the symmetric errors. Lai et al. (2009) introduce correlation between ε_i (rather than its components) and w_i through a copula function. Departing from the usual approach, Kumbhakar et al. (2009) propose an SF model with an endogenous switching equation in which $d_i^* = z_i\alpha + \delta u_i + w_i$. That is, they include the inefficiency term as a determinant of regime membership.[25] The various error components are assumed to be independent of one another, and both the log-likelihood of the model and $E[u_i|(\varepsilon_i|d_i)]$ are obtained by quadrature.

---

[25]Kumbhakar et al. (2009), using panel data, also include a lagged regime membership (i.e. technology choice) dummy in their selection equation.

8.3   Two-Tiered Models

SF methods have been widely applied outside of the context of production and cost frontier estimation. Most applications have utilised standard cross-section or panel data SF specifications, or some of the variants discussed above. However, one area of application which has seen its own distinct methodological developments is modelling of earnings determination. Polachek and Yoon (1987) proposed a 'two-tiered' SF (2TSF) model in which

$$\varepsilon_i = v_i - u_i + w_i, \tag{31}$$

where v_i is again a normally distributed noise component, and u_i and w_i follow exponential distributions with means σ_u and σ_w, respectively.[26] The dependent variable is a worker's actual wage. The u_i component captures deviations from the firm's reservation wage—i.e. the maximum wage offers the firm would make—as a result of incomplete information on the part of the employee. Similarly, w_i captures deviations from the worker's reservation wage—i.e. the minimum wage offer the worker would accept—as a result of incomplete information on the part of the employer. The inclusion of these two terms therefore allows estimation of the extent of average employee and employer incomplete information, and even observation-specific predictions of these. The assumption of exponentially distributed u_i and w_i makes derivation of f_ε, and therefore the log-likelihood, straightforward. However, as in the standard SF model, alternative distributional assumptions have been proposed: Papadopoulos (2015) derive a closed form for f_ε when u_i and w_i follow half-normal distributions, and Tsionas (2012) estimates the model assuming that they follow gamma distributions via inverse fast Fourier transform of the characteristic function as discussed in Sect. 3.

In general, developments of the 2TSF model have tended to parallel those of the standard SF model. A panel data 2TSF model was proposed by Polachek and Yoon (1996), in which

$$\varepsilon_{ift} = v_{ift} - u_{it} + w_{ft} \tag{32}$$

where the subscript f denotes the firm. The employee incomplete information component u_{it} and the employer incomplete information component

---

[26]The authors actually use u_i to denote the noise term and v_i and w_i for the one-sided errors. In the interest of consistency and to avoid confusion, we use v_i to refer to the noise term and u_i and w_i for the one-sided errors.

w_f, which is assumed to be constant across all employees, are further decomposed such that $u_{it} = u_i + u_{it}^*$ and $w_{ft} = w_f + w_{ft}^*$, where u_i and w_f are time-invariant fixed effects and u_{it}^* and w_{ft}^* follow independent exponential distributions. It is clear that many alternative panel data specifications could be proposed, particularly considering the numerous possible extensions of the models discussed in Sect. 4.

In addition, and analogous to the models discussed in Sect. 5, modelling of u_i and w_i in terms of vectors of explanatory variables has been proposed. Assuming exponential u_i and w_i, Groot and Oosterbeek (1994) propose modelling the inverse signal-to-noise ratios σ_v/σ_u and σ_v/σ_w as linear functions of vectors z_{ui} and z_{wi}. This specification introduces heteroskedasticity of each of the error components, but in rather an odd way, and is problematic in that it does not restrict σ_u or σ_w to be positive. This issue is resolved by Kumbhakar and Parmeter (2010), who propose a specification in which $\sigma_{ui} = \exp(z_{ui}d_u)$ and $\sigma_{wi} = \exp(z_{wi}d_w)$. Note that this model has the scaling property. Parmeter (2018) proposes estimating a 2TSF model with the scaling property, avoiding explicit distributional assumptions, by non-linear least squares.

Finally, tying back to the previous section, Blanco (2017) proposes an extension of the basic Polachek and Yoon (1987) model to account for sample selection, assuming that the symmetric error components follow a bivariate normal distribution, while the one-sided errors follow independent univariate exponential distributions.

9   Conclusion

The methodological literature on SFA has developed considerably since the first SF models were developed by Aigner et al. (1977) and Meeusen and van Den Broeck (1977). The defining feature of SFA models is the focus on determining observation-specific predictions for inefficiency. This in turn requires a prediction of an inefficiency error terms which is present in tandem with a noise error. Hence, there is a deconvolution problem associated with the error in the model. As such, distributional assumptions are not just required to get 'best' estimates of the underlying frontier relationship (cost frontier, production frontier, etc.), but also essential for enabling appropriate predictions of the quantity of interest: firm inefficiency.

This review has considered numerous ways in which SFA has been innovated, which in turn has involved the use of differing distributional forms. One strand of literature concerns alternative distributional assumptions for

the inefficiency error term, and more recently, the noise error term. This raises the obvious question as to which to choose. Given economic theory only requires the inefficiency error to be one-sided, it is generally an empirical matter as to which is to be preferred. Formulations which nest other forms as special cases have obvious appeal; however, there are also non-nested tests, such as those developed by Wang et al. (2011) to aid selection.

Another strand of literature considers alternative distributions in the presence of specific empirical issues. The 'wrong-skew' problem is a good example, where it is entirely plausible that inefficiency could be found to have skewness counter to the direction imposed by the use of the common, half-normal, truncated-normal or exponential inefficiency distributions. Without a change to the distributional assumptions, the model estimation would indicate no evidence of inefficiency which is often difficult to justify in the context of knowledge and other available evidence of the performance of the industries that these techniques are applied to.

Other innovations include models for sample selection, the presence of efficient firms and two-tier SF models. Panel data is a data structure which greatly increases the scope of modelling possibilities. It potentially allows for construction of predictors of inefficiency without appeal to 'full' distributional assumptions on the noise and inefficiency (instead only requiring moment assumptions), by exploiting time persistency in inefficiency. Alternatively, full parametric approaches can be adopted, with the benefit of being able to obtain separate predictions for inefficiency—which may have both time-invariant and time-varying components—and time-invariant unobserved heterogeneity.

Finally, a strand of literature has developed characterising heteroskedasticity in the error components. This is of particular interest as it allows for quantification of the determinants of inefficiency, which is important in beginning to explain why there is a performance gap for a firm in addition to providing a prediction of the size of such a gap. This, in turn can be used by stakeholders to guide implementation of better performance.

Overall it is misleading to think of SFA as representing a single approach to efficiency analysis. Instead, SFA characterises a broad set of models, where different approaches will be relevant given the empirical context. The limited scope of this review has excluded several topics such as nonparametric SF models, Bayesian SF models, metafrontiers, and estimation of distance functions. Inefficiency is an unobserved error component, and so by definition, the predictor of such an error will be sensitive to distributional assumptions regarding inefficiency and the other unobserved error components, such as noise and unobserved heterogeneity. Thus, the conclusion is that for any

given empirical application of efficiency analysis, several SFA models will need to be considered in order to establish the sensitivity of the efficiency predictions to the distributional assumptions adopted. This review should provide a useful starting point for such an exercise.

References

Afriat, S.N. 1972. Efficiency estimation of production functions. *International Economic Review* 13 (3): 568–598.

Ahn, S.C., Y.H. Lee, and P. Schmidt. 2007. Stochastic frontier models with multiple time-varying individual effects. *Journal of Productivity Analysis* 27 (1): 1–12.

Ahn, S.C., Y.H. Lee, and P. Schmidt. 2013. Panel data models with multiple time-varying individual effects. *Journal of Econometrics* 174 (1): 1–14.

Aigner, D.J., and S.F. Chu. 1968. On estimating the industry production function. *The American Economic Review* 58 (4): 826–839.

Aigner, D.J., T. Amemiya, and D.J. Poirier. 1976. On the estimation of production frontiers: Maximum likelihood estimation of the parameters of a discontinuous density function. *International Economic Review* 17 (2): 377–396.

Aigner, D., C.A.K. Lovell, and P. Schmidt. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6 (1): 21–37.

Almanidis, P., and R.C. Sickles. 2012. The skewness issue in stochastic frontiers models: Fact or fiction? In *Exploring research frontiers in contemporary statistics and econometrics: A festschrift for Léopold Simar*, ed. I. Van Keilegom and W.P. Wilson, 201–227. Heidelberg: Physica-Verlag HD.

Almanidis, P., J. Qian, and R.C. Sickles. 2014. Stochastic frontier models with bounded inefficiency. In *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications*, ed. R.C. Sickles and W.C. Horrace, 47–81. New York, NY: Springer.

Alvarez, A., C. Amsler, L. Orea, and P. Schmidt. 2006. Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. *Journal of Productivity Analysis* 25 (3): 201–212.

Amemiya, T., and T.E. MaCurdy. 1986. Instrumental-variable estimation of an error-components model. *Econometrica* 54 (4): 869–880.

Amsler, C., A. Prokhorov, and P. Schmidt. 2014. Using copulas to model time dependence in stochastic frontier models. *Econometric Reviews* 33 (5–6): 497–522.

Amsler, C., P. Schmidt, and W.-J. Tsay. 2015. A post-truncation parameterization of truncated normal technical inefficiency. *Journal of Productivity Analysis* 44 (2): 209–220.

Andrews, D.W.K. 1993a. An introduction to econometric applications of empirical process theory for dependent random variables. *Econometric Reviews* 12 (2): 183–216.

Andrews, D.W.K. 1993b. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61 (4): 821–856.

Azzalini, A. 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12 (2): 171–178.

Azzalini, A., and A. Capitanio. 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65 (2): 367–389.

Badunenko, O., and S.C. Kumbhakar. 2016. When, where and how to estimate persistent and transient efficiency in stochastic frontier panel data models. *European Journal of Operational Research* 255 (1): 272–287.

Bandyopadhyay, D., and A. Das. 2006. On measures of technical inefficiency and production uncertainty in stochastic frontier production model with correlated error components. *Journal of Productivity Analysis* 26 (2): 165–180.

Barrow, D.F., and A.C. Cohen. 1954. On some functions involving Mill's ratio. *The Annals of Mathematical Statistics* 25 (2): 405–408.

Battese, G.E., and T.J. Coelli. 1988. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics* 38 (3): 387–399.

Battese, G.E., and T.J. Coelli. 1992. Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis* 3 (1): 153–169.

Battese, G.E., and T.J. Coelli. 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics* 20 (2): 325–332.

Battese, G.E., and G.S. Corra. 1977. Estimation of a production frontier model: With application to the pastoral zone of Eastern Australia. *Australian Journal of Agricultural Economics* 21 (3): 169–179.

Battese, G.E., A.N. Rambaldi, and G.H. Wan. 1997. A stochastic frontier production function with flexible risk properties. *Journal of Productivity Analysis* 8 (3): 269–280.

Baumol, W.J. 1967. *Business behavior, value and growth,* revised ed. New York: Macmillan.

Beckers, D.E., and C.J. Hammond. 1987. A tractable likelihood function for the normal-gamma stochastic frontier model. *Economics Letters* 24 (1): 33–38.

Belotti, F., and G. Ilardi. 2018. Consistent inference in fixed-effects stochastic frontier models. *Journal of Econometrics* 202 (2): 161–177.

Bera, A.K., and S.C. Sharma. 1999. Estimating production uncertainty in stochastic frontier production function models. *Journal of Productivity Analysis* 12 (3): 187–210.

Blanco, G. 2017. Who benefits from job placement services? A two-sided analysis. *Journal of Productivity Analysis* 47 (1): 33–47.

Bonanno, G., D. De Giovanni, and F. Domma. 2017. The 'wrong skewness' problem: A re-specification of stochastic frontiers. *Journal of Productivity Analysis* 47 (1): 49–64.

Bracewell, R.N. 1978. *The Fourier transform and its applications*, 2nd ed. New York: McGraw-Hill.

Bradford, W.D., A.N. Kleit, M.A. Krousel-Wood, and R.N. Re. 2001. Stochastic frontier estimation of cost models within the hospital. *The Review of Economics and Statistics* 83 (2): 302–309.

Brorsen, B.W., and T. Kim. 2013. Data aggregation in stochastic frontier models: the closed skew normal distribution. *Journal of Productivity Analysis* 39 (1): 27–34.

Carree, M.A. 2002. Technological inefficiency and the skewness of the error component in stochastic frontier analysis. *Economics Letters* 77 (1): 101–107.

Caudill, S.B., and J.M. Ford. 1993. Biases in frontier estimation due to heteroscedasticity. *Economics Letters* 41 (1): 17–20.

Caudill, S.B., J.M. Ford, and D.M. Gropper. 1995. Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business & Economic Statistics* 13 (1): 105–111.

Chamberlain, G. 1984. Panel data. In *Handbook of econometrics*, ed. Z. Griliches and M.D. Intriligator, 1247–1318. Amsterdam: Elsevier.

Chen, Y.-Y., P. Schmidt, and H.-J. Wang. 2014. Consistent estimation of the fixed effects stochastic frontier model. *Journal of Econometrics* 181 (2): 65–76.

Coelli, T. 1995. Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis. *Journal of Productivity Analysis* 6 (3): 247–268.

Coelli, T.J., D.S.P. Rao, and G.E. Battese. 2005. *An introduction to efficiency and productivity analysis*, 2nd ed. New York: Springer.

Colombi, R., S.C. Kumbhakar, G. Martini, and G. Vittadini. 2014. Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *Journal of Productivity Analysis* 42 (2): 123–136.

Cornwell, C., P. Schmidt, and R.C. Sickles. 1990. Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics* 46 (1): 185–200.

Cuesta, R.A. 2000. A production model with firm-specific temporal variation in technical inefficiency: With application to Spanish dairy farms. *Journal of Productivity Analysis* 13 (2): 139–158.

Debreu, G. 1951. The coefficient of resource utilization. *Econometrica* 19 (3): 273–292.

Deprins, D., and L. Simar. 1989a. Estimating technical inefficiencies with correction for environmental conditions. *Annals of Public and Cooperative Economics* 60 (1): 81–102.

Deprins, D., and L. Simar. 1989b. Estimation de frontières déterministes avec facteurs exogènes d'inefficacitè. *Annales d'Économie et de Statistique* 14: 117–150.

El Mehdi, R., and C.M. Hafner. 2014. Inference in stochastic frontier analysis with dependent error terms. *Mathematics and Computers in Simulation* 102 (Suppl. C): 104–116.

Fan, J. 1991. On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics* 19 (3): 1257–1272.

Fan, J. 1992. Deconvolution with supersmooth distributions. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* 20 (2): 155–169.

Farrell, M.J. 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)* 120 (3): 253–290.

Filippini, M., and W. Greene. 2016. Persistent and transient productive inefficiency: a maximum simulated likelihood approach. *Journal of Productivity Analysis* 45 (2): 187–196.

Gómez-Déniz, E., and J.V. Pérez-Rodríguez. 2015. Closed-form solution for a bivariate distribution in stochastic frontier models with dependent errors. *Journal of Productivity Analysis* 43 (2): 215–223.

González-Farías, G., J.A. Domínguez-Molina, and A.K. Gupta. 2004a. Additive properties of skew normal random vectors. *Journal of Statistical Planning and Inference* 126 (2): 521–534.

González-Farías, G., J.A. Domínguez-Molina, and A.K. Gupta. 2004b. The closed skew-normal distribution. In *Skew-elliptical distributions and their applications*, ed. M.G. Genton, 25–42. Boca Raton: Chapman and Hall/CRC.

Gouriéroux, C., A. Holly, and A. Monfort. 1982. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* 50 (1): 63–80.

Greene, W.H. 1980. Maximum likelihood estimation of econometric frontier functions. *Journal of Econometrics* 13 (1): 27–56.

Greene, W.H. 1990. A gamma-distributed stochastic frontier model. *Journal of Econometrics* 46 (1): 141–163.

Greene, W.H. 2003. Simulated likelihood estimation of the normal-gamma stochastic frontier function. *Journal of Productivity Analysis* 19 (2): 179–190.

Greene, W.H. 2004. Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems. *Health Economics* 13 (10): 959–980.

Greene, W.H. 2005a. Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis* 23 (1): 7–32.

Greene, W.H. 2005b. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics* 126 (2): 269–303.

Greene, W.H. 2008. The econometric approach to efficiency analysis. In *The measurement of productive efficiency and productivity growth*, 2nd ed., ed. H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, 92–159. Oxford: Oxford University Press.

Greene, W.H. 2010. A stochastic frontier model with correction for sample selection. *Journal of Productivity Analysis* 34 (1): 15–24.

Greene, W.H. 2016. *LIMDEP Version 11.0 econometric modeling guide.* Econometric Software.

Greene, W.H., and S. Misra. 2003. Simulated maximum likelihood estimation of general stochastic frontier regressions. Working Paper, William Simon School of Business, University of Rochester.

Griffin, J.E., and M.F.J. Steel. 2007. Bayesian stochastic frontier analysis using WinBUGS. *Journal of Productivity Analysis* 27 (3): 163–176.

Griffin, J.E., and M.F.J. Steel. 2008. Flexible mixture modelling of stochastic frontiers. *Journal of Productivity Analysis* 29 (1): 33–50.

Griffiths, W.E., and G. Hajargasht. 2016. Some models for stochastic frontiers with endogeneity. *Journal of Econometrics.* 190 (2): 341–348.

Groot, W., and H. Oosterbeek. 1994. Stochastic reservation and offer wages. *Labour Economics* 1 (3): 383–390.

Grushka, E. 1972. Characterization of exponentially modified Gaussian peaks in chromatography. *Analytical Chemistry* 44 (11): 1733–1738.

Gupta, A.K., and N. Nguyen. 2010. Stochastic frontier analysis with fat-tailed error models applied to WHO health data. *International Journal of Innovative Management, Information & Production* 1 (1): 43–48.

Hadri, K. 1999. Estimation of a doubly heteroscedastic stochastic frontier cost function. *Journal of Business & Economic Statistics* 17 (3): 359–363.

Hajargasht, G. 2014. The folded normal stochastic frontier model. Working Paper.

Hajargasht, G. 2015. Stochastic frontiers with a Rayleigh distribution. *Journal of Productivity Analysis* 44 (2): 199–208.

Hansen, B.E. 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64 (2): 413–430.

Hausman, J.A. 1978. Specification tests in econometrics. *Econometrica* 46 (6): 1251–1271.

Hausman, J.A., and W.E. Taylor. 1981. A generalized specification test. *Economics Letters* 8 (3): 239–245.

Heckman, J.J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5 (4): 17.

Heckman, J.J. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46 (4): 931–959.

Heckman, J.J. 1979. Sample selection bias as a specification error. *Econometrica* 47 (1): 153–161.

Heshmati, A., and S.C. Kumbhakar. 1994. Farm heterogeneity and technical efficiency: Some results from Swedish dairy farms. *Journal of Productivity Analysis* 5 (1): 45–61.

Hicks, J.R. 1935. Annual survey of economic theory: The theory of monopoly. *Econometrica* 3 (1): 1–20.

Hochberg, Y., and A.C. Tamhane. 1987. *Multiple comparison procedures*, 1st ed. New York: Wiley.

Horrace, W.C. 2005. Some results on the multivariate truncated normal distribution. *Journal of Multivariate Analysis* 94 (1): 209–221.

Horrace, W.C. 2015. Moments of the truncated normal distribution. *Journal of Productivity Analysis* 43 (2): 133–138.

Horrace, W.C., and C.F. Parmeter. 2011. Semiparametric deconvolution with unknown error variance. *Journal of Productivity Analysis* 35 (2): 129–141.

Horrace, W.C., and C.F. Parmeter. 2018. A Laplace stochastic frontier model. *Econometric Reviews* 37 (3): 260–280.

Horrace, W.C., and P. Schmidt. 2000. Multiple comparisons with the best, with economic applications. *Journal of Applied Econometrics* 15 (1): 1–26.

Hsu, J.C. 1981. Simultaneous confidence intervals for all distances from the "best". *The Annals of Statistics* 9 (5): 1026–1034.

Hsu, J.C. 1984. Constrained simultaneous confidence intervals for multiple comparisons with the best. *The Annals of Statistics* 12 (3): 1136–1144.

Huang, C.J., and J.-T. Liu. 1994. Estimation of a non-neutral stochastic frontier production function. *Journal of Productivity Analysis* 5 (2): 171–180.

Johnston, J. 1960. *Statistical cost analysis*, 1st ed. New York: McGraw-Hill.

Jondrow, J., C.A. Knox Lovell, I.S. Materov, and P. Schmidt. 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics* 19 (2): 233–238.

Jovanovic, B., and Y. Nyarko. 1996. Learning by doing and the choice of technology. *Econometrica* 64 (6): 1299–1310.

Jung, H. 2017. Adaptive LASSO for stochastic frontier models with many efficient firms. Working Paper, Maxwell School of Citizenship and Public Affairs, Syracuse University.

Just, R.E., and R.D. Pope. 1978. Stochastic specification of production functions and economic implications. *Journal of Econometrics* 7 (1): 67–86.

Kneip, A., R.C. Sickles, and W. Song. 2012. A new panel data treatment for heterogeneity in time trends. *Econometric Theory* 28 (3): 590–628.

Kodde, D.A., and F.C. Palm. 1986. Wald criteria for jointly testing equality and inequality restrictions. *Econometrica* 54 (5): 1243–1248.

Koopmans, T.C. 1951. Efficient allocation of resources. *Econometrica* 19 (4): 455–465.

Kumbhakar, S.C. 1990. Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics* 46 (1): 201–211.

Kumbhakar, S.C. 1991. Estimation of technical inefficiency in panel data models with firm- and time-specific effects. *Economics Letters* 36 (1): 43–48.

Kumbhakar, S.C. 1993. Production risk, technical efficiency, and panel data. *Economics Letters* 41 (1): 11–16.

Kumbhakar, S.C., and A. Heshmati. 1995. Efficiency measurement in Swedish dairy farms: An application of rotating panel data, 1976–88. *American Journal of Agricultural Economics* 77 (3): 660–674.

Kumbhakar, S.C., and L. Hjalmarsson. 1995. Labour-use efficiency in Swedish social insurance offices. *Journal of Applied Econometrics* 10 (1): 33–47.

Kumbhakar, S.C., and C.A.K Lovell. 2000. *Stochastic frontier analysis*, 1st ed. Cambridge: Cambridge University Press.

Kumbhakar, S.C., and C.F. Parmeter. 2010. Estimation of hedonic price functions with incomplete information. *Empirical Economics* 39 (1): 1–25.

Kumbhakar, S.C., and K. Sun. 2013. Derivation of marginal effects of determinants of technical inefficiency. *Economics Letters* 120 (2): 249–253.

Kumbhakar, S.C., S. Ghosh, and J.T. McGuckin. 1991. A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms. *Journal of Business & Economic Statistics* 9 (3): 279–286.

Kumbhakar, S.C., E.G. Tsionas, and T. Sipiläinen. 2009. Joint estimation of technology choice and technical efficiency: an application to organic and conventional dairy farming. *Journal of Productivity Analysis* 31 (3): 151–161.

Kumbhakar, S.C., C.F. Parmeter, and E.G. Tsionas. 2013. A zero inefficiency stochastic frontier model. *Journal of Econometrics* 172 (1): 66–76.

Kumbhakar, S.C., G. Lien, and J.B. Hardaker. 2014. Technical efficiency in competing panel data models: A study of Norwegian grain farming. *Journal of Productivity Analysis* 41 (2): 321–337.

Lai, H.-P. 2015. Maximum likelihood estimation of the stochastic frontier model with endogenous switching or sample selection. *Journal of Productivity Analysis* 43 (1): 105–117.

Lai, H.-P., and C.J. Huang. 2013. Maximum likelihood estimation of seemingly unrelated stochastic frontier regressions. *Journal of Productivity Analysis* 40 (1): 1–14.

Lai, H.-P., S.W. Polachek, and H.-J. Wang. 2009. Estimation of a stochastic frontier model with a sample selection problem. Working Paper, Department of Economics, National Chung Cheng University.

Lee, L.-F. 1983. A test for distributional assumptions for the stochastic frontier functions. *Journal of Econometrics* 22 (3): 245–267.

Lee, L.-F. 1993. Asymptotic distribution of the maximum likelihood estimator for a stochastic frontier function model with a singular information matrix. *Econometric Theory* 9 (3): 413–430.

Lee, Y.H. 1996. Tail truncated stochastic frontier models. *Journal of Economic Theory and Econometrics* 2: 137–152.

Lee, L.-F., and A. Chesher. 1986. Specification testing when score test statistics are identically zero. *Journal of Econometrics* 31 (2): 121–149.

Lee, S., and Y.H. Lee. 2014. Stochastic frontier models with threshold efficiency. *Journal of Productivity Analysis* 42 (1): 45–54.

Lee, Y.H., and P. Schmidt. 1993. A production frontier model with flexible temporal variation in technical efficiency. In *The measurement of productive efficiency: Techniques and applications*, ed. H.O. Fried, S.S. Schmidt, and C.A.K. Lovell, 237–255. Oxford: Oxford University Press.

Lee, L.-F., and W.G. Tyler. 1978. The stochastic frontier production function and average efficiency. *Journal of Econometrics* 7 (3): 385–389.

Leibenstein, H. 1966. Allocative efficiency vs. "X-efficiency". *The American Economic Review* 56(3), 392–415.

Leibenstein, H. 1975. Aspects of the X-efficiency theory of the firm. *The Bell Journal of Economics* 6 (2): 580–606.

Leibenstein, H. 1978. X-inefficiency Xists: Reply to an Xorcist. *The American Economic Review* 68 (1): 203–211.

Li, Q. 1996. Estimating a stochastic production frontier when the adjusted error is symmetric. *Economics Letters* 52 (3): 221–228.

Lukacs, E. and R.G. Laha. 1964. *Applications of characteristic functions.* London: Charles Griffin and Company.

Maddala, G.S. 1983. *Limited-dependent and qualitative variables in econometrics.* Cambridge: Cambridge University Press.

Marris, R.L. 1964. *The economic theory of managerial capitalism*, 1st ed. London: Macmillan.

Meeusen, W., and J. van Den Broeck. 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review* 18 (2): 435–444.

Migon, H.S., and E.V. Medici. 2001. Bayesian hierarchical models for stochastic production frontier. Working Paper, Universidade Federal do Rio de Janeiro.

Mundlak, Y. 1978. On the pooling of time series and cross section data. *Econometrica* 46 (1): 69–85.

Murillo-Zamorano, L.R. 2004. Economic efficiency and frontier techniques. *Journal of Economic Surveys* 18 (1): 33–77.

Nelsen, R.B. 2006. *An introduction to copulas*, 2nd ed. New York: Springer.

Nerlove, M. 1963. Returns to scale in electricity supply. In *Measurement in economics: Studies in mathematical economics and econometrics*, ed. C.F. Christ, M. Friedman, L.A. Goodman, Z. Griliches, A.C. Harberger, N. Liviatan, J. Mincer, Y. Mundlak, M. Nerlove, D. Patinkin, L.G. Telser, and H. Theil, 167–198. Stanford: Stanford University Press.

Nguyen, N. 2010. Estimation of technical efficiency in stochastic frontier analysis. PhD thesis, Bowling Green State University.

Oikawa, K. 2016. A microfoundation for stochastic frontier analysis. *Economics Letters* 139: 15–17.

Ondrich, J., and J. Ruggiero. 2001. Efficiency measurement in the stochastic frontier model. *European Journal of Operational Research* 129 (2): 434–442.

Orea, L., and S.C. Kumbhakar. 2004. Efficiency measurement using a latent class stochastic frontier model. *Empirical Economics* 29 (1): 169–183.

Pal, M. 2004. A note on a unified approach to the frontier production function models with correlated non-normal error components: The case of cross section data. *Indian Economic Review* 39 (1): 7–18.

Pal, M., and A. Sengupta. 1999. A model of FPF with correlated error components: An application to Indian agriculture. *Sankhyā: The Indian Journal of Statistics, Series B* 61 (2): 337–350.

Papadopoulos, A. 2015. The half-normal specification for the two-tier stochastic frontier model. *Journal of Productivity Analysis* 43 (2): 225–230.

Parmeter, C.F. 2018. Estimation of the two-tiered stochastic frontier model with the scaling property. *Journal of Productivity Analysis* 49 (1): 37–47.

Parmeter, C.F., and S.C. Kumbhakar. 2014. Efficiency analysis: A primer on recent advances. *Foundations and Trends in Econometrics* 7 (3–4): 191–385.

Pitt, M.M., and L.-F. Lee. 1981. The measurement and sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics* 9 (1): 43–64.

Polachek, S.W., and B.J. Yoon. 1987. A two-tiered earnings frontier estimation of employer and employee information in the labor market. *The Review of Economics and Statistics* 69 (2): 296–302.

Polachek, S.W., and B.J. Yoon. 1996. Panel estimates of a two-tiered earnings frontier. *Journal of Applied Econometrics* 11 (2): 169–178.

Qian, J., and R.C. Sickles. 2008. Stochastic frontiers with bounded inefficiency. Mimeo. Department of Economics, Rice University.

Reifschneider, D., and R. Stevenson. 1991. Systematic departures from the frontier: A framework for the analysis of firm inefficiency. *International Economic Review* 32 (3): 715–723.

Rho, S., and P. Schmidt. 2015. Are all firms inefficient? *Journal of Productivity Analysis* 43 (3): 327–349.

Richmond, J. 1974. Estimating the efficiency of production. *International Economic Review* 15 (2): 515–521.

Ritter, C., and L. Simar. 1997. Pitfalls of normal-gamma stochastic frontier models. *Journal of Productivity Analysis* 8 (2): 167–182.

Schmidt, P. 1976. On the statistical estimation of parametric frontier production functions. *The Review of Economics and Statistics* 58 (2): 238–239.

Schmidt, P., and C.A.K. Lovell. 1980. Estimating stochastic production and cost frontiers when technical and allocative inefficiency are correlated. *Journal of Econometrics* 13 (1): 83–100.

Schmidt, P., and R.C. Sickles. 1984. Production frontiers and panel data. *Journal of Business & Economic Statistics.* 2 (4): 367–374.

Self, S.G., and K.-Y. Liang. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 82 (398): 605–610.

Shephard, R.W. 1953. *Cost and production functions*, 1st ed. Princeton: Princeton University Press.

Simar, L., and P.W. Wilson. 2010. Inferences from cross-sectional, stochastic frontier models. *Econometric Reviews* 29 (1): 62–98.

Simar, L., C.A.K. Lovell, and P. Vanden Eeckaut. 1994. Stochastic frontiers incorporating exogenous influences on efficiency. STAT Discussion Papers no. 9403, Institut de Statistique, Université Catholique de Louvain.

Sipiläinen, T., and A. Oude Lansink. 2005. Learning in switching to organic farming. *NJF Report* 1 (1): 169–172.

Smith, M.D. 2008. Stochastic frontier models with dependent error components. *Econometrics Journal* 11 (1): 172–192.

Smith, A.S.J., and P. Wheat. 2012. Estimation of cost inefficiency in panel data models with firm specific and sub-company specific effects. *Journal of Productivity Analysis* 37 (1): 27–40.

Stead, A.D. 2017. Regulation and efficiency in UK public utilities. PhD thesis, University of Hull.

Stead, A.D., P. Wheat, and W.H. Greene. 2018. Estimating efficiency in the presence of extreme outliers: A logistic-half normal stochastic frontier model with application to highway maintenance costs in England. In *Productivity and inequality*, ed W.H. Greene, L. Khalaf, P. Makdissi, R.C. Sickles, M. Veall, and M. Voia, 1–19. Springer.

Stevenson, R.E. 1980. Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics* 13 (1): 57–66.

Stigler, G.J. 1976. The Xistence of X-efficiency. *The American Economic Review* 66 (1): 213–216.

Tancredi, A. 2002. Accounting for heavy tails in stochastic frontier models. Working Paper no. 2002.16, Department of Statistical Sciences, University of Padua.

Tchumtchoua, S., and D.K. Dey. 2007. Bayesian estimation of stochastic frontier models with multivariate skew *t* error terms. *Communications in Statistics - Theory and Methods* 36 (5): 907–916.

Terza, J.V. 2009. Parametric nonlinear regression with endogenous switching. *Econometric Reviews* 28 (6): 555–580.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58 (1): 267–288.

Timmer, C.P. 1971. Using a probabilistic frontier production function to measure technical efficiency. *Journal of Political Economy* 79 (4): 776–794.

Train, K.E. 2009. Discrete choice methods with simulation, 2nd ed. Cambridge: Cambridge University Press.

Tsagris, M., C. Beneki, and H. Hassani. 2014. On the folded normal distribution. *Mathematics* 2 (1): 12–28.

Tsionas, E.G. 2002. Stochastic frontier models with random coefficients. *Journal of Applied Econometrics* 17 (2): 127–147.

Tsionas, E.G. 2007. Efficiency measurement with the Weibull stochastic frontier. *Oxford Bulletin of Economics and Statistics* 69 (5): 693–706.

Tsionas, E.G. 2012. Maximum likelihood estimation of stochastic frontier models by the Fourier transform. *Journal of Econometrics* 170 (1): 234–248.

Tsionas, M.G. 2017. Microfoundations for stochastic frontiers. *European Journal of Operational Research* 258 (3): 1165–1170.

Tsionas, E.G., and S.C. Kumbhakar. 2014. Firm heterogeneity, persistent and transient technical inefficiency: a generalized true random-effects model. *Journal of Applied Econometrics* 29 (1): 110–132.

Waldman, D.M. 1984. Properties of technical efficiency estimators in the stochastic frontier model. *Journal of Econometrics* 25 (3): 353–364.

Wang, H.-J. 2002. Heteroscedasticity and Non-monotonic efficiency effects of a stochastic frontier model. *Journal of Productivity Analysis* 18 (3): 241–253.

Wang, J., and P. Schmidt. 2002. One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis* 18: 129–144.

Wang, W.S., and P. Schmidt. 2009. On the distribution of estimated technical efficiency in stochastic frontier models. *Journal of Econometrics* 148 (1): 36–45.

Wang, W.S., C. Amsler, and P. Schmidt. 2011. Goodness of fit tests in stochastic frontier models. *Journal of Productivity Analysis* 35 (2): 95–118.

Weinstein, M.A. 1964. The sum of values from a normal and a truncated normal distribution. *Technometrics* 6 (1): 104–105.

Wheat, P., and A. Smith. 2012. Is the choice of (t−T) in Battese and Coelli (1992) type stochastic frontier models innocuous? Observations and generalisations. *Economics Letters* 116 (3): 291–294.

Wheat, P., W. Greene, and A. Smith. 2014. Understanding prediction intervals for firm specific inefficiency scores from parametric stochastic frontier models. *Journal of Productivity Analysis* 42 (1): 55–65.

Wheat, P., A.D. Stead, and W.H. Greene. 2017. Allowing for outliers in stochastic frontier models: A mixture noise distribution approach. 15th European Workshop on Efficiency and Productivity Analysis, London, UK.

Wheat, P., A.D. Stead, and W.H. Greene. 2019. Robust stochastic frontier analysis: A student's t-half normal model with application to highway maintenance costs in England. *Journal of Productivity Analysis* 51 (1): 21–38. https://doi.org/10.1007/s11123-018-0541-y.

Williamson, O.E. 1963. Managerial discretion and business behavior. *The American Economic Review* 53 (5): 1032–1057.

Zellner, A. 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* 57 (298): 348–368.