INTERPRETING AND TESTING THE SCALING PROPERTY
IN MODELS WHERE INEFFICIENCY DEPENDS ON FIRM CHARACTERISTICS

Antonio Alvarez
Dpto. de Economia
Universidad de Oviedo
Avenida del Cristo s.n.
33071 Oviedo, Spain
Phone: (34) 985-104-859
FAX: (34) 985-104-871
alvarez@uniovi.es

Christine Amsler
Department of Economics
Michigan State University
East Lansing, MI 48824, USA
Phone: 517-355-3774
FAX: 517-432-1068
amsler@msu.edu

Luis Orea
Dpto. de Economia
Universidad de Oviedo
Avenida del Cristo s.n.
33071 Oviedo, Spain
Phone: (34) 985-106-243
FAX: (34) 985-104-871
lorea@uniovi.es

Peter Schmidt
Department of Economics
Michigan State University
East Lansing, MI 48824, USA
Phone: 517-355-8381
FAX: 517-432-1068
schmidtp@msu.edu

The authors gratefully acknowledge the financial support of the Spanish Ministry of Science and Technology.

ABSTRACT

Let $u \geq 0$ be technical inefficiency, let z be a set of variables that affect u, and let δ be the parameters of this relationship. The model satisfies the *scaling property* if $u(z,\delta)$ can be written as a scaling function $h(z, \delta)$ times a random variable u^* that does not depend on z. This property implies that changes in z affect the scale but not the shape of $u(z,\delta)$. This paper reviews the existing literature and identifies models that do and do not have the scaling property. It also discusses practical advantages of the scaling property. The paper shows how to test the hypothesis of scaling, and other interesting hypotheses, in the context of the model of Wang, *Journal of Productivity Analysis*, 2002. Finally, two empirical examples are given.

1. Introduction

In this paper, we are interested in a stochastic frontier model in which observable characteristics of the firms affect their levels of technical inefficiency. To be more precise, let $u \geq 0$ be the one-sided error reflecting technical inefficiency, and let z be a set of variables that affect u. Then we can write u as $u(z,\delta)$ to reflect its dependence on z and some parameters δ. Various models in the existing literature specify the distribution of $u(z,\delta)$. We will be interested in models that satisfy the *scaling property*, which says that $u(z,\delta)$ can be written as a scaling function $h(z, \delta)$ times a random variable u^* that does not depend on z. This property implies that changes in z affect the scale but not the shape of $u(z,\delta)$.

We discuss scaling in the context of the stochastic frontier model but it is also of relevance in the semiparametric (DEA) context. There is a large literature attempting to relate DEA efficiency scores to "environmental variables" z. Simar and Wilson (2003) give a survey of this literature, and they introduce a data generating process in which the (population) output efficiency is random and depends in a parametric way on some variables z. Their assumption about the way efficiency scores depend on z corresponds to the KGMHLBC model discussed below. They could instead have assumed a scaling model.

Some (but not all) of the models in the literature have the scaling property. However, there is really no previous systematic treatment of scaling as a unifying principle. (The article that comes closest is Simar, Lovell and van den Eeckaut (1994).) In this paper, we provide a comprehensive treatment of the scaling property, and a review of the relevant literature. We identify models in the literature that do and do not have this property, and we propose a specific model that may be empirically useful. We discuss the practical advantages of models with the scaling property. We also show how to test the scaling hypothesis, and other interesting hypotheses, in the context of the model of Wang (2002).

The paper also makes the following important observation, which is known in the econometric literature but appears to have been missed in the production frontier literature. Maximum likelihood estimates of models that assume independence of technical efficiency over time (such as our model, and most models that deal with determinants of inefficiency) remain consistent even if the independence assumption is false. However, the estimated variances of the parameters using the usual formulas that assume independence are incorrect. We show in this paper how they can be corrected in a simple way. This affects tests for scaling, as well as any other inference based on the estimated coefficients.

The plan of the paper is as follows. In section 2, we present our basic framework. In section 3, we review some of the existing literature and we identify models that do and do not have the scaling property. In section 4, we discuss the practical advantages of models with this property. We also discuss corrections for non-independence when the MLE is based on an incorrect assumption of independence. In section 5, we discuss tests of the hypothesis that the scaling property holds, and of other interesting hypotheses that allow us to distinguish between various competing models. In section 6, we give two empirical examples involving Spanish banks and Indian farms. Finally, our concluding remarks are in section 7.

2. The scaling property

Our basic setup and notation follows Wang and Schmidt (2002). We suppose that we have panel data, in which firms are indexed by $i = 1,\ldots,N$ and time is indexed by $t = 1,\ldots,T$. Let y_{it} be log output; let x_{it} be a vector of variables that affect the position of the frontier; and let z_{it} be a vector of variables that affect the magnitude of technical inefficiency. Generally the x_{it} are inputs and the z_{it} are either functions of inputs or measures of the environment in which the firm operates. The x_{it} and z_{it} can overlap. Because the z_{it} (like the x_{it}) are treated as "fixed," they cannot be functions of y_{it}.

Let $y_{it}^* \geq y_{it}$ be the unobserved frontier. The linear stochastic frontier model asserts that, conditional on x_{it} and z_{it}, y_{it}^* is distributed as $N(x_{it}'\beta, \sigma_v^2)$. Then we can write the frontier as:

(1) $$y_{it}^* = x_{it}'\beta + v_{it}$$

where v_{it} is distributed as $N(0,\sigma_v^2)$ and is independent of x_{it} and z_{it}. Finally, the model asserts that (conditional on x_{it}, z_{it} and y_{it}^*) the actual output level y_{it} equals y_{it}^* minus a one-sided error whose distribution depends only on z_{it}. Therefore we can write the model as:

(2) $$y_{it} = x_{it}'\beta + v_{it} - u_{it}(z_{it},\delta) \quad , \quad u_{it}(z_{it},\delta) \geq 0 \quad .$$

Here u_{it} and v_{it} are independent of each other and of x_{it}, and in addition v_{it} is independent of z_{it}.

We will say that the model has the *scaling property* if

(3) $$u_{it}(z_{it},\delta) = h(z_{it},\delta)\bullet u_{it}^* \quad ,$$

where $h(z_{it},\delta) \geq 0$, and where $u_{it}^* \geq 0$ has a distribution that does not depend on z_{it}. We will call $h(z_{it},\delta)$ the *scaling function* and u_{it}^* the *basic random variable*, while the distribution of u_{it}^* will be called the *basic distribution*.

The essential feature of the scaling property is the fact that changes in z_{it} change the scale but not the shape of the distribution of u_{it}. This is so because the shape is determined by the basic distribution, which does not depend on z_{it}, whereas the scaling function $h(z_{it},\delta)$ determines the scale. More precisely, suppose that u^* has density $f(u^*)$, and that $u = h\bullet u^*$ where h is treated as a constant. Then the density of u equals $(1/h)f(u/h)$. The sense in which this has the same shape as

$f(u^*)$ is that, with proper rescaling of the axes, the graphs of the two densities would be identical.

The scaling property can be viewed as a purely statistical matter, but it can also be given the following economic interpretation, which we find attractive. The basic random variable u_{it}^* can be seen as the firm's base efficiency level which captures things like the manager's natural skills, which we view as random. How well these natural skills are exploited to manage the firm efficiently depends on other variables z_{it}, which might include the manager's education or experience, or measures of the environment in which the firm operates, for example. So the ultimate level of efficiency depends on u_{it}^* and also on some function of z_{it}, $h(z_{it},\delta)$. The scaling property then corresponds to a multiplicative decomposition of $u_{it}(z_{it},\delta)$ into these two logically independent parts.

An alternative that has sometimes been proposed (Huang and Liu (1994), Battese and Coelli (1995), Simar and Wilson (2003)) is an additive decomposition of the form $u_{it}(z_{it},\delta) = h(z_{it},\delta) + \tau_{it}$. However, this can never actually be a decomposition into independent parts, because $u_{it}(z_{it},\delta) \geq 0$ requires $\tau_{it} \geq -h(z_{it},\delta)$.

3. Review of the literature

In the stochastic frontier literature, the first models with the scaling property appear to be models that were designed to allow time-varying inefficiency, as opposed to the effects of firm characteristics on inefficiency. Kumbhakar (1990) and Battese and Coelli (1992) suggested models of the form $u_{it} = h(t,\delta)\cdot u_i^*$. These models fit our framework above, as a special case corresponding to $z_{it} = t$ and $u_{it}^* = u_i^*$ for all t.

An example of a model that has the scaling property is the scaled half-normal model, or RSCFG model, of Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill, Ford and Gropper (1995). In this model it is assumed that u_{it} is distributed as $N(0, \sigma_{it}^2)^+$, where $\sigma_{it}(z_{it},\theta)$ depends in a specific way on z_{it} and some parameters θ. (Here and throughout the paper, superscript "+" indicates truncation of a random variable or distribution from the left at zero.) The RSCFG model has the scaling property because it is equivalent to say that the distribution of u_{it} is $N[0, \sigma(z_{it},\theta)^2]^+$, or to say that u_{it} is distributed as $\sigma(z_{it},\theta)\bullet N(0,1)^+$. The above papers make different suggestions for the function $\sigma(z_{it},\theta)$. Caudill, Ford and Gropper specify $\sigma_{it} = \sigma_u \bullet h(z_{it},\gamma)$ with the exponential scaling function $h(z_{it},\gamma) = \exp(z_{it}'\gamma)$, a choice that we will also use in this paper. So θ consists of σ_u and γ.

A detail that is important is that we need to be careful to specify only one constant. Here we have chosen to separate the constant (in this case, the standard deviation σ_u) rather than to incorporate it into the function $h(z_{it},\gamma)$. In the case of the exponential scaling function, this means that z_{it} does not contain an intercept. It would be equivalent to omit the constant σ_u and to introduce an intercept into z_{it}

Other distributions would also lead naturally to models with the scaling property. For example, if the u_{it} are exponentially distributed, with parameter $\lambda(z_{it},\theta)$, the model has the scaling property because an exponential distribution with parameter λ is the same as λ times an exponential distribution with parameter equal to one. This model has been considered by Simar, Lovell and van den Eeckaut (1994).

A well known and popular model that does not have the scaling property is the KGMHLBC model of Kumbhakar, Ghosh and McGuckin (1991), Huang and Liu (1994), and Battese and Coelli (1995). (This model has also recently been suggested for the parameterization of the output inefficiency measure in DEA, by Simar and Wilson (2003).) This is a truncated normal model in which the mean of the pre-truncation normal depends on z_{it} and some parameters θ. That is, u_{it} is distributed as $N(\mu_{it}, \sigma_u^2)^+$ where $\mu_{it} = \mu(z_{it},\theta)$. Since the degree of truncation varies with μ_{it}, the shape of the distribution changes when z_{it} changes. All three of the papers listed above suggest a linear specification of μ_{it}: $\mu_{it} = \alpha + z_{it}'\delta$. Our implementation of the model (reported later) will have $\mu_{it} = \mu \bullet \exp(z_{it}'\delta)$, with μ unrestricted (i.e., it can be positive or negative).

An alternative but equivalent presentation of the KGMHLBC model, followed by Huang and Liu (1994), Battese and Coelli (1995) and Simar and Wilson (2003), is to write $u_{it} = \mu_{it} + \tau_{it}$, where τ_{it} is $N(0, \sigma_\tau^2)$ truncated from the left at $-\mu_{it}$. As argued in Section 2, this decomposition is not into logically independent parts because we must have $\tau_{it} \geq -\mu_{it}$. In other words, the failure of scaling in this model is not just a result of a focus on a multiplicative decomposition.

In the RSCFG model, the expectation of u_{it} is monotonic in z_{it} so long as the specification for σ_{it} is monotonic in z_{it}. Similarly, in the KGMHLBC model, the expectation of u_{it} is monotonic in z_{it} (though the relationship is complicated) so long as the specification of μ_{it} is monotonic in z_{it}. Wang (2002) proposes a model in which the relationship of the expectation of u_{it} to z_{it} could be non-monotonic. He does this by assuming that the distribution of u_{it} is $N(\mu_{it}, \sigma_{it}^2)^+$, where both μ_{it} and σ_{it} depend on z_{it} and parameters. Specifically, he assumes that $\mu_{it} = z_{it}'\delta$ and $\sigma_{it}^2 = \exp(z_{it}'\gamma)$ where z_{it} contains intercept. Our implementation of this model will have $\mu_{it} = \mu \cdot \exp(z_{it}'\delta)$ and $\sigma_u \cdot \exp(z_{it}'\gamma)$. In either case, this model does not generally have the scaling property.

In Wang's model the z_{it} have two different coefficients. One set (δ) is for the mean and the other (γ) is for the variance of the pre-truncation normal. In the RSCFG model and the KGMHLBC model, the z_{it} have only one set of coefficients. If one wishes to restrict attention to models in which each of the z_{it} have only a single set of coefficients, scaling models may be attractive. In particular, a reasonable competitor to the RSCFG and KGMHLBC models would be the *scaled Stevenson model*, which is simply the scaled version of the truncated normal model of Stevenson (1980). In this model the distribution of u_{it} would be $h(z_{it},\delta)\bullet N(\mu,\sigma_u^2)^+$. Therefore the mean and standard deviation of the pre-truncation normal both depend on z_{it}: $\mu_{it} = \mu\bullet h(z_{it}, \delta)$ and $\sigma_{it} = \sigma_u\bullet h(z_{it},\delta)$. However, the degree of truncation depends on $\mu_{it}/\sigma_{it} = \mu/\sigma_u$, which does not depend on $z_{it.}$ Our implementation of this model will use the exponential scaling function $h(z_{it},\delta) = \exp(z_{it}'\delta)$. This model is the assumed data generating process in the simulations of Wang and Schmidt (2002), and it is discussed by Simar, Lovell and van den Eeckaut (1994).

There are not many general discussions of scaling. The earliest is apparently Simar, Lovell and van den Eeckaut (1994). (See also the extensive summary of this article in Kumbhakar and Lovell (2000, section 7.3).) They consider several different basic distributions, and they also discuss nonlinear least squares estimation without any distributional assumption. Wang and Schmidt (2002) also discuss scaling as a general principle upon which to base models, and they list some advantages of the scaling property, which we will discuss further in the next section.

4. Advantages of the scaling property

So far as we are aware, no theory indicates that the scaling property should hold, and it is ultimately an empirical question whether or not models with this property are useful. However, the scaling property has some features that we find attractive, and which we will now discuss.

1. The defining feature of models with the scaling property is that firms differ in their mean efficiencies, but not in the shape of the distribution of inefficiency. We find this intuitively appealing (though, of course, others may not).

2. The question of whether the effects of the z_{it} on efficiency are monotonic can be handled

easily by the choice of scaling function. If one wishes to impose monotonicity, simply use a monotonic scaling function, such as the exponential scaling function $\exp(z_{it}'\delta)$. If not, use a non-monotonic scaling function.

3. As noted by Simar, Lovell and van den Eeckaut (1994) and Wang and Schmidt (2002), at least some portions of the model can be estimated by non-linear least squares (NLLS), without making a distributional assumption on the basic random variable u_{it}^*. More specifically, if we define $\mu^* = E(u_{it}^*)$, then taking expectations in (2) yields:

(4) $$E(y_{it} \mid x_{it}, z_{it}) = x_{it}'\beta - \mu^* \cdot h(z_{it},\delta) .$$

We can then obtain consistent estimates of the parameters in (4) by NLLS. If we define $\sigma_{u^*}^2 = \mathrm{var}(u_{it}^*)$, then we can also note that

(5) $$\mathrm{Var}(y_{it} \mid x_{it}, z_{it}) = \sigma_v^2 + \sigma_{u^*}^2 \cdot h(z_{it},\delta)^2$$

so that the error in the regression indicated by (4) is heteroskedastic. Therefore generalized NLLS would be needed to obtain estimates that are efficient (in the class of estimates that do not impose distributional assumptions).

4. As noted by Wang and Schmidt (2002), the interpretation of δ does not depend on the distribution of inefficiency, and simple scaling functions yield simple expressions for the effect of the z_{it} on mean efficiency. For example, if we use the exponential scaling function, so that $u_{it} = \exp(z_{it}'\delta) \bullet u_{it}^*$, then $\delta = \partial \ln(u_{it})/\partial z_{it}$, and the coefficients δ are just the derivatives of log inefficiency with respect to the z_{it}. By contrast, in the KGMHLBC model or Wang's model the expression for the effect of z_{it} on u_{it} or $E(u_{it})$ is complicated, and depends on features of the truncated normal distribution. See, e.g., Wang (2002, p. 244).

5. An important but currently unsolved technical question is how to allow correlation over

time in the inefficiencies u_{it}. The Battese-Coelli (1995) and Wang (2002) models assume that the u_{it} are independent (conditional on the z_{it}) over time. This is widely recognized as an unrealistic assumption, but it is not clear how to relax it. However, under scaling we have the possibility of the following alternative model:

(6) $$u_{it} = h(z_{it},\delta) \bullet u_i^*$$

where u_i^* is time-invariant. This model is a slight generalization of the models of Kumbhakar (1990) and Battese and Coelli (1992).

A related point that seems not to be recognized in the stochastic frontier literature is that maximum likelihood estimates based on the assumption of independent observations are consistent even if the observations are not independent, so long as the (marginal) distribution of each observation is correctly specified. Thus, for example, estimates of the Battese-Coelli (1995) model will be consistent even if the u_{it}^* are not independent over time, so long as the model is otherwise correctly specified. However, the estimated variances (or standard errors) of the estimated parameters, calculated under the assumption of independence, will not be correct if independence does not hold. It is possible to calculate asymptotically valid "corrected" estimated variances that allow for non-independence of unspecified form. These points are known in the econometric literature. For example, see Hayashi (2000), section 8.7. Some details are given in the Appendix.

6. Suppose we wish to test the adequacy of the half-normal model against the alternative of a truncated normal with non-zero mean (Stevenson (1980)). There are some technical difficulties in this problem that can be avoided under scaling. This point is developed further in the next section of the paper.

5. Testing scaling and other interesting hypotheses

In this section, we will discuss tests of the scaling hypothesis, as well as some other hypotheses of interest. We assume that u_{it} is distributed as $N(\mu_{it}, \sigma_{it}^2)^+$, where $\mu_{it} = \mu \bullet \exp(z_{it}'\delta)$ and $\sigma_{it} = \sigma_u \bullet \exp(z_{it}'\gamma)$, with $\sigma_u > 0$ but μ unrestricted. Conditional on the z_{it}, the u_{it} are independent over both i and t. This is the same as the model presented in Wang (2002) except for the specification of μ_{it}, which he took to be linear.

Various special cases of this general model are of interest, and we can test the restrictions that lead to these special cases. This point has been made by Wang (2002, 2003), though his list of special cases is not the same as ours. To address the testing question in some generality, let $\xi = 0$ denote the hypothesis of interest. (For example, the scaling hypothesis corresponds to $\delta = \gamma$, and so $\xi = \delta - \gamma$.) We will consider the following three standard testing principles, all of which are typically based on the results of maximum likelihood estimation. **(i) Likelihood ratio test.** Let ln L be generic notation for the logarithm of the likelihood function. Suppose that L_U is the maximized value of the likelihood when the model is estimated ignoring the restriction (i.e. U is for unrestricted), and let L_R be the maximized value of the likelihood when the mode is estimated with the restriction imposed (i.e. R is for restricted). Then the test statistic is $LR = 2(\ln L_U - \ln L_R)$. The statistic is asymptotically (for large N) distributed as χ_p^2, where p is the number of restrictions being tested. **(ii) Wald test.** Let $\hat{\xi}$ be the unrestricted estimate of ξ and let $V(\hat{\xi})$ be the asymptotic variance-covariance matrix of this estimate. Then the test statistic is $\hat{\xi}'V(\hat{\xi})^{-1}\hat{\xi}$, and the statistic is also asymptotically distributed as χ_p^2. **(iii) LM test.** Let θ be the vector of all of the parameters in the model, and let $\tilde{\theta}$ be the restricted maximum likelihood estimate. (For example, in Wang's model, θ contains β, μ, σ_v, σ_u, δ and γ; or, equivalently, β, μ, σ_u, δ and ξ, where $\xi = \delta -$

γ. The scaling hypothesis is $\xi = 0$, which means that the restricted estimates are $\tilde{\beta}$, $\tilde{\mu}$, $\tilde{\sigma}_v$, $\tilde{\sigma}_u$, $\tilde{\delta}$ and $\tilde{\xi}$, where $\tilde{\xi} = 0$ and the rest of $\tilde{\theta}$ consists of the other restricted estimates.) Let $I(\tilde{\theta})$ be the information matrix evaluated at $\theta = \tilde{\theta}$, and let $[\partial \ln L(\tilde{\theta})/\partial\theta]$ be the vector of partial derivatives of the logarithm of the likelihood function, evaluated at $\theta = \tilde{\theta}$. (The vector of partial derivatives is available from the authors on request.) Then the test statistic is LM = $[\partial \ln L(\tilde{\theta})/\partial\theta]'$ $I(\tilde{\theta})^{-1}[\partial \ln L(\tilde{\theta})/\partial\theta]$, and once again the asymptotic distribution is χ_p^2.

In the previous section we mentioned the possibility of allowing for non-independence of technical inefficiency over time by using "corrected" estimated variances for the parameters, as discussed further in the Appendix. The Wald test easily accommodates this possibility since one can simply take the asymptotic variance-covariance matrix $V(\hat{\xi})$ to be the "corrected" matrix that is valid under non-independence. We will refer to the Wald test using this corrected variance-covariance matrix as the Robust Wald test. The LR and LM tests are not subject to such easy correction.

We now consider various specific interesting restrictions (hypotheses to test). Each of these corresponds to a simplified version of the general model in which u_{it} is distributed as $N(\mu_{it}, \sigma_{it}^2)^+$, with $\mu_{it} = \mu \cdot \exp(z_{it}'\delta)$ and $\sigma_{it} = \sigma_u \cdot \exp(z_{it}'\gamma)$. Many of these simpler models appear previously in the literature.

A. $\delta = \gamma$. As noted above, this is the scaling hypothesis. When $\delta = \gamma$, the distribution of u_{it} is $N[\mu \bullet \exp(z_{it}'\delta), \sigma_u^2 \bullet (\exp(z_{it}'\delta))^2]^+$, or equivalently $\exp(z_{it}'\delta) \bullet N(\mu, \sigma_u^2)^+$, which is the scaled Stevenson model. The scaled Stevenson model obviously has the scaling property.

B. $\gamma = 0$. In this case the distribution of u_{it} becomes $N(\mu \bullet \exp(z_{it}'\delta), \sigma_u^2)^+$, which is a version of the KGMHLBC model. Wang (2002, 2003) tested essentially the same hypothesis. This model does not satisfy the scaling property (unless $\mu = 0$).

C. $\delta = 0$. In this case the distribution of u_{it} becomes $N(\mu, \sigma_u^2 \bullet (\exp(z_{it}'\gamma))^2)^+$, a model that does not appear (so far as we are aware) in the literature, at least with μ unrestricted. With $\mu = 0$ this would be the RSCFG model. Therefore we will call this model the RSCFG-μ model. Unless $\mu = 0$, it does not satisfy the scaling property. However, it is important to note that the hypothesis $\delta = 0$ can only be tested in a meaningful sense if μ is not equal to zero. The reason is that if $\mu = 0$, then δ is not identified, since then $\mu \bullet \exp(z_{it}'\delta) = 0$ for any value of δ.

D. $\delta = \gamma = 0$. In this case the distribution of u_{it} is $N(\mu, \sigma_u^2)^+$, which is the model of Stevenson (1980). It does not contain any variables (z_{it}) that influence the distribution of inefficiency, so the question of scaling does not arise.

E. $\mu = 0$. This is the restriction that yields the RSCFG model, since now the distribution of u_{it} becomes $N[0, \sigma_u^2 \bullet (\exp(z_{it}'\gamma))^2]^+$, or equivalently $\exp(z_{it}'\gamma) \bullet N(0, \sigma_u^2)^+$, which is the RSCFG model. An interesting and relevant observation (which also appears to be original) is that, in the context of the Wang model, this is a non-standard test. The reason is that the "nuisance parameter" δ is not identified under the null hypothesis ($\mu = 0$). As a result the asymptotic distribution of the likelihood ratio, Wald or LM test would not be chi-squared. See Hansen (1996) for a general discussion of this phenomenon. We can solve this problem by making some specific assumption about δ that identifies it even when $\mu = 0$. So, for example, we could assert that $\delta = 0$, in which case the test of $\mu = 0$ becomes a standard (asymptotically chi-squared) test. In this case we are simply testing whether $\mu = 0$ in the RSCFG-μ model. We suggest that another useful mechanism for identifying δ under the null is to assert that the scaling hypothesis is correct, so that $\delta = \gamma$. Then we are simply testing whether $\mu = 0$ in the scaled Stevenson model, and this is a standard (asymptotically chi-squared) test. Naturally the test of $\mu = 0$, that is of the adequacy of the RSCFG model, may differ depending on whether the alternative is the RSCFG-μ model or the scaled Stevenson model. What we cannot do (at least as a standard test) is test the joint hypothesis that $\mu = 0$ and $\delta = 0$.

F. $\mu = 0, \gamma = 0$. These are the restrictions that yield the Aigner, Lovell and Schmidt (1977) (ALS) model, since now u_{it} is distributed as $N(0, \sigma_u^2)^+$. For the reasons given in the preceding paragraph, this is a non-standard test in the context of Wang's model. It becomes a standard test under an identifying restriction on δ. Alternatively, we can obtain the ALS model as the special case of the RSCFG model corresponding to $\gamma = 0$. That would be a standard test. Like Stevenson's model, the ALS model does not contain any variables that affect the distribution of inefficiency, and so the question of scaling does not arise.

Table 1 gives a brief summary of these restrictions and the resulting models.

6. Empirical examples

In this section we give two empirical examples, involving previously-analyzed data sets on Spanish savings banks and Indian farms.

6.1 Spanish savings banks

We estimate a stochastic cost frontier using data from N = 118 Spanish banks for T = 7 years (1992-1998). These data were previously analyzed by Cuesta and Orea (2002) and Han, Orea and Schmidt (2004), to which the reader is referred for detail. However, those articles use somewhat different specifications than the specification that we use here.

We have three outputs: loans to firms and households; other financial assets; and non-interest income. We have two input prices: the loanable funds price, defined as the ratio of interest expenses to the total amount of deposits and other loanable funds; and an operational inputs price, defined as the ratio of total operating expenses to the total number of employees. In order to impose linear homogeneity in input prices, we divide total cost and total loanable funds price by the operational inputs price. We also include a linear time trend in the specification, and an intercept. We use a translog specification, so there are 21 explanatory variables (5 linear terms, 5 quadratic terms, 10 cross product terms, and the intercept).

We have five z_{it} variables that parameterize the inefficiency component: TREND, a time trend; MERGERS, which starts at zero and increases by one every time the bank acquires another bank; SIZE, measured by total assets; NBL, the ratio of non-bank loans to total assets; and DEPOSITS, the ratio of deposits to total assets.

We estimate six different models, as follows. 1. Wang's model, with δ and γ both estimated. 2. Stevenson's model ($\delta = \gamma = 0$). 3. The scaled Stevenson model ($\delta = \gamma$). This model is estimated by MLE and by NLLS. 4. The KGMHLBC model ($\gamma = 0$). 5. The RSCFG-μ model ($\delta = 0$). 6. The RSCFG model ($\mu = 0$). In all cases u_{it} is truncated normal with $\mu_{it} = \exp(z_{it}'\delta)$ and $\sigma_{it} = \exp(z_{it}'\gamma)$.

In Table 2 we report the estimates of δ and γ. (To save space we do not report the parameter estimates (β's) for the frontier.) We also report their standard errors, based on the usual MLE formula; the standard errors that are robust to autocorrelation, as discussed in the Appendix; and the maximized log likelihood values.

For the scaled Stevenson model, the KGMKLBC model and the RSCFG model, the effects of the z_{it} variables are known to be monotonic, so that a positive coefficient means that an increase in that variable increases mean inefficiency, and conversely. For each of the z_{it} variables, its coefficient has the same sign in each of these three models, so in that sense the results are qualitatively similar across models. Wang's model is not directly comparable in this respect because the effects of the z_{it} on mean inefficiency can be non-monotonic. The level of statistical significance of the individual coefficients varies across coefficients and models, and depends on whether the ordinary or robust standard errors are used. The sets of coefficients are always jointly significant at any reasonable level, in the sense that the restrictions that would reduce any of these models to the Stevenson model (in which no z_{it} appear) are always clearly rejected. For example, for Wang's model, the restriction would be that $\delta = \gamma = 0$, and this yields the following statistics: Likelihood ratio (LR), 686.0; Wald, 169.3; Robust Wald, 107.8. These are significant at any reasonable level (e.g. the 99% critical value of a chi-squared with 10 degrees of freedom is 23.2).

In fact, the restrictions that would reduce Wang's model into any of the simpler models are also always rejected. (1) If we test the scaling hypothesis that $\delta = \gamma$, we have LR = 252.5, Wald = 96.0, and Robust Wald = 60.9. (2) If we test $\gamma = 0$, which is a test of the adequacy of the KGMHLBC model, we have LR = 363.2, Wald = 82.2, Robust Wald = 49.1. (3) If we test $\delta = 0$, which is a test of the adequacy of the RSCFG-μ model, we have LR = 115.2, Wald = 80.5, and Robust Wald = 33.6. All of these values are significant at any reasonable level (e.g. the 99% critical value of a chi-squared with five degrees of freedom is 15.1).

To test the adequacy of the RSCFG model, we test $\mu = 0$. As discussed above, δ is not identified when $\mu = 0$ and so we need to specify something about δ to conduct the test. (1) If we test $\mu = 0$ while asserting that $\delta = \gamma$, we are interested in the significance of μ in the scaled Stevenson model. Here the LR statistic equals 255.1, the Wald statistic equals 18.1, and the

Robust Wald statistic equals 4.29. (These are chi-squared statistics with one degree of freedom. For the Wald statistics they are just the square of the asymptotic t-statistic given by the coefficient divided by its standard error. For the LR statistic, we compare the likelihood values of the RSCFG and scaled Stevenson model.) The Robust Wald statistic is significant at about the 95% level, whereas the other two are significant at any reasonable level. (2) If we test $\mu = 0$ while asserting that $\delta = 0$, we are interested in the significance of μ in the RSCFG-μ model. Now we have LR = 392.5, Wald = 69.5, and Robust Wald = 137.8, all of which are significant at any reasonable level. Thus the RSCFG model also appears to be rejected by the data, though the example indicates the ambiguity in the test due to the need to pick the model against which it is tested (i.e. to maintain an assumption that identifies δ).

Clearly the data favor Wang's model over any of its simpler competitors, including the scaled Stevenson model. It is also interesting to ask how these simpler models compare to each other. The scaled Stevenson model, the KGMHLBC model and the RSCFG-μ model have the same number of parameters. These models are non-nested but we can legitimately compare likelihoods because they have the same number of parameters. The RSCFG-μ model is clearly favored by the data among this set of models because its likelihood value is considerably higher than for the other models in the set. The RSCFG model has one less parameter than these other models, and it is nested in both the RSCFG-μ model and the scaled Stevenson model, as the special case that $\mu = 0$. This restriction is decisively rejected by the data. So, if we were not considering Wang's model, we would conclude that the RSCFG-μ model is preferred.

6.2 Indian farms

Next we analyze the data used by Wang (2002), which was previously used by Battese and

Coelli (1995). This is an unbalanced panel of N = 34 Indian farmers, observed for a maximum of 10 years. The actual number of observations is 271. The model is a production frontier with five inputs plus a time trend. There are three z_{it} variables: AGE, the age of the farmer; SCHOOL, the number of years of schooling of the farmer; and YEAR, a time trend. More precise definitions of these variables and additional detail about the data are given in Wang (2002).

We estimate the same set of six models as for the previous data set. The results are in Table 3, which has essentially the same format as Table 2. We note that the variables AGE and SCHOOL individually do not generally have very high levels of statistical significance, whereas YEAR does achieve significance in most cases.

We now test to the restrictions that would reduce Wang's model to the simpler specifications. (1) We can reject Stevenson's model ($\delta = \gamma = 0$). We have LR = 15.3, Wald = 27.1 and Robust Wald = 44.2, whereas the 95% critical value of a chi-squared with six degrees of freedom is 12.6 and the 99% critical value is 16.8. (2) We can also reject the KGMHLBC model, with LR = 9.85, Wald = 16.2 and Robust Wald = 20.0, whereas the 95% critical value of a chi-squared with three degrees of freedom is 7.81 and the 99% critical value is 11.3. (3) If we test $\delta = 0$, which is a test of the adequacy of the RSCFG-μ model, we have LR = 5.56, Wald = 4.51 and Robust Wald = 25.7, so we would reject the hypothesis using the Robust Wald test but accept it using the other two tests. (4) If we test $\delta = \gamma$, which is the scaling hypothesis, we have LR = 5.87, Wald = 4.46 and Robust Wald = 28.6, so again we would reject the hypothesis using the Robust Wald test but accept it using the other two tests.

If we use the Robust Wald tests, therefore, we are simply left with the Wang model. If we use the LR or Wald tests, we can simplify to either the scaled Stevenson model or the RSCFG-μ

model. In either case it is interesting to ask whether we can simplify further to the RSCFG model. It turns out that we can. The estimate of μ is very insignificant in the scaled Stevenson model, which means that we cannot reject the hypothesis that $\mu = 0$ while maintaining $\delta = \gamma$. The estimate of μ is also very insignificant in the RSCFG-μ model, so that we cannot reject the hypothesis that $\mu = 0$ while maintaining $\delta = 0$. Thus the RSCFG model is not rejected by the data. Note that this is a scaling model.

5. Concluding remarks

This paper did four rather different things. First, it gave a systematic discussion of the scaling property and identified its advantages. Basically these are advantages of convenience. Second, it identified some new models that may be empirically useful, notably the scaled Stevenson model. Third, it showed how the scaling property, and other interesting hypotheses, can be tested. Wang (2003) noted the prospect of some of these tests, but this paper provides some important additional technical detail, especially on testing the adequacy of the RSCFG model. Fourth, it pointed out that maximizing a likelihood based on an incorrect assumption of independence over time still leads to consistent estimates, but the standard errors of the estimates need to be corrected. This point is known in the econometric literature but its relevance to the frontiers literature seems to have previously not been appreciated.

An important topic for further research would be to construct tests of the scaling property that do not hinge on a specific distributional assumption for technical inefficiency. This should be possible in principle.

APPENDIX

Maximum Likelihood with Non-Independent Observations

Let the observed data be w_{it}, i = 1,…,N, t = 1,…,T, and suppose that the density of w_{it} is $f(w_{it},\theta)$. We assume random sampling over i (that is, $(w_{i1}, w_{i2}, \ldots, w_{iT})$ is independent and identically distributed over i) but we do not assume independence over different values of t. We define the "quasi-likelihood" as

(A1) $\ln L = \Sigma_i \Sigma_t \ln f(w_{it},\theta)$.

If we have independence over t, then L is the joint density of $(w_{i1}, w_{i2}, \ldots, w_{iT})$ and we would call ln L the likelihood. However, if we do not have independence over t, then L is not the joint density of $(w_{i1}, w_{i2}, \ldots, w_{iT})$ and the phrase quasi-likelihood is used. We will use the notation $\hat{\theta}$ to represent the "quasi-MLE," which is the value that maximizes ln L in (A1), and θ_0 will represent the true population value of θ.

We define the following notation:

(A2) $S_{it}(\theta) = \partial \ln f(w_{it},\theta)/ \partial \theta \quad , \quad S_i(\theta) = \Sigma_t S_{it}(\theta)$

(A3) $H_{it}(\theta) = \partial S_{it}(\theta)/ \partial \theta' = \partial^2 \ln f(w_{it},\theta)/ \partial \theta \ \partial \theta' \quad , \quad H_i(\theta) = \Sigma_t H_{it}(\theta)$

(A4) $C_i(\theta) = S_i(\theta)S_i(\theta)' = \Sigma_t \Sigma_r S_{it}(\theta) S_{ir}(\theta)'$

(A5) $H = E\, H_i(\theta_0) \quad , \quad C = E\, C_i(\theta_0)$

The quasi-MLE solves the equation: $\Sigma_i S_i(\hat{\theta}) = 0$. The key to understanding its properties is to view it as a generalized method of moments (GMM) estimator based on the moment condition:

(A6) $E\, S_i(\theta_0) = 0$.

The validity of this moment condition depends only on correct specification for each t separately, in the sense that $f(w_{it},\theta_0)$ is the density of w_{it}. This implies that $E\, S_{it}(\theta_0) = 0$ for all t, and thus (A6) holds, whether or not there is correlation over time. Therefore the GMM estimator based on (A6), which is the quasi-MLE, is consistent if we have correct specification, even if there is correlation over time. Standard GMM results indicate that the asymptotic variance matrix of $\hat{\theta}$ is given by

(A7) $$V(\hat{\theta}) = N^{-1} H^{-1} C H^{-1} \ .$$

This is the "sandwich form" expression with C sandwiched between the H^{-1} terms. Now let H and C be estimated by their sample equivalents:

(A8) $$\hat{H} = N^{-1} \Sigma_i H_i(\hat{\theta}) \ , \ \hat{C} = N^{-1} \Sigma_i C_i(\hat{\theta}) = N^{-1} \Sigma_i \Sigma_t \Sigma_r S_{it}(\hat{\theta}) S_{ir}(\hat{\theta})' \ .$$

Then we can estimate $V(\hat{\theta})$ consistently by

(A9) $$\hat{V}(\hat{\theta}) = N^{-1} \hat{H}^{-1} \hat{C} \hat{H}^{-1} = [\Sigma_i H_i(\hat{\theta})]^{-1} [\Sigma_i C_i(\hat{\theta})] [\Sigma_i H_i(\hat{\theta})]^{-1} .$$

(The N^{-1} terms have cancelled.) This is our "corrected" asymptotic variance formula.

The sense in which this is a "corrected" formula is the following. The standard (uncorrected) formula is the one that would assume independence over time. In this case we would have $E\, S_{it}(\theta_0) S_{ir}(\theta_0)' = 0$ when t is not equal to r. Correspondingly the double sum in the definition of $\hat{C}$ in (A8) would reduce to a single sum:

(A10) $$\hat{C} = N^{-1} \Sigma_i \Sigma_t S_{it}(\hat{\theta}) S_{it}(\hat{\theta})' \ .$$

(It is also the case that, under independence, H = C, so that the term $H^{-1} C H^{-1}$ in (A7) can be reduced to either H^{-1} or C^{-1}. However, it is still not the case that $\hat{H} = \hat{C}$. Therefore the "sandwich form" estimator as in (A9) is still recommended; just the form of $\hat{C}$ simplifies.)

The most plausible departures from independence would involve positive correlations over t, for a given i. For example, we expect technical inefficiency to correlate positively over time, because firms that are very inefficient (relative to other firms) this time period will probably also be very inefficient in other time periods. So we expect S_{it} to correlate positively with S_{ir} (with r not equal to t). If so, $\hat{C}$ in (A8) will tend to be larger than $\hat{C}$ in (A10), so that the "corrected" asymptotic variances will tend to be larger than the uncorrected ones. Another way to say this is that, if there is correlation over time, the (uncorrected) estimated variances one obtains under the assumption of independence will be too small.

REFERENCES

Aigner, D. J., C.A.K. Lovell and P. Schmidt (1977), “Formulation and Estimation of Stochastic Frontier Production Functions,” *Journal of Econometrics*, 6, 21-37.

Battese, G.E. and T.J. Coelli (1988), "Prediction of Firm-level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data," *Journal of Econometrics*, 38, 387-399.

Battese, G.E. and T.J. Coelli (1992), “Frontier Production Functions, Technical Efficiency and Panel Data: With Applications to Paddy Farmers in India,” *Journal of Productivity Analysis*, 3, 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, 325-332.

Caudill, S.B. and J.M. Ford (1993), "Biases in Frontier Estimation Due to Heteroskedasticity," *Economics Letters*, 41, 17-20.

Caudill, S.B., J.M. Ford and D.M. Gropper (1995), "Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroskedasticity," *Journal of Business and Economic Statistics*, 13, 105-111.

Cuesta, R.A. and L. Orea (2002), “Mergers and Technical Efficiency in Spanish Savings Banks: A Stochastic Distance Function Approach,” *Journal of Banking and Finance*, 26, 2231-2247.

Hansen, B.E. (1996), “Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis,” *Econometrica*, 64, 413-430.

Hayashi, F. (2000), *Econometrics*, Princeton University Press.

Huang, C.J. and J.-T. Liu (1994), "Estimation of a Non-Neutral Stochastic Frontier Production Function," *Journal of Productivity Analysis*, 5, 171-180.

Kumbhakar, S.C. (1990), “Production Frontiers, Panel Data, and Time-Varying Technical Efficiency,” *Journal of Econometrics*, 46, 201-211.

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 and Economic Statistics*, 9, 279-286.

Kumbhakar, S.C. and C.A.K. Lovell (2000), *Stochastic Frontier Analysis*, Cambridge University Press.

Reifschneider, D. and R. Stevenson (1991), "Systematic Departures from the Frontier: A Framework for the Analysis of Firm Inefficiency," *International Economic Review*, 32, 715-723.

Simar, L., C.A.K. Lovell and P. van den Eeckaut (1994), "Stochastic Frontiers Incorporating Exogenous Influences on Efficiency," Discussion Paper No. 9403, Institut de Statistique, Université Catholique de Louvain.

Simar, L, and P. Wilson (2003), "Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes," Discussion Paper No. 0307, Institut de Statistique, Université Catholique de Louvain.

Stevenson, R.E. (1980), "Likelihood Functions for Generalized Stochastic Frontier Estimation," *Journal of Econometrics*. 13, 57-66.

Wang, H.-J. (2002), "Heteroscedasticity and Non-Monotonic Efficiency Effects of a Stochastic Frontier Model," *Journal of Productivity Analysis*, 18, 241-253.

Wang, H.-J. (2003), "A Stochastic Frontier Analysis of Financing Constraints on Investment: The Case of Financial Liberalization in Taiwan," *Journal of Business and Economic Statistics*, 21, 406-419.

Wang, H.-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.

TABLE 1

List of Models

Model	Restrictions	$N(\mu_{it}, \sigma_{it}^2)^+$	
		Mean	**Std. Deviation**
General Model (Wang)		$\mu_{it} = \mu \cdot \exp(z_{it}'\delta)$	$\sigma_{it} = \sigma_u \cdot \exp(z_{it}'\gamma)$
RSCFG Model	$\mu = 0$	$\mu_{it} = 0$	$\sigma_{it} = \sigma_u \cdot \exp(z_{it}'\gamma)$
KGMHLBC Model	$\gamma = 0$	$\mu_{it} = \mu \cdot \exp(z_{it}'\delta)$	$\sigma_{it} = \sigma_u$
Scaled Stevenson Model	$\delta = \gamma$	$\mu_{it} = \mu \cdot \exp(z_{it}'\delta)$	$\sigma_{it} = \sigma_u \cdot \exp(z_{it}'\delta)$
RSCFG-μ Model	$\delta = 0$	$\mu_{it} = \mu$	$\sigma_{it} = \sigma_u \cdot \exp(z_{it}'\gamma)$
Stevenson Model	$\delta = \gamma = 0$	$\mu_{it} = \mu$	$\sigma_{it} = \sigma_u$
ALS Model	$\mu = \gamma = 0$	$\mu_{it} = 0$	$\sigma_{it} = \sigma_u$

TABLE 2
Results for Spanish Banks

		Model 1 (Wang)			Model 2 (Stevenson)			Model 3 (Scaled Stevenson)				
								MLE			NLLS	
		Coeff.	s.e.	Robust s.e.	Coeff.	s.e.	Robust s.e.	Coeff.	s.e.	Robust s.e.	Coeff.	s.e.
	TREND	-0.002	0.021	0.018				0.027	0.011	0.015	0.032	0.084
	MERGERS	0.122	0.032	0.069				0.018	0.014	0.030	0.100	0.020
δ	SIZE	-0.217	0.031	0.051				-0.221	0.025	0.051	-0.276	0.066
	NBL	0.072	0.138	0.197				-0.239	0.063	0.150	-0.048	0.038
	DEPOSITS	0.485	0.080	0.130				0.259	0.052	0.098	0.986	0.122
	TREND	0.016	0.049	0.103				0.027	0.011	0.015	0.032	0.084
	MERGERS	0.268	0.448	0.689				0.018	0.014	0.030	0.100	0.020
γ	SIZE	-0.409	0.076	0.164				-0.221	0.025	0.051	-0.276	0.066
	NBL	-0.936	0.357	0.735				-0.239	0.063	0.150	-0.048	0.038
	DEPOSITS	-2.464	0.342	0.636				0.259	0.052	0.098	0.986	0.122
	σ	0.286	0.064	0.117	0.128	0.003	0.011	0.087	0.004	0.008		
	μ	0.392	0.069	0.088	0.547	4.726	0.075	1.090	0.256	0.507	0.163	0.293
	lnL		867.160			524.168			740.897			

		Model 4 (KGMHLBC)			Model 5 (RSCFG-μ)			Model 6 (RSCFG)		
		Coeff.	s.e.	Robust s.e.	Coeff.	s.e.	Robust s.e.	Coeff.	s.e.	Robust s.e.
	TREND	0.036	0.023	0.045						
	MERGERS	0.092	0.068	0.175						
δ	SIZE	-0.286	0.037	0.092						
	NBL	-0.094	0.149	0.497						
	DEPOSITS	0.953	0.188	0.486						
	TREND				0.031	0.032	0.046	0.047	0.044	0.076
	MERGERS				0.577	0.342	0.801	0.702	0.350	0.589
γ	SIZE				-0.408	0.060	0.106	-0.717	0.066	0.093
	NBL				-0.012	0.288	0.467	-1.230	0.357	0.805
	DEPOSITS				-2.869	0.283	0.532	2.861	0.388	0.397
	σ	0.106	0.003	0.008	0.291	0.049	0.079	0.101	0.003	0.009
	μ	0.172	0.060	0.177	1.208	0.145	0.103			
	lnL		685.582			809.584			613.342	

TABLE 3
Results for Indian Farms

		Model 1 (Wang)			Model 2 (Stevenson)			Model 3 (Scaled Stevenson)				
								MLE			NLLS	
		Coeff.	s.e.	Robust s.e.	Coeff.	s.e.	Robust s.e.	Coeff.	s.e.	Robust s.e.	Coeff.	s.e.
	AGE	-0.078	0.042	0.020				-0.004	0.005	0.006	-0.168	0.000
δ	SCHOOL	-0.281	0.213	0.138				-0.005	0.026	0.031	-0.268	0.137
	YEAR	0.256	0.199	0.111				-0.082	0.030	0.030	1.763	0.680
	AGE	0.002	0.005	0.005				-0.004	0.005	0.006	-0.168	0.000
γ	SCHOOL	0.011	0.024	0.027				-0.005	0.026	0.031	-0.268	0.137
	YEAR	-0.100	0.026	0.023				-0.082	0.030	0.030	1.763	0.680
	σ	0.709	0.222	0.199	1.118	1.000	0.678	1.389	0.911	0.724		
	μ	1.212	1.660	1.447	-3.223	7.491	5.190	-1.721	3.611	2.714	0.000	2.135
	lnL		-83.944			-91.610			-86.879			

		Model 4 (KGMHLBC)			Model 5 (RSCFG-μ)			Model 6 (RSCFG)		
		Coeff.	s.e.	Robust s.e.	Coeff.	s.e.	Robust s.e.	Coeff.	s.e.	Robust s.e.
	AGE	0.027	0.027	0.025						
δ	SCHOOL	0.175	0.119	0.108						
	YEAR	-0.490	0.240	0.121						
	AGE				-0.003	0.004	0.004	-0.004	0.004	0.005
γ	SCHOOL				-0.004	0.019	0.023	-0.004	0.022	0.027
	YEAR				-0.060	0.025	0.025	-0.076	0.024	0.027
	σ	0.502	0.036	0.029	1.177	0.479	0.364	0.935	0.241	0.265
	μ	0.100	0.204	0.199	-0.898	1.564	1.082			
	lnL		-88.871			-86.722			-87.376	