Neoclassical versus frontier production models? Testing for the skewness of regression residuals

Kuosmanen, Timo and Fosgerau, Mogens

Department of Economics, MTT, Technical University of Denmark

2009

Neoclassical versus frontier production models?

Testing for the skewness of regression residuals*

Timo Kuosmanen[1] and Mogens Fosgerau[2]

1) Department of Economics, MTT, Luutnantintie 13, 00410 Helsinki, Finland. timo.kuosmanen@mtt.fi.

2) Technical University of Denmark, Knuth-Winterfeldts Allé 116V, 2800 Kgs. Lyngby, Denmark. mf@transport.dtu.dk.

Abstract

The empirical literature on production and cost functions is divided into two strands: 1) the neoclassical approach that concentrates on model parameters, 2) the frontier approach that decomposes the disturbance term to a symmetric noise term and a positively skewed inefficiency term. We propose a theoretical justification for the skewness of the inefficiency term, arguing that this skewness is the key testable hypothesis of the frontier approach. We propose to test the regression residuals for skewness to distinguish the two competing approaches. Our test builds directly upon the asymmetry of regression residuals and does not require any prior distributional assumptions.

Key Words: *Firms and production, Frontier estimation, Hypotheses testing, Production function, Productive efficiency analysis*

JEL classification: C12, C14, D24

---

* This paper emerged from the discussions at the Nordic Econometric Meeting, 24-26 May 2007 in Tartu, Estonia. We would like to thank Professor Martin Browning for encouraging us to write this paper. We are also grateful to the two anonymous reviewers of this journal for providing relevant and exceptionally useful comments and suggestions.

I. Introduction

There is a large literature on the estimation of production and cost functions. A part of this literature, building upon the neoclassical theory of the firm, focuses on the estimation of the parameters of the specified production (or cost) function by regression techniques, and attributes no particular importance to the disturbance term (see e.g. Nerlove, 1963; Christensen and Greene, 1976). The disturbance term represent the errors of specification and measurement. The disturbance term will also capture differences in unobserved or omitted variables, such as management skills or productive efficiency of firms. In general, the disturbances are assumed to have zero mean (or median).

In contrast, there is another large literature that focuses on the measurement of inefficiencies at the level of the individual firm (originating from Koopmans, 1951; Debreu, 1951; and Farrell, 1957), referred to here as the frontier approach. Common to this literature is the notion of an efficient frontier, representing the best practice technology. Firms operating below this frontier are deemed as inefficient, and the degree of inefficiency is measured by the distance to the frontier. In this approach, the disturbance term consists of two components (e.g., Greene, 2008a). The first component represents errors of measurement and specification, and is referred to as the noise term. The noise term is assumed to have some symmetric distribution with zero mean. The second component is supposed to represent the degree of inefficiency. This component is assumed to have a signed, asymmetric distribution. Although the inefficiency and noise terms are both seen as unobservable random variables, Jondrow et al. (1982) have shown that it is possible to use the observed regression residuals for estimating the conditional expected value of the inefficiency term, provided that the inefficiency and noise terms have certain postulated distributions (see Kumbhakar and Lovell, 2000; and Greene, 2008a; for a more detailed exposition).

From the econometric point of view, a failure of the zero mean assumption of the disturbance term will generally lead to biased and inconsistent estimates (Greene, 2008b). Under the assumptions of the frontier approach, the disturbance term has an expected value that is strictly negative. Thus, estimating the neoclassical production/cost function with standard regression techniques when the underlying disturbance term includes a non-positive inefficiency term tends to underestimate the production function (or overestimate the cost function). This may affect the conclusions and policy recommendations drawn from the estimates. By contrast, wrongly assuming the frontier model when a non-positive inefficiency term is not present can lead to overestimation of the production function (or underestimation of the cost function). Obviously, the estimated efficiency scores and rankings are meaningless in the latter situation. In light of these observations, we argue that the specification of disturbance term in the production and cost function models is of such critical importance that it deserves to be subjected to a rigorous statistical test

We may note at this point that the key characteristic of the frontier approach is the division of the disturbance term into symmetric and asymmetric components. This paper contributes first by providing a formalization of how differences in managerial skill may translate into asymmetric inefficiencies. The assumptions of the frontier approach imply that disturbances must be asymmetric. We then propose that testing residuals for asymmetry may be used to indicate whether the frontier approach is relevant.

In the voluminous literature on the frontier approach (see e.g. Fried, Lovell, and Schmidt, 2008, for a review and references), statistical testing of the properties of the disturbance term has attracted surprisingly little attention. The parametric frontier estimation methods such as stochastic frontier analysis (SFA: e.g., Kumbhakar and Lovell, 2000; Greene, 2008a) would readily enable one to apply conventional methods of statistical inference for testing the significance of the inefficiency term, but such tests are conditional on the strong distributional assumptions regarding both the inefficiency and the noise terms. In SFA, the noise term is usually assumed to be normally distributed with zero mean and a constant variance, while the specification of the inefficiency term varies; the most common specifications involve half-normal, truncated normal, exponential, and gamma distributions. The choice of a particular inefficiency distribution is ad hoc; see e.g. Ondrich and Ruggiero (2001) for a critical discussion. Therefore, the objective of this paper is to test for the presence of an inefficiency component in the disturbance term without imposing any parametric assumptions about its distribution.

In the nonparametric literature, Afriat (1972), Hanoch and Rothschild (1972), Diewert and Parkan (1983), and Varian (1984) have proposed a series of axiomatic tests to check whether the observed sample of data is consistent with the hypotheses of profit maximization or cost minimization, but these tests are non-statistical in the sense that they do not allow for any measurement errors or other random noise. Varian (1985) has first proposed a statistical test for cost minimizing behavior that accounts for stochastic noise, but this test requires that the variance of the noise term is known a priori. Kuosmanen, Post and Scholtes (2007) have extended Varian's approach to a more general Pareto-Koopmans efficiency criterion in the general multi-output setting, but their test similarly requires prior specification of the variance of disturbance term. Unfortunately, this variance parameter cannot be estimated from the same sample of firms for which the inefficiency hypothesis is tested.

This paper follows up on Varian (1984) and Kuosmanen et al. (2007) by approaching the notion of inefficiency as an empirically testable hypothesis at the sample level. We propose to test for the neoclassical versus the frontier specification in a nonparametric fashion based on the hypothesis that the disturbance terms have a negatively skewed distribution. We first test the null hypothesis of a normally distributed disturbance term against the alternative hypothesis of a negatively skewed composite disturbance term. The powerful $\sqrt{b_1}$ test

(Shapiro, Wilk, and Chen, 1968) of this hypothesis is based on the third central moment of the residual distribution. We complement the skewness test by applying the b_2 test based on the fourth central moment of residuals to test for violations of normality due to unusual kurtosis or fat/thin tails, which are difficult to interpret as signs of inefficiency. We then relax the normality assumption, and test the null hypothesis of a symmetrically distributed disturbance term against the alternative hypothesis of a negatively skewed composite disturbance term. The $\sqrt{b_1}$ test is adapted to the more general null hypothesis by applying the consistent bootstrap procedure devised by Pérez-Alonso (2007). In both tests, the acceptance of the null hypothesis is interpreted as evidence in favor of the neoclassical model, whereas its rejection is seen as evidence in favor of the frontier model.

We apply the proposed skewness tests to the classic data set of U.S. electricity firms examined by Nerlove (1963) and Christensen and Greene (1976). This data has been used as a textbook example in a number of neoclassical and frontier production studies (see e.g. Greene, 2008a,b), and it hence qualifies as an ideal test case for examining the results of the skewness test. We apply four different methods for estimating the production technology, and apply the proposed skewness tests to the residuals obtained with each method. The application shows that the proposed tests can be useful for identifying the specific assumptions, model types and estimation methods for which the use of the frontier model is consistent with the data.

The rest of the paper is organized as follows. Section 2 describes the production model and the related notation, terminology, and assumptions. Section 3 presents the test of normality against the skewed alternative. Section 4 relaxes the normality assumption, and presents the test of symmetry against the skewed alternative. Section 5 applies the test in a number of data sets reported in the literature. Section 6 contains our concluding remarks.

II. Neoclassical versus frontier models

Consider the standard econometric production model of type

$$\ln y_i = f(\ln \mathbf{x}_i) + \varepsilon_i, \quad i = 1,...,n, \tag{1}$$

where y_i represents the output of firm *i*, *f* denotes the production function that characterizes the technology, $\mathbf{x}_i \in \mathbb{R}_+^m$ is the vector of inputs, and ε_i denotes the disturbance term (assumed to be exogenous in the sense that $E(\varepsilon_i | \mathbf{x}_1,...,\mathbf{x}_n) = 0 \ \forall i = 1,...,n$). Alternatively, *f* could be interpreted as the cost function, in which case y_i would be the observed total cost of firm *i*, and $\mathbf{x}_i \in \mathbb{R}_+^m$ would be a vector of input prices and outputs. Moreover, model (1) could be interpreted in terms of profit or distance functions in an analogous fashion. The particular interpretation of the model will be immaterial for the analysis that follows. For brevity, we will henceforth restrict attention on the production function interpretation of model (1).

The production function f is referred to as the deterministic part and the disturbance term ε_i as the stochastic part of the model.[1] The disturbance term captures the effects of measurement errors, specification errors, and any other deviations from the otherwise stable deterministic technology. It will also capture effects of omitted variables, such as unobserved differences in management skill. Therefore, it must be emphasized that whereas the neoclassical theory of the firm postulates that rational firms operate with full efficiency, the empirical production models referred to above as "neoclassical" can assimilate inefficiencies in the disturbance terms ε, provided that these inefficiencies are symmetrically distributed. We elaborate to the symmetry requirement in more detail below.

Formatted: Font color: Black, English (U.S.), Lowered by 3 pt

[1] Inputs **x** may be considered as deterministic or stochastic. At the firm level, inputs may be seen as deterministic decision variables specified by the management: observed input demands **x** are optimal choices to the firm's profit maximization problem. If we aggregate across firms and/or sectors, then aggregate inputs could be treated as random variables. Deterministic or stochastic specification of inputs does not make any difference to the tests we propose.

The frontier models are identical to the above neoclassical model with respect to the deterministic part; the only difference concerns the disturbance term ε_i. The frontier models treat ε_i as a composite disturbance that can be decomposed as

$$\varepsilon_i = v_i - u_i, \quad i = 1,...,n, \tag{2}$$

where $u_i \geq 0$ is a one-sided, asymmetric inefficiency component with $E(u_i) > 0$, and v_i is a symmetric noise component with $E(v_i) = 0$. To disentangle inefficiency from noise, more detailed assumptions are necessary; we will return to the specific assumptions in the next sections. However, it is worth emphasizing the negative skewness of the disturbance term ε_i as the key testable implication of the frontier model. Note that the neoclassical model is obtained as a restricted special case of (2) if $u_i = 0$. Moreover, the inefficiency term u_i is indistinguishable from v_i if it has a symmetric distribution. Importantly, introducing the asymmetric inefficiency term violates the exogeneity assumption: $E(\varepsilon_i | \mathbf{x}_1,...,\mathbf{x}_n) = -E(u_i) < 0 \ \forall i = 1,...,n$. Moreover, we note that $E(y_i | \mathbf{x}_i) = f(\mathbf{x}_i) - E(u_i) < f(\mathbf{x}_i)$. Thus, assuming the neoclassical model when the disturbance term contains the inefficiency component will result in biased and inconsistent estimates regarding *f*.

The positive skewness of the inefficiency component u_i can be motivated as follows.[2] Let us think of the inefficiency component u_i as a function of unobserved management skill s_i. More specifically, the inefficiency component is a bounded function $u_i = u(s_i)$, such that $\lim_{s \to \infty} u(s) = 0$. The upper bound of $u(s) = 0$ represents the best practice: however good the manager, one cannot achieve better than the best practice. It seems meaningful to assume that inefficiency decreases as skill increases, that is, $u'(s) < 0$ for all *s*. Furthermore, achieving

[2] We are grateful to Professor Nils Gottfries for suggesting this idea in his editorial comments.

higher levels of performance gets progressively more difficult, in other words, skill has diminishing marginal returns: $u''(s) > 0$ for all s. Now, suppose the distribution of management skill s is unimodal and symmetric (measurable skills, such as running skill or IQ test results, tend to follow this pattern). Thus, differences in unobserved management skill could be directly attributed to the standard symmetric disturbance term of the neoclassical model. However, if management skill is converted to inefficiency through a decreasing and convex function $u(s)$, as assumed above, then the symmetric, unimodal distribution of s will result in a right-skewed unimodal distribution of the inefficiency term u_i, in line with the frontier model.

The conversion of symmetric skill to right-skewed inefficiency through a convex inefficiency function u is illustrated by a numerical example in Figure 1. Suppose management skill s is a normally distributed random variable, with the density function graphed in the bottom-right quarter of Figure 1. The inefficiency function is assumed to take the form of $u(s)$ = exp(-s), which satisfies all of the postulated properties (i.e., u is decreasing, concave, and has the limit zero as s approaches to infinity). The inefficiency function is plotted in the top-right quarter of the diagram. Now, converting the normally distributed skill through this inefficiency function u results as a skewed inefficiency distribution, as illustrated in the top-left quarter of Figure 1. The thin grey lines further illustrate the conversion by showing the correspondence between the lower and upper quartiles of the skill and inefficiency distributions. Note that the inefficiency distribution obtained in this example is log-normal, with the distinct right-skewed shape of the density function that closely resembles that of the standard half-normal inefficiency distribution typically assumed in the literature of frontier estimation.

(Figure 1 around here)

The question of whether skill influences performance directly, or through a convex inefficiency function $u(s)$, need not be taken by faith: it can be tested empirically. In the following sections we propose some statistical test procedures for this purpose. As a preliminary step, we must estimate the conditional expectation $E(y_i | \mathbf{x}_i) = f(\mathbf{x}_i) - E(u_i)$ by some parametric or nonparametric regression technique. The parametric approach requires that the functional form of f is specified a priori, and the unknown parameters of f are estimated, e.g., by means of maximum likelihood or least squares techniques (see e.g. Greene, 2008b). Nonparametric estimation of (1) can build upon local averaging or shape constraints. A prime example of local averaging is kernel regression; see Fan, Li, and Weersink (1996) for an application of kernel regression to frontier estimation. If one assumes that f satisfies certain regularity conditions (monotonicity, concavity) then one can estimate f by nonparametric least squares (NLS) subject to shape constraints (Hildreth 1954; Hanson and Pledger 1976; Groneboom et al., 2001; Kuosmanen 2008); see Kuosmanen (2006) and Kuosmanen and Kortelainen (2007) for applications of NLS to frontier estimation. Other than the regularity conditions, the NLS approach does not require any prior assumptions about the functional form of f or its smoothness. However, it may be sensible to impose some additional smoothness conditions to alleviate the curse of dimensionality.

As a result of the regression analysis, we obtain a vector of residuals denoted by $\mathbf{e} = (e_1 ... e_n)'$. The tests developed in the next sections use the residuals conditional on the estimated models. The residuals are the object of interest in the frontier approach. We conclude by emphasizing that the estimation of model (1) by regression techniques does not involve any loss of generality as such. Moreover, the efficiency tests to be developed in the next section are not limited to any particular regression technique.

III. Testing normality against skewness

In regression analysis, normality of the disturbance term is a standard assumption that is frequently invoked to facilitate maximum likelihood estimation and the conventional methods of statistical inference. In the SFA literature, the normality of the noise component is widely used; virtually all cross-sectional SFA studies known to us assume a normally distributed noise component (one notable exception is Goldstein, 2003). Therefore, we start by formulating the null hypothesis as:

H_0: Disturbances $\boldsymbol{\varepsilon}$ of model (1) are normally distributed.

Acceptance of the null hypothesis is interpreted as evidence in favor of the neoclassical model. The normality assumption will be relaxed in Section 4.

While it might seem desirable to test normality against a general non-normal alternative, it is well-known that narrowing the class of alternatives can substantially improve the statistical power of the test (see, e.g., Poitras 2006). In the present context, violations of normality due to non-normal kurtosis or too fat or thin tails of the residual distribution are difficult to interpret as signs of the presence of an inefficiency term. Therefore, we exploit the asymmetric structure of the inefficiency component, and specify the alternative hypothesis as

H_1: Disturbances $\boldsymbol{\varepsilon}$ of model (1) are negatively skewed.

As the negative sign of skewness matches with the theoretical model of inefficiency presented in Section 2, we interpret the rejection of H_0 in favor of H_1 as evidence supporting the frontier model.

Few subjects in applied statistics have attracted as much attention as the tests of normality, and the main use of these tests concerns normality of the regression residuals. As a result, a number of alternative procedures are available. The statistical power of the normality tests has been investigated over a wide range of different distributions by means of Monte Carlo simulations (see e.g. D'Agostino, 1986, and Throde, 2002, for reviews). A general conclusion from these power studies is that the classic chi-squared and Kolmogorov-Smirnov tests have poor power properties and should not be used for testing normality. The recommended tests include the Shapiro-Wilk W test (Shapiro and Wilk, 1965), the third sample moment ($\sqrt{b_1}$) tests (Shapiro, Wilk, and Chen, 1968), the fourth sample moment (b_2) tests (D'Agostino and Pearson, 1973), and the D'Agostino-Pearson K^2 test that combines them. The W and K^2 tests are general purpose (omnibus) tests that have good power against a range of non-normal distributions, while the third and fourth sample moment tests ($\sqrt{b_1}$, b_2) are particularly good at detecting non-normal skewness and kurtosis, respectively. Given the particular specification of H_1 above, the third moment $\sqrt{b_1}$ test is our preferred choice.[3]

The third moment $\sqrt{b_1}$ test is simple to implement. Given the central moments of the residual distribution, defined as $m_j = \sum_{i=1}^{n} (e_i - \overline{e})^j / n$, j=2,3, where $\overline{e}$ denotes the mean $\overline{e} = \sum_{i=1}^{n} e_i / n$, the test statistic is computed as

$$\sqrt{b_1} = m_3 / (m_2)^{3/2} . \tag{3}$$

The test statistic is invariant to location and scale, so the critical values of $\sqrt{b_1}$ under the null hypothesis can be computed for any sample size and desired level of significance by means of a simple Monte Carlo simulation where random pseudo-samples are drawn from the standard normal distribution. More specifically, we draw M random pseudo-samples of n observations independently from N(0,1), and compute the $\sqrt{b_1}$ statistic for each pseudo-sample. Given

[3] See Poitras (2006) for some recent evidence from Monte Carlo simulations that demonstrates the high power of the $\sqrt{b_1}$ test against skewed alternative hypotheses compared to the omnibus tests such as D'Agostino-Pearson K^2 test as well as different variants of the Jarque and Bera (1980) test.

sufficiently large M, the α percentile of the thus obtained simulated distribution of the $\sqrt{b_1}$ statistics can be used as the critical value of the test statistic at the significance level α.[4]

Although the $\sqrt{b_1}$ test is the most powerful known method for detecting non-normal skewness, it may reject the null hypothesis even if the true distribution is perfectly symmetric but has non-normal kurtosis (see e.g. Poitras, 2006). In contrast to skewness, fat or thin tails of the residual distribution are difficult to interpret as signs of inefficiency. Indeed, in the context of H_0, non-normal kurtosis can be viewed as a sign of some data problems (e.g., outliers) or some sort of model misspecification. Hence, testing for non-normal kurtosis can provide useful supplementary information.

The standard b_2 kurtosis test, based on the fourth central moment of the residual distribution can be implemented similarly to the $\sqrt{b_1}$ test. The test statistic is computed as

$$b_2 = m_4 /(m_2)^2 . \tag{4}$$

The critical values of the b_2 statistic can be computed by the Monte Carlo method as described above.

The combined use of the $\sqrt{b_1}$ and b_2 tests can result in four possible outcomes with the following interpretations: 1) the null hypothesis is maintained in both $\sqrt{b_1}$ and b_2 tests: this supports the neoclassical model. 2) the null is rejected by the $\sqrt{b_1}$ test but is maintained by the b_2 test: this supports the frontier model. 3) both tests reject the null hypothesis: both neoclassical and frontier models are plausible but the evidence is inconclusive due to data problems or model misspecification. 4) the null is maintained by the $\sqrt{b_1}$ test but is rejected

[4] The critical values of $\sqrt{b_1}$ for certain sample sizes and significance levels have been tabulated by Pearson and Hartley (1966). D'Agostino (1970) has derived a simple transformation of $\sqrt{b_1}$ that follows approximately $N(0,1)$. With present computers it is, however, easy to compute critical values by simulation.

by the b_2 test: the frontier model is rejected, but the neoclassical model assuming normality also suffers from data problems or model misspecification.

In conclusion, the $\sqrt{b_1}$ and b_2 tests are easy to implement, and have proved as powerful methods for detecting non-normal skewness and kurtosis in a large number of Monte Carlo studies (see e.g. D'Agostino, 1986; Throde, 2002; Poitras, 2006; and references therein). Since skewness is the distinguishing feature of the presence of an asymmetric inefficiency component in the disturbance term, a test directed at skewness will be more powerful than any alternative omnibus test against an unspecified non-normal alternative. However, the null hypothesis of a normally distributed disturbance term can be questioned. If the true distribution of disturbances is symmetric but non-normal, the wrongly imposed normality assumption can yield misleading results. In the next section we relax the normality assumption and devise a fully nonparametric test based on the symmetry of the disturbance term under the null hypothesis.

IV. Testing symmetry against skewness

Normality of the disturbance term is not an innocuous assumption. In economic applications, distributions of regression residuals often exhibit fat tails that violate normality, but this cannot be interpreted as evidence in favor of the frontier model. Therefore, in this section we consider the neoclassical model from a broader perspective, relaxing the strong normality assumption.

Specifically, we test the following hypotheses:

H_0': Disturbances $\boldsymbol{\varepsilon}$ of model (1) are symmetrically distributed.

H_1': Disturbances $\boldsymbol{\varepsilon}$ of model (1) are negatively skewed.

Acceptance of the null hypothesis is again interpreted as evidence in favor of the neoclassical model, while its rejection is viewed as evidence in favor of the frontier model. However, it should be emphasized that H_0 considered in Section 3 represents a more restrictive interpretation of the neoclassical model than H_0'. While the test of Section 3 is based on the null hypothesis of normally distributed disturbances, the test in this section is based on the null hypothesis of symmetrically distributed disturbances. These are two different hypotheses that should not be confused. We emphasize that acceptance of the null hypothesis, like in section 3, does not imply that firms are fully efficient. It merely implies that disturbances do (probably) not contain an asymmetric component, which is the distinguishing feature of the frontier approach.

Deleted: approach

Deleted: ;

Deleted: can distinguish inefficiency from noise provided that inefficiency is non-normal, the test of this section additionally requires that inefficiency is non-symmetric.

Deleted: t

Deleted:

Deleted: residual

Testing symmetry of the distribution of residuals has attracted a lot of attention in econometrics and statistics in the recent decades (e.g. Godfrey and Orme, 1991; Ahmad and Li, 1997; Bai and Ng, 2001; and Hyndman and Yao, 2002). The test by Godfrey and Orme (1991) is based on the asymptotic distributions of the central moments. However, the higher central moments are known to converge slowly, which may result in low power of the asymptotic tests even in moderately sized samples, e.g., n = 100 (see Poitras, 2006, for discussion). Therefore, we resort to the $\sqrt{b_1}$ test described in Section 3 and extend its scope to the non-normal case by applying the bootstrap approach proposed by Pérez-Alonso (2007).

The challenge of applying the $\sqrt{b_1}$ test to the more general, nonparametric H_0' arises from the fact that the true distribution under the null remains unspecified. Pérez-Alonso (2007) has devised a nonparametric bootstrap approach, which can be adapted to our purposes. The main idea is to find a bootstrap distribution that mimics the actual distribution of disturbances under the null hypothesis. Pérez-Alonso proposes a simple resampling scheme where the bootstrap data-generating process respects the null hypothesis and mimics the observed empirical distribution as closely as possible. She also proves the consistency of the proposed bootstrap procedure.

Pérez-Alonso's (2007) bootstrap procedure can be adapted for our purposes as follows. We compute the critical values of the $\sqrt{b_1}$ statistic under the H_0' of symmetry in five stages.

Stage 1: Construct re-centered versions of the residuals $\tilde{e}_i = e_i - \overline{e}$.

Stage 2: Assign random signs to the centered residuals $\tilde{e}_i$ according to independent realizations of a Rademacher random variable s_i, independent of $\tilde{e}_i$, taking values -1 and +1 with equal probability of 1/2 each. As a result, we obtain a set of symmetrized residuals $\{s_1\tilde{e}_1,...,s_n\tilde{e}_n\}$.

Stage 3: Apply a random number generator to independently draw integers $\{i_1,...,i_n\} \in \mathbb{N}_+^n$, where $\Pr(i_i = z) = 1/n \ \forall z \in \{1,...,n\}$. Use the thus obtained sequence of integers to select elements of $\{s_1\tilde{e}_1,...,s_n\tilde{e}_n\}$ to the bootstrap sample $\{b_1,...,b_n\}$. Note that we sample with replacement to allow any residual $s_i\tilde{e}_i$ to appear more than once in the sample.

Stage 4: Compute the $\sqrt{b_1}$ statistic (5) using the bootstrap sample $\{b_1,...,b_n\}$.

Stage 5: Repeat Stages 2-4 *M* times to construct a bootstrap distribution of the $\sqrt{b_1}$ statistic under the null hypothesis of symmetry. The 1-α percentile of the thus obtained distribution can be used as the critical value of the test statistic at the significance level α.

Note that the only difference between the proposed $\sqrt{b_1}$ test of symmetry and the classic $\sqrt{b_1}$ test of normality concerns the true distribution under the null, which is used for computing the critical values. In the symmetry test we sample from the symmetrized residual distribution, whereas in the normality test we sample from the standard normal density. The power of the bootstrap method is evident from the Monte Carlo simulations reported by Pérez-Alonso (2007).

V. Application

To illustrate how the proposed tests work in practice, we re-examined the classic production data of the U.S. electricity companies from years 1955 and 1970, reported by Nerlove (1963) and Christensen and Greene (1976), respectively.[5] The data sets include one output (electricity) and three inputs (labor, capital, and fuel). Both volume and cost data are available, which facilitates the estimation of both production and cost functions. The sample sizes are 145 and 123 in years 1955 and 1970, respectively. The sample sizes and dimensionality of these data are well representative of empirical studies reported in the literature. These particular data are widely used as a textbook example case (Greene, 2008b), and they have been re-examined by frontier methods (see e.g. Greene, 2008a). Hence, these data provide an ideal test case for the proposed tests.

We estimated both production and cost function models for years 1955 and 1970. All four models were estimated by four different regression methods: 1) OLS with Cobb-Douglas functional form, 2) OLS with the flexible translog functional form, 3) kernel regression, and 4) nonparametric least squares (NLS) subject to monotonicity and concavity constraints.[6] Given the residuals, we tested for the skewness of the disturbance terms against the symmetric and normal alternatives. We also tested for the non-normal kurtosis. This results in 48 different tests in total. The values of the test statistics and the p-values (i.e., the probability of obtaining the observed value of the test statistic or lower when the null hypothesis is true) are summarized in Tables 1 and 2. Table 1 relates to the production function and Table 2 to the cost function estimations, respectively.

(Table 1 around here)

(Table 2 around here)

[5] The data are available online at: http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm.

[6] In the kernel method, the Gaussian kernel function was used. The bandwidth parameters were selected by cross-validation to minimize the mean squared error.

The first columns of Tables 1 and 2 report the $\sqrt{b_1}$ test statistics. According to the inefficiency hypothesis, we should expect negative skewness in the case of the production function, and positive skewness in the case of the cost function. All eight production function models have the expected sign, but in the case of the cost function, four out of eight models have the unexpected (negative) sign. The production function models can only capture technical inefficiencies (i.e., output falling short of its maximum value), whereas the cost function model can also identify allocative inefficiencies (i.e., excess costs due to the wrong input mix).

Two possible interpretations of the neoclassical hypothesis have been considered: in Section 3 we interpreted normality of the residuals as evidence in favor of the neoclassical model, in Section 4 we considered symmetry of residuals as sufficient evidence in its favor. The conclusions drawn from the tests depend on how much we require from the neoclassical hypothesis.

If normality of the disturbance term is viewed as an integral part of the neoclassical model, then we recommend the combined use of the $\sqrt{b_1}$ and b_2 tests, as discussed in Section 3. The relevant p-values of these normality tests are reported in Tables 1 and 2 in the second and third columns from the right. From the b_2 test we see that the normality is rejected due to excess kurtosis in all 16 cases considered. Regarding the $\sqrt{b_1}$ test, normality is rejected in favor of skewness, in the half of the cases (five production functions and three cost functions). In those cases we conclude that the frontier model is plausible, but the evidence is inconclusive due to data problems or model misspecification. In those eight cases where the $\sqrt{b_1}$ test maintains normality, the evidence does not support either the frontier or the neoclassical models. Neither model explains the fat tails of the residual distribution, observed across all model types and specifications. Since the fat tails occur both in parametric and nonparametric models, we suspect it is an inherent feature of the data.

If we interpret the neoclassical model more broadly and accept symmetry of the disturbances as its distinguishing feature compared to the frontier model, then the symmetry test described in Section 4 is the relevant test to consider. The same $\sqrt{b_1}$ test statistic is used as in the case of the normality test, but the p-values of the symmetry tests should be read from the right-most columns of Tables 1 and 2. The results of the symmetry test broadly support the neoclassical view; the null hypothesis of symmetry is only rejected in two cases (production functions estimated by the kernel regression). For most model types and estimation methods, the evidence does not support the application of the frontier model to these specific data.

In conclusion, the most robust finding in our tests was the rejection of normality due to excess kurtosis. Proponents of the frontier approach might argue that non-normal kurtosis is just an artifact of an inefficiency term that is present in the disturbances, but the fact that excess kurtosis occurs even when skewness is insignificant or has an unexpected sign speaks against this interpretation. Neither of the competing paradigms can explain the fat tails of the distribution. To our knowledge, the problem of fat tails has gone unnoticed in the previous studies that have analyzed these data. This should be taken into account in the statistical inferences; the conventional methods of statistical inference can yield misleading results if the disturbances are non-normal.

VI. Concluding remarks

We have shown that the question of neoclassical versus frontier production model is amenable to statistical testing. We first outlined a new theoretical model where a symmetric unimodal skill distribution is transformed to a positively skewed inefficiency distribution through a decreasing and convex inefficiency function. Using the resulting negative skewness of the composite disturbance term as the key testable implication of the frontier model, we formulate pairs of hypotheses that are empirically testable at the sample level. Thus, it is no longer necessary to choose a neoclassical or frontier model based on faith: it is ultimately an empirical question.

We performed the proposed tests to the classic data sets of the U.S. electricity firms from years 1955 and 1970, using a variety of different regression methods to estimate the production technology. Our analysis pinpoints the specific assumptions, model types and estimation methods for which the use of the frontier model can be justified. For most models considered, the empirical evidence supports the neoclassical model when symmetry of the disturbance term is broadly interpreted as its distinguishing feature. However, the most robust finding of our tests was the identification of non-normal fat tails in the residuals, which might affect the skewness tests as well. Fat tails occurred in all different parametric and nonparametric production and cost function models considered, which strongly suggests that it is a real feature of the data.

We consider it relevant to perform the proposed skewness and kurtosis tests in empirical studies of production and cost functions. Whatever the preferred paradigm, the test results provide an empirical basis for choosing between the conventional and the frontier specification of production/cost functions. We hope that evidence will be accumulated to enable a more informed assessment of the relative merits of the two competing paradigms prevailing in the literature.

References

Afriat, S. (1972), Efficiency Estimation of Production Functions, *International Economic Review 13*, 568-598.

Ahmad, I.A. and Li Q. (1997), Testing the Symmetry of an Unknown Density Function by the Kernel Method, *Journal of Nonparametric Statistics 7*, 279–293.

Bai, J. and Ng S. (2001), A Consistent Test for Conditional Symmetry in Time Series Models, *Journal of Econometrics 103*, 225–258.

Christensen, L.R. and Greene W.H. (1976), Economies of Scale in U.S. Electric Power Generation, *Journal of Political Economy 84*, 655-676.

D'Agostino, R.B. (1970), Transformation to Normality of the Null Distribution of g_1, *Biometrika 57*, 679-681.

D'Agostino, R.B. (1986), Tests for the Normal Distribution, Ch. 9 in R.B. D'Agostino and M.A. Stephens (eds.), *Goodness-of-fit techniques*, Statistics: textbooks and monographs 68, Marcel Dekker, New York.

D'Agostino, R.B. and Pearson E. (1973), Tests for Departures from Normality: Empirical Results for the Distribution of $\sqrt{b_1}$ and b_2, *Biometrika 60*, 613-622.

Debreu, G. (1951), The Coefficient of Resource Utilization, *Econometrica 19*, 273-292.

Diewert, W.E. and Parkan C. (1983), Linear Programming Tests of Regularity Conditions for Production Frontiers, in W. Eichhorn et al. (eds.), *Quantitative studies on production and prices*, Physica-Verlag, Würzburg.

Fan, Y., Li, Q. and Weersink A. (1996), Semiparametric Estimation of Stochastic Production Frontier Models, *Journal of Business and Economic Statistics 14*, 460-468.

Farrell, M.J. (1957), The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society Series A. General* 120, 253-282.

Fried, H., Lovell C.A.K. and Schmidt S. (2008), *The Measurement of Productive Efficiency and Productivity Change*, Oxford University Press, New York.

Goldstein, H. (2003), On the COLS and CGMM Moment Estimation Methods for Frontier Production Functions, Ch. 14 in B.P. Stigum (ed.), *Econometrics and the Philosophy of Economics*, Princeton University Press, Princeton.

Godfrey, L.G. and Orme C.D. (1991), Testing for Skewness of Regression Disturbances, *Economics Letters 37*, 31-34.

Greene, W.H. (2008a), The Econometric Approach to Efficiency Analysis, Ch. 2 in H. Fried et al. (eds.), *The Measurement of Productive Efficiency and Productivity Change*, Oxford University Press, New York.

Greene, W.H. (2008b), *Econometric Analysis*, 6th Edition, Pearson Education, New Jersey.

Groeneboom, P., Jongbloed G. and Wellner J.A. (2001), Estimation of a Convex Function: Characterizations and Asymptotic Theory, *Annals of Statistics 29*, 1653-1698.

Hanoch G. and Rothschild M. (1972), Testing Assumptions of Production Theory: A Nonparametric Approach, *Journal of Political Economy 80*, 256-275.

Hanson, D.L. and Pledger G. (1976), Consistency in Concave Regression, *Annals of Statistics 4*, 1038-1050.

Hildreth, C. (1954), Point Estimates of Ordinates of Concave Functions, *Journal of the American Statistical Association 49*, 598-619.

Hyndman, R.J. and Yao Q. (2002), Nonparametric Estimation and Symmetry Tests for Conditional Density Functions, *Journal of Nonparametric Statistics 14*, 259-278.

Jarque, C.M. and Bera A.K. (1980), Efficient Tests for Normality, Heteroskedasticity, and Serial Independence of Regression Residuals, *Economics Letters 6*, 255–259.

Jondrow, J. Lovell, C.A.K., Materov, I.S. and Schmidt P. (1982) On Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model, *Journal of Econometrics 19*, 233-238.

Koopmans, T.C. (1951), An Analysis of Production as an Efficient Combination of Activities, in, T.C. Koopmans (ed.), *Activity Analysis of Production and Allocation*, Cowles Commision for Research in Economics, monograph no. 13, John Wiley & Sons Inc., New York.

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

Kuosmanen, T. (2006), Stochastic Nonparametric Envelopment of Data: Combining Virtues of SFA and DEA in a Unified Framework, MTT Discussion Paper 3/2006.

Kuosmanen, T. (2008), Representation Theorem for Convex Nonparametric Least Squares, *Econometrics Journal 11,* 308-325.

Kuosmanen T. and Kortelainen M. (2007), Stochastic Nonparametric Envelopment of Data: Cross-Sectional Frontier Estimation Subject to Shape Constraints, University of Joensuu, Economics Discussion Paper 46.

Kuosmanen, T., Post, G.T. and Scholtes S. (2007), Nonparametric Tests of Productive Efficiency with Errors-in-Variables, *Journal of Econometrics 136*, 131-162.

Nerlove, M. (1963), Returns to Scale in Electricity Supply, Ch. 7 in C. Christ et al. (eds.), *Measurement in Economics*, Stanford University Press, Stanford.

Ondrich, J. and Ruggiero J. (2001), Efficiency Measurement in the Stochastic Frontier Model, *European Journal of Operational Research 129*, 434-442.

Pearson, E. and Hartley H. (1966), *Tables for statisticians*, 3rd ed. Biometrika Vol. I, Cambridge University Press, Cambridge.

Pérez-Alonso, A. (2007), A Bootstrap Approach to Test the Conditional Symmetry in Time Series Models, *Computational Statistics and Data Analysis 51*, 3484–3504.

Poitras, G. (2006), More on the Correct Use of Omnibus Tests for Normality, *Economics Letters 90*, 304-309.

Shapiro, S.S. and Wilk M.B. (1965), An Analysis of Variance Test for Normality (Complete Samples), *Biometrika 52*, 591-611.

Shapiro, S.S., Wilk, M.B. and Chen H.J. (1968), A Comparative Study of Various Tests for Normality, *Journal of American Statistical Association 63*, 1343-78.

Throde, H. (2002), *Testing for Normality*, Marcel Dekker, New York.

Varian, H.R. (1984), The Nonparametric Approach to Production Analysis, *Econometrica 52*, 579-598.

Varian, H. (1985), Nonparametric Analysis of Optimizing Behavior with Measurement Error, *Journal of Econometrics 30*, 445-458.

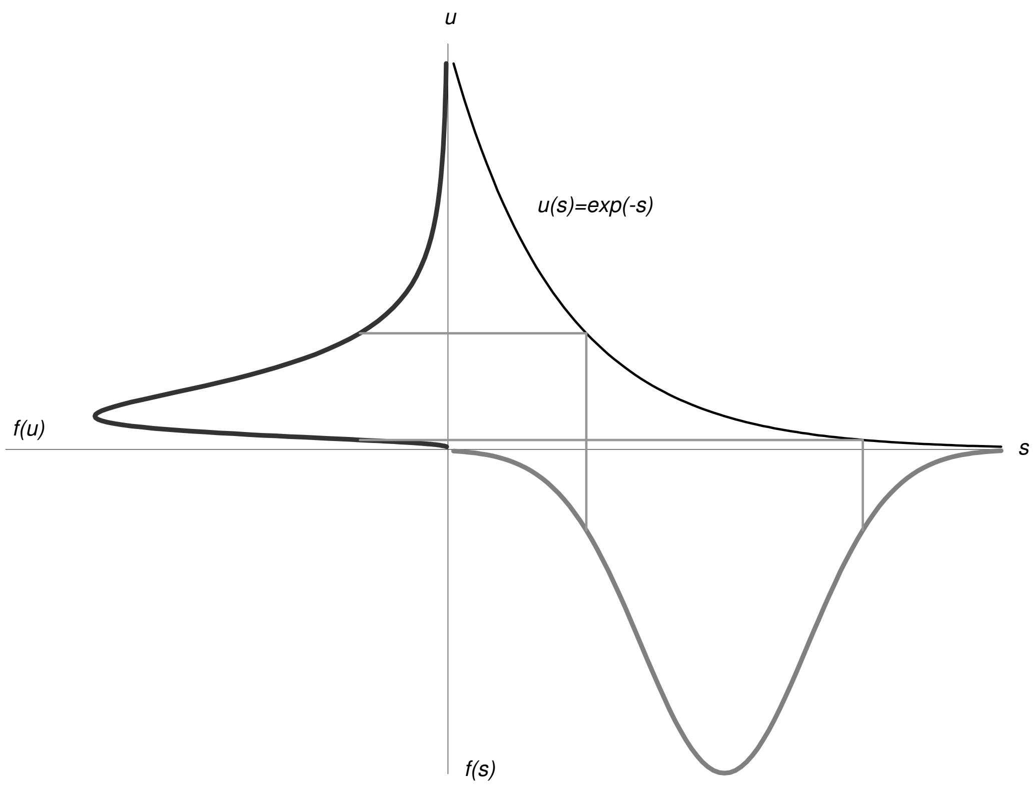

Figure 1: Illustration of how the decreasing convex inefficiency function u(s) = exp(-s) converts a symmetric skill distribution into a positively skewed inefficiency distribution.

Table 1: Results of the skewness and kurtosis tests; production function

year	Production function regression method	test statistics $\sqrt{b_1}$	b_2	p-values $\sqrt{b_1}$ normality	b_2 normality	$\sqrt{b_1}$ symmetry
1955	OLS Cobb-Douglas	-0.209	8.287	0.255	0.000	0.448
	OLS Translog	-0.530	7.630	0.044	0.000	0.339
	Kernel	-2.356	8.052	0.000	0.000	0.007
	Nonparametric least squares	-0.263	9.194	0.203	0.000	0.437
1970	OLS Cobb-Douglas	-1.256	8.127	0.000	0.000	0.195
	OLS Translog	-0.092	5.509	0.395	0.000	0.457
	Kernel	-3.835	18.511	0.000	0.000	0.047
	Nonparametric least squares	-2.045	10.029	0.000	0.000	0.089

Table 2: Results of the skewness and kurtosis tests; cost function

	Cost function	test statistics		p-values		
				$\sqrt{b_1}$	b_2	$\sqrt{b_1}$
year	regression method	$\sqrt{b_1}$	b_2	normality	normality	symmetry
1955	OLS Cobb-Douglas	1.302	7.614	0.000	0.000	0.153
	OLS Translog	0.289	6.944	0.187	0.000	0.413
	Kernel	-2.685	11.925	1.000	0.000	0.968
	Nonparametric least squares	2.054	9.283	0.000	0.000	0.058
1970	OLS Cobb-Douglas	0.991	6.078	0.002	0.000	0.188
	OLS Translog	-0.020	3.718	0.533	0.040	0.526
	Kernel	-2.607	9.581	1.000	0.000	0.983
	Nonparametric least squares	-0.128	3.987	0.648	0.017	0.568