SMOOTH CONSTRAINED FRONTIER ANALYSIS

CHRISTOPHER F. PARMETER AND JEFFREY S. RACINE

Abstract. Production frontiers (i.e., 'production functions') specify the maximum output of firms, industries, or economies as a function of their inputs. A variety of innovative methods have been proposed for estimating both 'deterministic' and 'stochastic' frontiers. However, existing approaches are either parametric in nature, rely on nonsmooth nonparametric methods, or rely on nonparametric or semiparametric methods that ignore theoretical axioms of production theory, each of which can be problematic. In this paper we propose a class of smooth constrained nonparametric and semiparametric frontier estimators that may be particularly appealing to practitioners who wish smooth (i.e., continuously differentiable) estimates that, in addition, are consistent with theoretical axioms of production.

1. Overview

Estimating production relationships is a key component of both applied micro- and macroeconomic research. Modern analysis traces its roots to the pioneering empirical work of Cobb & Douglas (1928), Klein (1947) and Arrow, Chenery, Minhas & Solow (1961). Though this work was notable for the application of statistical methodology to the estimation of a fundamental object in economics (the 'production function'), the field continues to evolve in innovative and sometimes controversial ways; see Leibenstein (1966) for a case in point. The theory of production has matured considerably since its early days. The modern theoretical framework stems from pathbreaking work by Debreu (1951), Shepard (1953, 1970), and Diewert (1971). The contemporary estimation of 'deterministic' and 'stochastic' production frontiers dates to the seminal work of Aigner & Chu (1968) and Aigner, Lovell & Schmidt (1977), respectively.

One popular approach to the estimation of deterministic frontiers is known as 'data envelopment analysis' (DEA); see Simar & Wilson (2008) and the references therein. DEA is a mathematical ('goal') programming approach designed to measure technical efficiency of decision making units

Date: November 6, 2009.
Key words and phrases. Efficiency, Concavity, Monotonicity, Constrained Kernel Estimator.
Preliminary and incomplete – kindly do not quote without permission. We would like to thank but not implicate Subal Kumbhakar, Peter Schmidt, Paul Wilson and Jeff Wooldridge for their insightful comments and suggestions. All errors in this paper are due to technical inefficiency and not noise. Racine would like to gratefully acknowledge support from Natural Sciences and Engineering Research Council of Canada (NSERC:www.nserc.ca), the Social Sciences and Humanities Research Council of Canada (SSHRC:www.sshrc.ca), and the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca).

(DMUs). DEA's chief advantage compared to econometric, regression-based techniques is its nonparametric (i.e., flexible) treatment of the production frontier. DEA relies on general axioms of production theory (e.g., monotonicity, concavity, homogeneity) but does not require nor assume specific functional forms. Moreover, given the deterministic nature of DEA, no *a priori* specification of the nature of efficiency is required.

Alternative methods for estimating deterministic frontiers that do not hinge on goal programming and are firmly entrenched in regression based techniques are 'corrected' and 'modified' ordinary least squares, COLS and MOLS, respectively. These methods 'shift' an estimated conditional mean so that the resulting shifted function lies (almost) everywhere on or above the data. The idea of shifting an estimated production function stems from Winsten's (1957) comment on Farrell's (1957) description of an industry 'envelope' isoquant, namely, that the estimated regression represents an average production function which could be shifted vertically to estimate the production frontier itself. Of course, potentially restrictive conditions may in general be required in order for the shifted mean function to mimic the frontier. In the COLS framework an OLS regression's intercept is increased (decreased) so that the estimated function lies on or above (below) all of the data. Greene (1980) presents a detailed treatment of COLS and shows that this estimator is consistent (in the no noise setting), albeit under the presumption of correct parametric specification and a homoscedastic inefficiency distributive law, both of which are potentially restrictive.

Afriat (1972) and Richmond (1974) were the first to propose MOLS. MOLS shifts the production function up by an amount equal to the expected valued of the assumed distribution of inefficiency as opposed to the amount necessary to envelope the data. That is, COLS shifts the conditional mean function up by the maximal residual whereas MOLS assumes a distribution and then uses the OLS residuals to estimate the mean of this distribution which is then used to adjust the conditional mean function upwards. This method is not guaranteed to envelope the data making it difficult for the researcher to interpret a firm which remains above the shifted regression function. Additionally, a distributional assumption is required to determine the appropriate 'shift' making the method heavily reliant on the particular distribution chosen and on the functional form of the conditional mean. These methods, popular in their day, were applied in a variety of settings (see Førsund & Hjalmarsson 1979*a*, Førsund & Hjalmarsson 1979*b*); however their appearance in empirical studies fell out of favor after the stochastic frontier literature gained momentum following the seminal work of Aigner et al. (1977), Battese & Corra (1977) and Meeusen & van den Broeck (1977). Both COLS and MOLS belong to a wider class of methods which could be called 'regression based envelopment'. One of the methods we propose below also fits into this class of 'shifted' techniques.

DEA is appealing because the method does not require distributional assumptions on producer inefficiency nor does it require a prespecified functional form. Recently, Kuosmanen & Johnson (2009) proposed an indirect deterministic frontier estimation procedure termed 'corrected concave nonparametric least squares' (C^2NLS) as the DEA inspired alternative to COLS. Like COLS, it too shifts the conditional mean upwards by the maximum residual, however, unlike COLS it is based on a nonsmooth nonparametric (piecewise linear) estimate of the conditional mean. This approach is appealing because, in addition to providing estimates of both inefficiency and the parameters of the production function, it imposes monotonicity and concavity on the production frontier as required by economic theory (positive marginal product and diminishing marginal returns). This approach is implemented in a quadratic programming framework that imposes monotonicity and concavity on the estimated production function, followed by the standard COLS correction to obtain the frontier and the estimates of firm level inefficiency. Additionally, Kuosmanen & Johnson (2009) show that C^2NLS is the DEA equivalent of a nonparametric least squares regression. One undesirable aspect of both DEA and C^2NLS is that while the frontier is continuous, it is nonsmooth at a finite number of points. This lack of differentiability across the support of the covariates is an undesirable feature of the estimator, particularly if one is interested in marginal effects (i.e., in measuring the responsiveness of output to changes in the input mix). Moreover, elasticities of substitution (Allen-Uzama, Morishima) are not calculable as they rely on second derivatives of the estimated production function. One could use a piecewise quadratic approach in place of piecewise linear, but the point we would like to emphasize is the additional flexibility afforded by the use of smooth nonparametric methods. The deterministic frontier approaches we advocate are twice continuously differentiable, unlike their C^2NLS and DEA counterparts.[1]

[1]See the work of Daraio & Simar (2005) and Cazals, Florens & Simar (2002) for recent work that generalizes the standard approach to conducting DEA. These methods involve calculating conditional probability functions and as such can be smooth.

Stochastic frontier analysis (SFA) constitutes a popular and widely used alternative to DEA. Unlike DEA, SFA does not envelope the data, rather, it proceeds by estimating a conditional mean model (i.e., like its 'shifted' deterministic frontier counterparts COLS, MOLS and C^2NLS, it is based on a model of average rather than maximum output) but with non-classical presumptions made about the model's residuals. SFA, unlike DEA, presumes that both noise ('stochastic error') and technical inefficiency are integral parts of the production process. By placing restrictions on the nature of the noise and inefficiency processes, inefficiency estimates can be backed out (Jondrow, Lovell, Materov & Schmidt 1982). Classical SFA methods are parametric in nature and involve functional form assumptions about the unknown conditional mean along with distributional assumptions about the inefficiency and the noise. Those who worry about potential misspecification of the conditional mean often gravitate towards methods that are robust to parametric misspecification; see by way of example Fan, Li & Weersink (1996) who proposed a semiparametric kernel-based SFA estimator and Kumbhakar, Park, Simar & Tsionas (2007) for a local maximum likelihood approach. Neither of these approaches, however, is constrained to satisfy general production axioms which limits their applicability. Yatchew (1985) provides a heuristic argument about how to use nonparametric methods to construct an optimization problem to estimate a nonparametric deterministic frontier, but no formal estimator is developed.

Before proceeding, a few words on constrained production frontiers are in order. Though concavity is a widely accepted classical theoretical property of a production function (Chambers 1988), practitioners frequently estimate models (e.g., the translog) which, by definition, cannot be concave without destroying their flexibility (Ryan & Wales 2000). The failure to verify whether concavity is satisfied using well-established tests (Hanoch & Rothschild 1972) is not uncommon; see the discussion in O'Donnell & Coelli (2005, p. 495) on the importance of imposing and checking theoretical constraints when estimating production relationships. Functional forms which can accommodate both long and short run behavior have been proposed in the literature, and by construction are not globally concave. By way of example, Duggal, Saltzman & Klein (1999, p. 47) note that their production function has a form which "allows for an 'S'-shaped production function which embodies not only the properties of a long-run production function but also those exhibited in the short run." We therefore wish to alert the reader to the fact that, though our method is capable of imposing concavity (in addition to a range of other constraints), in applied production settings it may be unwise to impose concavity without further investigation (Duggal, Saltzman & Klein 2007). Of course, it is trivial to drop this restriction while maintaining (weak) monotonicity.

In this paper we propose three kernel-based frontier estimators that satisfy requisite axioms of production and are continuously differentiable. Two are deterministic frontier estimators while the third is a stochastic frontier estimator. All three of these methods incorporate general axioms of production by exploiting recent advances in constrained kernel estimation; see Hall & Huang (2001) and Racine, Parmeter & Du (2009) for details. They therefore represent smooth nonparametric generalizations of Aigner & Chu's (1968) goal programming approach, Winsten's (1957) COLS method, and Aigner et al.'s (1977) stochastic frontier approach. The first deterministic frontier method envelopes the data directly and is termed 'smooth goal programming' (SGP) and may be of interest to those currently using DEA approaches. The second deterministic frontier method 'corrects' a smooth nonparametric conditional mean function and is termed 'smooth corrected programming' (SCP), and may be of interest to those currently using corrected deterministic frontier methods in the sense outlined above. As such, the second approach can be thought of as the smooth counterpart of Kuosmanen & Johnson (2009). The third method is a constrained version of Fan et al. (1996) but where the resulting estimator is guaranteed to satisfy general production axioms and is termed 'smooth stochastic frontier' (SSF), and this method may be of interest to those using flexible stochastic frontier methods. Additionally, we show how concavity can be imposed using simple linear constraints as opposed to more computationally demanding nonlinear constraints (O'Donnell & Coelli 2005, Henderson & Parmeter 2009), which constitutes an extension to constrained kernel methods not done elsewhere that may be of general interest.

Amsler, Lee & Schmidt (2009, p. 22) comment "Of course we can always estimate a regression consistently by purely nonparametric methods like kernels or nearest neighbors, but there ought to be advantages of imposing the restrictions that economic theory dictates." We demonstrate that indeed there are substantial advantages to imposing restrictions that economic theory dictates on nonparametric frontier methods, thus the constrained nonparametric kernel estimators we propose address one of the outstanding issues in applied frontier analysis.[2]

[2]Amsler et al. (2009) continue, "We predict that in the foreseeable future the methodology will exist for routine application of the stochastic frontier model without a parametric specification of the frontier."

The outline of the paper is as follows. Section 2 outlines the restricted nonparametric estimators upon which our proposed smooth constrained frontier estimators are based. Section 3 formally describes our approach and establishes the requisite theoretical properties. Section 4 provides several Monte Carlo simulations designed to examine the finite-sample performance of the methods while Section 5 provides an illustrative example. Section 6 provides concluding remarks.

2. Constrained Nonparametric Regression

The three frontier methods we outline below require as input a nonparametric model constrained to obey axioms of production theory. We adopt the approach of Racine et al. (2009) and direct the reader to that article for theoretical underpinnings and further details. Below we provide a brief sketch of their approach for the interested reader.

In what follows we let $\{x_i, y_i\}_{i=1}^n$ denote sample pairs of inputs and outputs and x a point of support at which we evaluate the frontier. Our goal is to nonparametrically estimate the unknown production frontier $m(x)$ subject to constraints on $m^{(\mathbf{s})}(x)$ where $\mathbf{s}$ is a k-vector corresponding to the dimension of x. The elements of $\mathbf{s}$ represent the order of the partial derivative corresponding to each element of x. Thus $\mathbf{s} = (0, 0, \ldots, 0)$ represents the function itself, while $\mathbf{s} = (1, 0, \ldots, 0)$ represents $\partial m(x)/\partial x_1$. In general, for $\mathbf{s} = (s_1, s_2, \ldots, s_k)$ we have

$$m^{(\mathbf{s})}(x) = \frac{\partial^{s_1} m(x)}{\partial x_1^{s_1}}, \ldots, \frac{\partial^{s_k} m(x)}{\partial x_k^{s_k}}. \tag{1}$$

We consider the class of kernel regression smoothers that can be written as linear combinations of the output y_i, i.e.,

$$\hat{m}(x) = \sum_{i=1}^{n} n^{-1} A_i(x) y_i, \tag{2}$$

which is a very broad class. For instance, the local constant or Nadaraya-Watson estimator uses

$$A_i(x) = \frac{n K_\gamma(x_i, x)}{\sum_{j=1}^{n} K_\gamma(X_j, x)}, \tag{3}$$

where $K_\gamma(\cdot)$ is a generalized product kernel that admits both continuous and categorical inputs, and γ is a vector of bandwidths; see Racine & Li (2004) for details.

In order to impose constraints on a nonparametric frontier, we shall require a nonparametric estimator that satisfies constraints of the form

$$l(x) \leq \hat{m}^{(\mathbf{s})}(x) \leq u(x) \tag{4}$$

for arbitrary $l(\cdot)$, $u(\cdot)$, and $\mathbf{s}$, where $l(\cdot)$ and $u(\cdot)$ represent (local) lower and upper bounds, respectively.

The constrained estimator is obtained by introducing an n-vector of weights p chosen so that the resulting estimator satisfies (4). We define the constrained estimator to be

$$\hat{m}(x|p) = \sum_{i=1}^{n} p_i A_i(x) y_i, \tag{5}$$

such that (4) is satisfied. Construction of (5) proceeds as follows. Let p_u be an n-vector with elements $1/n$ and let p be the vector of weights to be selected. In order to impose our constraints, we choose $p = \hat{p}$ to minimize the distance from p to the uniform weights $p_i = 1/n \ \forall i$ using the distance metric $D(p) = (p_u - p)'(p_u - p)$. The constrained estimator is then obtained by selecting those weights p that minimize $D(p)$ subject to constraints such as those given in (9), (10) and (11) below, which can be cast as a general nonlinear programming problem. For the constraints we need to impose (frontier behavior, monotonicity and concavity) we will have inequalities that are linear in p, which can be solved using standard quadratic programming methods and off-the-shelf software.[3] The appropriate bandwidth(s) for our unknown function can be estimated using any of the commonly available data-driven procedures and require estimation of the unrestricted function only. For notational simplicity we shall drop the '$|p$' notation with the understanding that the constrained estimator is that defined in (5) above.

2.1. A Digression on Derivative Estimation.

Before proceeding further we note that when one wishes to impose constraints on the surface $\hat{m}(x)$, such as monotonicity, the 'direct' derivatives may not be sufficient in finite samples. That is, the estimated derivatives that come directly from local polynomial estimation, say the local linear method, will not *exactly* correspond to the analytical

[3]For example, in the `R` language it is solved using the `quadprog` package, in GAUSS it is solved using the `qprog` command, and in MATLAB the `quadprog` command. Even when n is quite large the solution is computationally fast using any of these packages.

derivatives from direct computation from $\hat{m}(x)$, i.e., $\partial\hat{m}(x)/\partial x$. We therefore caution users about this issue as we suspect this result is not widely known.

By way of example, consider the local linear estimator of $m(x)$ and its derivatives,

$$\hat{\delta}(x) = \min_{\delta(x)} (\mathcal{Y} - \mathcal{X}\delta(x))'\mathcal{K}(x)(\mathcal{Y} - \mathcal{X}\delta(x)),$$

where $\mathcal{Y}$ is an $n \times 1$ vector with ith component y_i, $\mathcal{X}$ is an $n \times (1 + d)$ matrix with ith row $(1,\, (x_i - x)')$ and $\mathcal{K}(x)$ is the $n \times n$ diagonal matrix having ith diagonal element $K_\gamma(x_i, x)$. The vector $\delta(x)$ contains the conditional mean evaluated at x (first component) as well as the d first order derivatives of $m(x)$ (the 2 through $d+1$ components). The vector $\hat{\delta}(x)$ is a consistent estimator for $m(x)$ and its d first order derivatives (Li & Racine 2007, Theorem 2.7).

However, consider the exact form of the estimator for $\hat{m}(x)$ which, for the univariate case, can be shown to be

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} \left(S_2(x) - S_1(x)(x_i - x)\right) K_h(x_i, x) y_i}{S_2(X)S_0(x) - S_1(x)^2},$$

where

$$S_j(x) = \sum_{i=1}^{n} (x_i - x)^j K_h(x_i, x)$$

and $K_h(x_i, x) = h^{-1}K((x_i - x)/h)$ for a selected kernel function $K(\cdot)$; see Fan & Gijbels (1996). We can rewrite our unconstrained local linear estimator of the conditional mean as

$$\hat{m}(x) = \sum_{i=1}^{n} A_i(x) y_i,$$

where

$$A_i(x) = \frac{\left(S_2(x) - S_1(x)(x_i - x)\right) K_h(x_i, x)}{S_2(X)S_0(x) - S_1(x)^2} = \frac{B(x)}{D(x)}.$$

For example, if one wished to impose monotonicity on $\hat{m}(x)$ then one would need the derivative of $A_i(x)$, which in this case is

$$\frac{dA_i(x)}{dx} = \frac{D(x)K_h(x_i, x)R_1(x) + D(x)R_2(x)dK_h(x_i, x) - B(x)R_3(x)}{D(x)^2}, \tag{6}$$

where

$$
\begin{aligned}
R_1(x) =& S_1(x) + T_2(X) - [S_0(x) + T_1(x)(x_i - x)] \\
R_2(x) =& S_2(x) - S_1(x)(x_i - x) \\
R_3(x) =& S_0(x)T_2(x) + S_2(x)T_0(x) - 2S_1(x)T_1(x)
\end{aligned}
$$

and

$$
T_j(x) = \sum_{i=1}^{n} (x_i - x)^j dK_h(x_i, x).
$$

Comparing the analytical derivative of $\hat{m}(x)$ based on (6) with that arising directly from the local linear estimator (element two of $\hat{\delta}(x)$),

$$
\frac{\sum_{i=1}^{n} \left(S_0(x)(x_i - x) - S_1(x)\right) K_h(x_i, x) y_i}{D(x)},
$$

reveals that in finite samples these quantities will not be identical for finite non-zero h.

The upshot is that we caution users that imposing restrictions on the derivatives obtained directly from, say, the local linear estimator may result in a surface that in fact is not consistent with the constraints, however, this can be avoided by use of analytical derivatives of the fitted regression function $\hat{m}(x)$ itself.

We now deploy the constrained estimator outlined above for smooth constrained nonparametric estimation of both deterministic and stochastic frontier models.

3. Smooth Constrained Nonparametric Frontier Estimation

The starting point for modeling production frontiers is

$$
y_i = m(x_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{7}
$$

where y_i represents output, x_i a k-vector of inputs, $m(\cdot)$ the frontier (i.e., maximum output given x), and ε_i is either one-sided technical inefficiency (deterministic frontier) or a two-part composed error term (stochastic frontier) consisting of a one-sided term representing inefficiency (u_i) and a two-sided term representing statistical noise (v_i) such that $\varepsilon_i = v_i - u_i$.

We shall first consider constrained nonparametric extensions of two popular approaches used to estimate deterministic frontiers, and then proceed to estimate constrained semiparametric stochastic frontiers. Naturally, the constraints will be those dictated by the axioms of production theory.

3.1. **Constrained Nonparametric Deterministic Frontiers.** The deterministic frontier approach models ε_i as a one-sided process, i.e. $\varepsilon_i \leq 0$. One could estimate the frontier directly using programming type methods such as DEA or, by specifying the functional form for $m(\cdot)$, could proceed indirectly via COLS, which necessarily places joint restrictions on ε_i and $m(\cdot)$. We briefly outline COLS in a linear production function setting by way of illustration. Presume that the deterministic frontier model is

$$y_i = \alpha + x_i'\beta + \varepsilon_i, \quad i = 1, \ldots, n, \tag{8}$$

and that one estimates this model ignoring the error structure to obtain estimates $\hat{\alpha}$ and $\hat{\beta}$. Since $E[\varepsilon_i|x_i] \neq 0$, $\hat{\alpha}$ is a biased estimate of α. Moreover, the production function will not lie above all of the data which is inconsistent with the notion that $m(x_i) = \alpha + x_i'\beta$ is a frontier. To remedy this COLS corrects the estimated intercept to guarantee that the adjusted frontier function does indeed lie above the observed outputs. That is $\hat{\alpha}_c = \hat{\alpha} + \max \hat{\varepsilon}_i$ and the residuals are corrected in the opposite direction, $\hat{\varepsilon}_i^c = \hat{\varepsilon}_i - \max \hat{\varepsilon}_i$. Greene (1980) shows that this procedure will produce consistent estimates for α and ε_i^c *presuming the model is correctly specified and given that a set of regularity conditions hold*. Unfortunately, if the one-sided errors are heteroscedastic the conditions of Greene are violated and the COLS estimator will not be consistent. That is, when ones faces heteroscedastic error terms the conditional mean will not approach the conditional maximum (the 'frontier') via a constant (everywhere) shift. Clearly this limits the utility of this and of all such shifted approaches.

One could resort to maximum likelihood methods but in this case the one-sided distribution must satisfy additional regularity conditions in order to address asymptotic efficiency concerns. Specifically, the additional conditions required by Greene for maximum likelihood estimation to be viable given an exclusively one-sided error term are that the contact point on the end of the inefficiency distribution be zero, $f_\varepsilon(0) = 0$ and $\frac{\partial f_\varepsilon(\varepsilon)}{\partial \varepsilon}|_{\varepsilon=0} = 0$, where $f_\varepsilon(\cdot)$ is the distribution of the one-sided error. These conditions may appear stringent at first glance but as noted by Greene (1980)[page 40] "...these place virtually no restriction on the sorts of empirical models typically specified for production frontiers." In this simple setting, if the parametric model is correctly specified, one can consistently estimate inefficiency along with all parameters of the production function. However, in applied settings one may worry about misspecification, or one may test for correct parametric specification and reject one's model. In either case, models that are robust to misspecification would have obvious appeal.

In what follows we propose two alternate nonparametric deterministic frontier estimators, one that bounds the data and (unlike DEA) is smooth everywhere, requiring only one stage, and one that can be considered the smoothed version of the estimator proposed by Kuosmanen & Johnson (2009). Due to the fact that this setup will allow us to incorporate restrictions on the frontier and on its derivatives, imposing bounds, monotonicity and concavity is straightforward. Let our constrained estimator defined in (5) satisfy the following restrictions:

$$\sum_{i=1}^{n} p_i A_i(x_i) y_i - y_i \geq 0, \tag{9}$$

$$\sum_{i=1}^{n} p_i \left[\sum_{\mathbf{s} \in \mathbf{S}_1} A_i^{(\mathbf{s})}(x) \right] y_i \geq 0, \tag{10}$$

$$\sum_{i=1}^{n} p_i \left[\sum_{\mathbf{s} \in \mathbf{S}_2} A_i^{(\mathbf{s})}(x) \right] y_i \leq 0, \tag{11}$$

where $\mathbf{S}_1$ is

$$\Big[\; (1, 0, \dots, 0) \quad (0, 1, \dots, 0) \quad \cdots \quad (0, 0, \dots, 1) \; \Big]_k ,$$

while $\mathbf{S}_2$ is

$$\Big[\; (2, 0, \dots, 0) \quad (0, 2, \dots, 0) \quad \cdots \quad (0, 0, \dots, 2) \; \Big]_k .$$

These three conditions guarantee that the estimated frontier lies (weakly) above all observed output while respecting monotonicity and *necessary* conditions for concavity. This direct one-step estimator can be thought of as the smooth, nonparametric variant of Aigner & Chu's (1968) parametric goal programming approach which we term $\hat{m}^{SGP}(x_i)$ ('smooth goal programming').

Note that for $k = 1$ the above conditions are both necessary and sufficient for concavity, however, for $k \geq 2$ they may not be sufficient. To ensure that sufficiency is met we shall, in addition, impose Afriat's (1967) conditions. Note that if, instead, one were to focus attention on the matrix of second derivatives this would involve nonlinear constraints which are more complicated, computationally speaking, than Afriat's (1967) approach (the Afriat conditions are linear in the constraint weights).[4] Assuming that the unknown frontier is first order differentiable,[5] Afriat's (1967) conditions state that a function is (globally) concave if and only if

$$m(\mathbf{z}) - m(\mathbf{x}) \leq \frac{\partial m}{\partial x_1}(\mathbf{x})(z_1 - x_1) + \cdots + \frac{\partial m}{\partial x_k}(\mathbf{x})(z_k - x_k), \quad \forall \mathbf{z}, \mathbf{x}. \tag{12}$$

In our framework these inequalities can be handled directly without resorting to nonlinear constrained optimization, though of course this does not reduce the number of overall inequalities that must be imposed (a total of $n \times (n-1) \times k$ inequalities for $k \geq 2$).[6]

Powerful and flexible nonsmooth conditional mean-based deterministic frontier methods that satisfy requisite constraints have been proposed (Kuosmanen & Johnson 2009, Kuosmanen & Kortelainen 2009). The following corrected method will produce a smooth counterpart to Kuosmanen & Johnson (2009). Thus, as an alternative to the SGP estimator proposed above, if one is willing to impose a set of regularity conditions on the inefficiency and base the frontier on a conditional mean, then a simple two step estimator can be constructed by estimating the mean production function in equation (5) subject to the constraints in (10) and (12) (equation (11) would suffice if $k = 1$). Having estimated the constrained conditional mean (which we shall call $\hat{g}(x_i)$ to distinguish it from the frontier estimator $\hat{m}(x_i)$) we can obtain residuals (i.e., $\hat{\varepsilon}_i = y_i - \hat{g}(x_i)$) and then use $\max \hat{\varepsilon}_i$ to correct our estimate. The following procedure will yield our SCP ('smooth corrected programming') deterministic frontier estimator.

(i) First, estimate the conditional mean $\hat{g}(x_i)$ imposing the constraints in (10) and (12) (or (11)) and obtain the residuals, $\hat{\varepsilon}_i$.

---

[4]See Henderson & Parmeter (2009) for a detailed exposition on imposing concavity using the Hessian in this setting.
[5]Note that the class of nonparametric estimators considered herein also rely on this assumption.
[6]The use of the Afriat conditions to impose concavity is not uncommon and is used by Matzkin (1991) in a utility theoretic context and also by Kuosmanen & Johnson (2009) in the nonsmooth production context.

(ii) Second, shift the estimated conditional mean so that it envelopes the data, i.e. construct

$$\hat{m}^{SCP}(x_i) = \hat{g}(x_i) + \max_i \hat{\varepsilon}_i.$$

(iii) Finally, calculate estimates of producer inefficiency,

$$\hat{\varepsilon}_i^{SCP} = \hat{m}^{SCP}(x_i) - y_i = \hat{\varepsilon}_i - \max_i \hat{\varepsilon}_i.$$

We drop the lower bound constraint on $\hat{g}(x_i)$ since Kuosmanen & Johnson (2009, Theorem 4.2) show that the discriminatory power of C^{2}NLS is greater than that of a DEA estimator. The reason for this is as follows; input values in either the extreme lower or upper end of the support tend to reflect the frontier estimator downward, resulting in a biased estimate of a firm's efficiency level. The two-step procedure that corrects the entire frontier is not impacted to the same degree as the one step method since all observations are used in the smoothing. Below we show that this relationship holds between our SGP and SCP estimators.

3.2. **Constrained Semiparametric Stochastic Frontiers.** Unlike the fully nonparametric deterministic approach outlined above, the approach we now outline for stochastic frontiers is, strictly speaking, a semiparametric method as it relies on parametric structure for the composed error distribution. Smooth estimation of a stochastic frontier was proposed by Fan et al. (1996). They note that standard maximum likelihood methods are infeasible when one does not specify (i.e., parameterize) the production function. They further note that direct nonparametric estimation of the conditional mean would result in a biased estimate when one ignores the inefficiency term. Fan et al.'s (1996) solution is to correct this (downward) bias by retaining standard distributional assumptions from the SFA literature (e.g., normal noise, half-normal inefficiency) and estimating the corresponding distributional parameters via maximum likelihood on the nonparametric residuals from a standard kernel regression. Once these parameters are determined, the estimated conditional mean can be shifted (bias-corrected) by the estimated mean of the inefficiency distribution (mean correction factor). Under weak conditions Fan et al. (1996) show that the parameters of the composed error distribution can be estimated at the parametric $\sqrt{n}$ rate. Their simulations reveal that the semiparametric method produces estimates of the distributional parameters that are competitive with the same distributional parameter estimates produced from correctly specified production frontiers in the standard maximum likelihood framework. One drawback of their approach, however, is that they do not constrain the estimator so that it satisfies general axioms of production. As such, there is room for improvement.

A key distinction between the previous work of Fan et al. (1996) and that proposed here is that the proposed semiparametric estimator is guaranteed to satisfy the theoretical axioms of producer theory (production, cost, profit, etc.). This is especially important if one is interested in returns to scale or technical change. For example, returns to scale is defined as the sum of input elasticities (which are to be non-negative), and it is essential that these restrictions are satisfied at all data points. In an unconstrained semiparametric setting this is not guaranteed to be the case. Furthermore, empirical results may not be of much use for policy purposes if, for example, the scale measure is defined at the mean (which may not be indicative of any producer in the sample), or if production restrictions are violated for individual producers. These situations could arise in an unconstrained semiparametric framework which underscores the importance of deploying constrained nonparametric estimation in a production setting.

Our approach to estimating stochastic frontiers follows directly along the lines of Fan et al. (1996) thereby affording the researcher the same flexibility that the estimator of Fan et al. (1996) provides, but in addition we constrain the resulting stochastic frontier to satisfy general axioms of production as was done for the two deterministic approaches defined above. This is achieved by replacing the unknown conditional mean in Fan et al. (1996, (13), page 462) with one based upon (5) defined above. No further changes are necessary, and all results of Fan et al. (1996) follow without modification.

Fan et al. (1996) assume that the noise follows a mean zero normal distributive law and that the technical inefficiency stems from a half-normal distributive law. Given these presumptions and given the constrained estimate (5) one would construct the smooth constrained stochastic semiparametric frontier model as follows:

(i) Compute the (constrained) smooth conditional expectation, $E[y_i|x_i]$, as described above and call this $\hat{g}(x_i)$. Let the residuals be denoted $\hat{\varepsilon}_i = y_i - \hat{g}(x_i)$.

(ii) Define the concentrated variance of the composed error term $\sigma^2(\lambda)$ as a function of $\lambda = \sigma_u/\sigma_v$, $\sigma^2 = \sigma_u^2 + \sigma_v^2$, as follows:

$$\hat{\sigma}^2(\lambda) = \frac{n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{1 - \frac{2\lambda^2}{\pi(1+\lambda^2)}}. \tag{13}$$

(iii) Define the mean correction factor $\mu(\lambda)$ as a function of λ, i.e.,

$$\hat{\mu}(\lambda) = \frac{\sqrt{2}\hat{\sigma}\lambda}{\pi(1+\lambda^2)}. \tag{14}$$

(iv) Estimate λ by maximizing the concentrated log likelihood function consistent with the presumed distributional assumptions. In this setting we have

$$\hat{\lambda} = \max_{\lambda}\left(-n\ln\hat{\sigma}(\lambda) + \sum_{i=1}^{n}\ln(\Phi(-\tilde{\varepsilon}_i\lambda/\hat{\sigma}(\lambda)) - (2\hat{\sigma}^2(\lambda))^{-1}\sum_{i=1}^{n}\tilde{\varepsilon}_i^2\right), \tag{15}$$

where $\tilde{\varepsilon}_i = \hat{\varepsilon}_i - \hat{\mu}(\lambda)$ and $\Phi(\cdot)$ is the cumulative distribution function for a standard normal random variate.

(v) The constrained smooth stochastic production frontier $m(x_i)$ is consistently estimated by

$$\hat{m}^{SSF}(x_i) = \hat{g}(x_i) + \hat{\mu}, \tag{16}$$

where $\hat{\mu} = \sqrt{2}\hat{\sigma}\hat{\lambda}/\left(\pi(1+\hat{\lambda}^2)\right)$ and where $\hat{\sigma} = n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^2/\left(1 - \frac{2\hat{\lambda}^2}{\pi(1+\hat{\lambda}^2)}\right)$. See Fan et al. (1996) for further details.

Again, the sole difference between the approach of Fan et al. (1996) and $\hat{m}^{SSF}(x_i)$ defined above is that, in addition to being a semiparametric smooth estimate, $\hat{m}^{SSF}(x_i)$ will satisfy the axioms of production which it inherits from $\hat{g}(x_i)$ above. We show below that this difference is non-trivial.

3.3. **Theoretical Properties.** We now outline some elementary properties of the proposed estimators.

Theorem 3.1. *For independent and identically distributed inefficiency terms, $\varepsilon_1, \ldots, \varepsilon_n$, which are uncorrelated with the covariates X, if $f(\varepsilon_i) > 0$ at $\varepsilon_i = 0$, the SCP efficiency estimator is consistent. That is*

$$\operatorname*{plim}_{n\to\infty} \hat{\varepsilon}_i^{SCP} = \varepsilon_i, \quad \forall i = 1, \ldots, n.$$

Proof of Theorem 3.1. Letting $\mu = E[\varepsilon_i]$, Racine et al. (2009, Theorem 2.2(i)) guarantees that $\hat{\varepsilon}_i$ (obtained from the first stage) is a consistent estimator for $\varepsilon_i - \mu \;\forall\; i$ since $E[\varepsilon_i - \mu] = 0$. The arguments in Greene (1980, pgs. 32–34) show that $\operatorname*{plim}_{n\to\infty} \varepsilon_{(1)} = 0$, where $\varepsilon_{(1)}$ is the first order statistic of ε. This implies that $\operatorname*{plim}_{n\to\infty} \hat{\varepsilon}_{(1)} = \mu$. Therefore, $\operatorname*{plim}_{n\to\infty} \hat{\varepsilon}_i^{SCP} = \operatorname*{plim}_{n\to\infty} \hat{\varepsilon}_i - \operatorname*{plim}_{n\to\infty} \max \hat{\varepsilon}_j = (\varepsilon_i - \mu) - \operatorname*{plim}_{n\to\infty} \hat{\varepsilon}_{(1)} = \varepsilon_i$. □

This result implies that our corrected procedure will produce consistent estimates of producer inefficiency when regularity conditions on the inefficiency distribution are met. Moreover, this result implies that our efficiency estimates are asymptotically unbiased as well, albeit only in an iid framework. Additionally, in a comment to Schmidt (1985), Yatchew (1985, proposition 1) proves consistency of a nonparametric deterministic frontier assuming compact support of the vector of inputs and that the unknown function comes from a family of functions which are equicontinuous and bounded. Consistency is also obtained by replacing the equicontinuity assumption with an appropriate Lipschitz condition on the family of functions.

Theorem 3.2. *We have that i)* $\hat{\varepsilon}_i^{SCP} \leq \hat{\varepsilon}_i^{SGP} \leq 0 \;\forall i$ *and ii)* $\hat{\varepsilon}_i^{SCP} \leq \hat{\varepsilon}_i^{SSF} \;\forall i$.

Proof of Theorem 3.2. For i), suppose not. Then for an arbitrary firm j, $\hat{\varepsilon}_j^{SCP} \geq \hat{\varepsilon}_j^{SGP}$. Since our nonparametric goal programming method envelopes the data and both the SGP and SCP frontiers are concave there must exist at least one firm such that $\hat{\varepsilon}_i^{SGP} = 0$ and $\hat{\varepsilon}_i^{SCP} \geq 0$. However, this implies that $\hat{\varepsilon}_i^{SCP} = \hat{\varepsilon}_i - \max \hat{\varepsilon}_j \geq 0$ which implies that $\hat{\varepsilon}_i \geq \max \hat{\varepsilon}_j$ which is a contradiction.

For ii) we see immediately that since $\hat{\varepsilon}_i^{SCP} = \hat{\varepsilon}_i - \max \hat{\varepsilon}_j$ and $\hat{\varepsilon}_i^{SSF} = \hat{\varepsilon}_i - \hat{\mu}(\lambda)$, $\hat{\varepsilon}_i^{SCP} \leq \hat{\varepsilon}_i^{SSF}$ since the estimator of the mean of the one-sided distribution cannot be larger than the largest composed error residual. Moreover, since both $\hat{m}^{SCP}(x)$ and $\hat{m}^{SSF}(x)$ employ identical first stage estimates, they differ only in the amount of their (upward) correction. □

Theorem 3.2 states that our two stage corrected deterministic estimator lies everywhere above our single stage goal programming approach. This result does not, however, provide insight into the efficiency (statistically speaking) of the estimates which we therefore investigate via Monte Carlo simulation in the next section. Additionally, because the SCP estimator (as well as COLS, MOLS and C^{2}NLS) is based upon a (shifted) conditional mean, there is no guarantee that the microeconomic features of interest (returns to scale (henceforth RTS), technical change, elasticities of substitution) are equivalent among methods. This can be seen immediately in Figure 1. Here we have generated data from a single input frontier with one-sided inefficiency and fit the model using both SGP and SCP. If we were to shift the average production frontier so that it encapsulated all of the data, it would severely distort estimates of inefficiency.

The second part of Theorem 3.2 states that our SSF estimator is everywhere below our SCP estimator, which is intuitive since the SSF estimator and the SCP estimator use identical first stage estimates. We cannot generalize a result between the SSF estimator and the SGP estimator without further assumptions on the error terms (composed or not). Even though both the SGP and SSF estimators have the same smoothness constraints imposed (monotonicity and concavity), the SGP constraint is imposed at the frontier while the SSF constraint is imposed at the mean thus the shape of the SGP and the SSF curves could well differ. Further, this implies that there could exist a crossing of the SGP and SSF frontiers which rules out a strict relationship between their inefficiency estimates.

The point here is that while these estimators may be consistent, the behavior of shifted average estimators may not be consistent with direct frontier estimators in finite-sample settings. Unfortunately, the fact that SGP and DEA bound the data leaves them susceptible to outliers. While Timmer (1971) proposed an *ad hoc* upper bound to mitigate the presence of outliers in the data and though robust DEA measures have been developed, the basic issue of how one reconciles estimates from a frontier with those of an average in the presence of extreme observations remains. Kuosmanen & Johnson (2009, pages 17-18) note that "In contrast to DEA, however, all observations influence the shape of the C^2NLS frontier. Thus, a single outlier located above the frontier does not distort the shape of the C^2NLS frontier as severely as in DEA. Further, C^2NLS utilizes the information inefficient observations contain about the frontier." We note that SGP contains the same desirable features of C^2NLS, i.e. all observations influence the shape of the frontier (via local smoothing) but the presence of outliers can (and will) distort both the SGP and SCP estimators described here as well as the C^2NLS estimator. Again, these estimators are designed for estimation of deterministic frontiers and data with substantial noise should not be modelled using these methods. We do, however, wish to point out that kernel methods have recently been proposed that

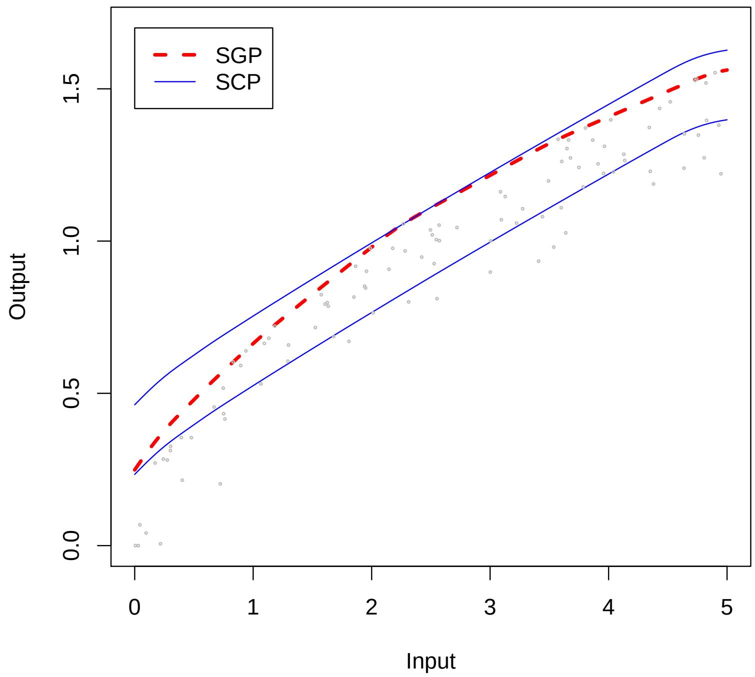

FIGURE 1. Deterministic Frontier. Conditional mean-based (SCP) versus maximal (SGP) output Estimation. The top two lines represent frontier estimates, the bottom line the conditional mean.

admit outliers and these approaches might be applicable in such instances; we direct the interested reader to Leung (2005) and the references therein.

Figure 2 presents the counterpart to Figure 1 for a stochastic frontier. In Figure 2 we compare the proposed smooth corrected semiparametric frontier (SSF) versus the smooth uncorrected semiparametric frontier (‘U-SSF’) of Fan et al. (1996). It is evident that imposing the requisite constraints of production theory can bring the semiparametric estimate in line with basic production axioms (one can observe negative marginal productivity estimates present in the U-SSF estimator, for example, i.e., a negative slope of the U-SSF frontier for some firms).

Theorem 3.3. *Given conditions 1 through 9 of Hardle & Stoker (1989, Appendix A.1) as well as condition 2.3 in Li & Racine (2007) we have the following:*

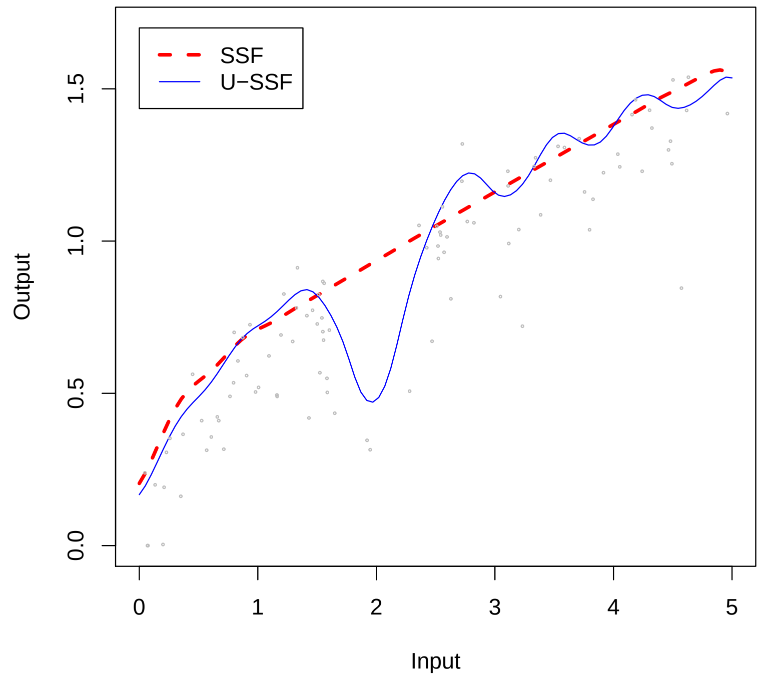

FIGURE 2. Stochastic Frontier. Smooth corrected semiparametric frontier (SSF) versus the smooth uncorrected semiparametric frontier (U-SSF).

(i) $\hat{m}^{SSF}(x_i)$ *is uniformly consistent with rate* $O\left((\ln(n)/(nh^k))^{1/2} + h^{p+1}\right)$*, where* p *is the order of the polynomial used to estimate the unknown conditional mean. In the local constant setting* $p = 0$*.*

(ii) We have that $\hat{\sigma}^2$ *and* $\hat{\lambda}$ *are* $\sqrt{n}$*-consistent estimates.*

Proof of Theorem 3.3. For part i) we direct the reader to Theorem 6 of Masry (1996). The only difference is the imposition of valid constraints which, as noted earlier, make the constrained and unconstrained estimators asymptotically equivalent; see Racine et al. (2009). The proof of ii) is trivial given the results of Fan et al. (1996, p. 462) and is not repeated here. □

We point out that the only difference between the SSF estimator discussed here and that of Fan et al. (1996) is the use of the constrained kernel estimator for the former. The conditions in Hardle & Stoker (1989) are not violated by using constrained kernel regression. Hence, the results stated in Theorem 3.3 follow directly from existing results and are not reproduced here.

4. Finite-Sample Behavior

In this section we undertake a series of Monte Carlo experiments designed to assess the finite-sample performance of the proposed approaches. The same underlying models are used for both the deterministic and stochastic frontier simulations though with differential treatment of ε. For simulations we make use of the `R` environment for statistical computing (R Development Core Team 2009) and the `R` packages `np` (Hayfield & Racine 2008), `DEA` (Diaz-Martinez & Fernandez-Menendez. 2008), `cobs` (Ng & Maechler 2009) and `quadprog` (original by Berwin A. Turlach R port by Andreas Weingessel ¡Andreas.Weingessel@ci.tuwien.ac.at¿ 2007).

4.1. Deterministic Frontiers.

First, we restrict our attention to the two deterministic frontier estimators considered above as well as COLS using a correctly specified parametric model ('COLS'), COLS using an incorrectly specified linear parametric model ('L-COLS'), COLS using an incorrectly specified quadratic parametric model ('Q-OLS'), the DEA approach ('DEA'), and where appropriate the SGP model that does not impose concavity but does impose monotonicity ('M-SGP'). We consider the following data generating processes (DGPs):

(i) $m(x) = 3 + 4 \times \ln(x) + 3 \times \sqrt{x}$,
(ii) $m(x) = 3 + \text{pnorm}(x - 3.5)$.

For each experiment we consider sample sizes of $n = 200, 400, 600, 800$ while x is distributed independently $\mathcal{U}[1, 10]$. Our one-sided error is generated as $|N(0, \sigma^2_\varepsilon)|$. For each scenario we conduct $M = 1,000$ Monte Carlo simulations. Note that DGP i) is globally concave while DGP ii) uses pnorm$(\cdot)$, the Standard Gaussian CDF which delivers a frontier having a sigmoidal shape consistent with parametric specifications outlined in Duggal et al. (1999, p. 47).

To assess the finite-sample performance of the deterministic frontier estimators, we consider the above DGPs and let σ^2_ε equal 0.2, 0.4 and 0.8. We use the local linear[7] kernel estimator with bandwidths obtained via least squares cross-validation. For DGP i) our monotonicity and concavity constraints are imposed on a grid of 100 equally spaced points while for DGP ii) we impose monotonicity only. We report ratios of median mean square error (MSE) taken over all M Monte Carlo replications, with the numeraire being that for the SGP estimator. For each run MSE

[7]Results for the local constant estimator are qualitatively similar and are excluded for space considerations.

is calculated as the average squared difference between each estimators' fit and the true frontier values on the same set of grid points used to impose the constraints. Results for the deterministic frontier simulation are reported in Table 1.

4.2. **Stochastic Frontiers.** In order to assess the finite-sample performance of the SSF estimator we consider the same DGPS as above, but now add noise in addition to inefficiency. In our setting we use the same values of λ and σ^2 as Fan et al. (1996) did in their simulations.[8] Here we estimate the unrestricted nonparametric stochastic frontier as in Fan et al. (1996) ('U-SSF'), the proposed smooth constrained semiparametric stochastic frontier ('SSF') as well as shifted parametric conditional mean models, both correctly and incorrectly specified ('COLS', 'L-COLS' and 'Q-COLS', respectively), and where appropriate the SSF model that does not impose concavity but does impose monotonicity ('M-SSF'). As with our previous simulations we consider sample sizes of $n = 200, 400, 600, 800$. Our one-sided error is generated as $|N(0, \sigma_u^2)|$ while our two-sided error, generated independently from u and x is $N(0, \sigma_v^2)$. For each scenario we conduct 1,000 Monte Carlo simulations and let $(\lambda, \sigma^2) = (1.66, 1.88)$, $(1.24, 1.63)$ and $(0.83, 1.35)$.[9]

As before, we use the local linear estimator with bandwidths obtained via least squares cross-validation. Our monotonicity and concavity constraints are imposed on a grid of 100 equally spaced points. We report the ratio between each estimator's MSE against that for the SSF estimator where the median is taken over all $M = 1,000$ replications. For each run MSE is calculated as the average squared difference between each of the estimators and the true frontier values on the same set of grid points used to impose the constraints.[10] Results for the stochastic frontier simulation are reported in Table 2.

4.3. **Discussion.** First, consider results for the deterministic frontier case summarized in Table 1. Of the two direct nonparametric frontier estimators (SGP and DEA), the proposed SGP approach dominates except in quite small samples. Of the shifted methods (SCP, COLS, L-COLS, Q-COLS), the proposed SCP estimator improves dramatically over the misspecified linear and quadratic COLS

[8] These are also identical to the values employed by Aigner et al. (1977).

[9] This corresponds to $\sigma_u^2 = 1.379$, 0.901, 0.536 and $\sigma_v^2 = 0.500$, 0.339, 0.294, respectively.

[10] Note that frontier behavior is evaluated at the sample realizations, not the grid points, and the DEA estimator is evaluated at the sample realizations for all constraints.

Table 1. Deterministic Frontier Monte Carlo. Ratio of median MSE for each estimator in the respective column heading relative to that for the SGP estimator. Numbers larger than 1 indicate superior MSE performance of the SGP method (results accurate to three significant digits).

DGP (i), Deterministic Frontier

	$m(x) = 3 + 4 \times \ln(x) + 3 \times \sqrt{x}$					
	SCP	COLS	L-COLS	Q-COLS	DEA	M-SGP
			$\sigma_u^2 = 0.2$			
200	1.01	0.147	250	36	0.833	1.48
400	1.68	0.181	603	91.9	0.879	1.44
600	2.62	0.245	1100	175	1.14	1.48
800	3.37	0.255	1590	263	1.27	1.45
			$\sigma_u^2 = 0.4$			
200	1.19	0.2	163	23	0.919	1.44
400	2.04	0.273	409	60.7	1.07	1.44
600	2.96	0.343	717	110	1.38	1.43
800	3.69	0.399	1040	162	1.54	1.41
			$\sigma_u^2 = 0.8$			
200	1.21	0.228	96	13.3	0.89	1.38
400	2.51	0.401	279	41.1	1.29	1.42
600	3.5	0.444	478	72.2	1.62	1.44
800	4.32	0.492	746	114	2.27	1.45

DGP (ii), Deterministic Frontier

	$m(x) = 3 + \text{pnorm}(x - 3.5)$			
	SCP	COLS	L-COLS	Q-COLS
		$\sigma_u^2 = 0.2$		
200	5.44	0.399	151	28.8
400	9.44	0.476	398	92.8
600	12.7	0.668	714	189
800	14.8	0.621	1030	300
		$\sigma_u^2 = 0.4$		
200	5.99	0.531	97.6	18.8
400	10.3	0.685	259	53.4
600	13.8	0.904	458	106
800	17.2	1.15	720	191
		$\sigma_u^2 = 0.8$		
200	6.05	0.593	56.4	11.8
400	11.1	0.889	164	35.9
600	13.7	1.04	274	62.3
800	18.1	1.27	427	100

TABLE 2. Stochastic Frontier Monte Carlo. Ratio of median MSE for each estimator in the respective column heading relative to that for the SSF estimator. Numbers larger than 1 indicate superior MSE performance of the SSF approach (results accurate to three significant digits).

DGP (i), Stochastic Frontier

	$m(x) = 3 + 4 \times \ln(x) + 3 \times \sqrt{x}$				
	U-SSF	COLS	L-COLS	Q-COLS	M-SSF
	$\sigma_u^2 = 0.536, \sigma_v^2 = 0.294$				
200	1.51	0.525	30.9	3.22	1.12
400	1.42	0.513	50.6	4.88	1.13
600	1.33	0.585	58.2	5.5	1.1
800	1.37	0.606	75.9	7.02	1.12
	$\sigma_u^2 = 0.901, \sigma_v^2 = 0.339$				
200	1.33	0.582	18.1	2.29	1.09
400	1.34	0.609	28.5	3.26	1.13
600	1.32	0.673	34	3.79	1.1
800	1.24	0.705	36.2	3.99	1.07
	$\sigma_u^2 = 1.379, \sigma_v^2 = 0.500$				
200	1.35	0.544	12.3	1.76	1.12
400	1.32	0.617	19.2	2.53	1.1
600	1.25	0.714	20.1	2.66	1.07
800	1.18	0.743	22.9	2.95	1.06

DGP (ii), Stochastic Frontier

	$m(x) = 3 + \text{pnorm}(x - 3.5)$			
	U-SSF	COLS	L-COLS	Q-COLS
	$\sigma_u^2 = 0.536, \sigma_v^2 = 0.294$			
200	1.25	0.527	1.95	0.919
400	1.27	0.578	2.68	1.07
600	1.24	0.615	3.27	1.14
800	1.12	0.672	3.32	1.23
	$\sigma_u^2 = 0.901, \sigma_v^2 = 0.339$			
200	1.23	0.596	1.53	0.888
400	1.22	0.676	1.99	1.02
600	1.17	0.718	2.23	1.08
800	1.14	0.794	2.2	1.09
	$\sigma_u^2 = 1.379, \sigma_v^2 = 0.500$			
200	1.19	0.683	1.3	0.937
400	1.24	0.711	1.6	0.956
600	1.18	0.755	1.75	1.02
800	1.17	0.826	1.89	1.06

estimators (L-COLS, Q-COLS) that are prevalent in applied settings. Furthermore, the nonparametric SGP estimator can even outperform a correctly specified parametric model as n increases which some may consider impossible (DGP ii), COLS). The key to interpreting these entries is to recognize that the proposed SGP estimator is a direct estimator of the frontier, while COLS and its ilk (including Kuosmanen & Johnson (2009)) involve shifting a conditional mean.[11] These entries simply highlight potential benefits of direct estimation of the frontier. Finally, imposing concavity where appropriate appears to improve on that imposing monotonicity only (M-SGP).

It is worth noting that the performance gains of SGP relative to DEA are to be expected given the theoretical results of Banker & Maindiratta (1988) which show that DEA delivers a lower bound on the family of production possibilities sets which rationalize the observed data. Given the concavity of the SGP estimator it cannot lie below the DEA estimator which may therefore result in improved estimates of the frontier.

Next, consider results for the stochastic frontier case summarized in Table 2. Recall that each of these methods involve shifting a conditional mean model, hence in this case the correctly specified parametric model cannot be beat. However, as n increases the proposed nonparametric method converges to the correctly specified parametric model for both DGPs considered (consider the COLS column as n increases). Furthermore, the proposed smooth SSF method outperforms the popular linear and quadratic specifications (except for small samples for DGP ii) for Q-OLS) with the relative performance improving as n increases. Finally, as expected, the restricted SSF estimator proposed here dominates the unrestricted nonparametric stochastic frontier ('U-SSF').

The simulation results comparing the performance of unrestricted and restricted nonparametric conditional mean models are also novel for the class of estimators proposed in Racine et al. (2009). While Hall & Huang (2001) did quantify the gains from imposing monotonicity on a conditional mean, their simulations were limited to a single simulated example. Our work here shows that imposing conditions consistent with economic theory can have large impacts on the performance of a nonparametric estimator, regardless of the context.

[11] Though we do not include Kuosmanen & Johnson's (2009) approach in our simulation, theirs is a 'shifted' (indirect) method that relies on a nonparametric estimate of the conditional mean, thus its performance would be comparable to the SCP approach and it too would be dominated by the (direct) SGP approach.

5. Application

To illustrate how our methods work in applied settings we use the classic production data of U.S. electricity companies in 1970 studied by Christensen & Greene (1976).[12] The data consist of a single output, millions of kilowatt hours of electricity generated (y), and three inputs, labor (l), capital (c) and fuel (f). Overall, though this data set it small ($n = 123$), it provides us with a well known setting in which we can evaluate the proposed methods. The dimensionality of the data is consistent with a large number of applied production studies; Kumbhakar & Tsionas (2009) analyze electricity generation using the same three inputs to the production process.

Our production frontier is

$$y_i = m(l_i, c_i, f_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{17}$$

which we estimate using a local constant estimator with cross-validated bandwidths. Our primary interest is in deviations from the frontier (inefficiency) and RTS. Since we are not using a logarithmic transformation, RTS is defined here as

$$\widehat{RTS}_i = \left(\frac{\partial \hat{m}(l_i, c_i, f_i)}{\partial l} \cdot l_i + \frac{\partial \hat{m}(l_i, c_i, f_i)}{\partial c} \cdot c_i + \frac{\partial \hat{m}(l_i, c_i, f_i)}{\partial f} \cdot f_i \right) \frac{1}{\hat{m}(l_i, c_i, f_i)}. \tag{18}$$

We use both of the frontier methods described above, SGP and SCP, as well as SSF to analyze RTS and inefficiency. We impose both monotonicity and concavity across the inputs and bandwidths are selected using least squares cross validation (Li & Racine 2004). We summarize the distribution of RTS and inefficiency for both methods by plotting their (smooth) CDFs, which are provided in Figures 3 and Figure 4, respectively.

Figure 3 displays the CDF of estimated RTS[13] using a standard unrestricted kernel estimator of the production function, a monotonically restricted (in all inputs) kernel estimator, the SCP estimator (which is equivalent to imposing monotonicity and concavity since the frontier is a neutral shift) and our SGP estimator (which also bounds the data). As noted above, the distribution of RTS differs considerably across the SGP and SCP methods. Also, it appears that imposing *both* monotonicity and concavity severely limits the RTS across all firms in the sample. This is

[12] This data is freely available in cost function form in the Ecdat library (Croissant 2006) in R (R Development Core Team 2009).

[13] We use Silverman's rule-of-thumb bandwidth for all smoothed distribution plots.

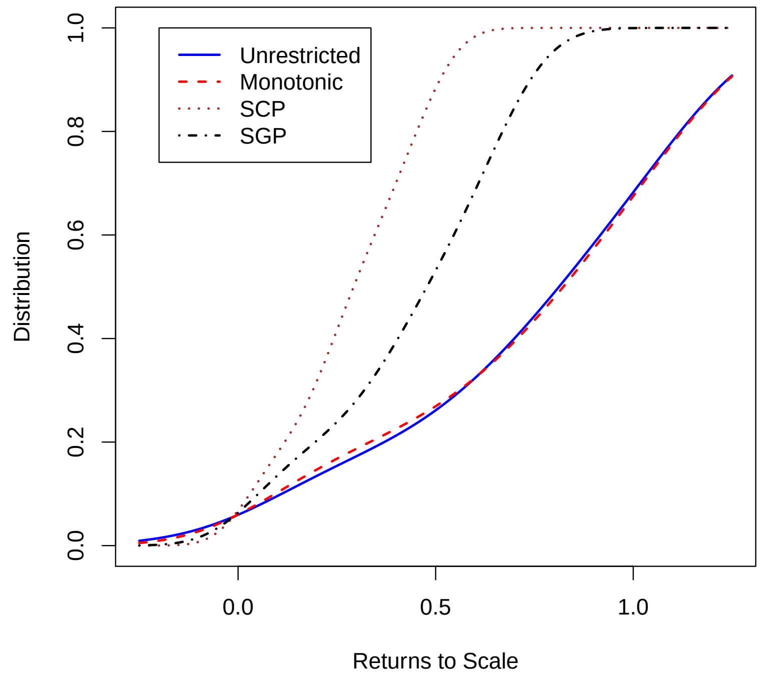

Figure 3. Distribution of RTS for four nonparametric production function/frontier estimators.

consistent with the parametric cost function findings of Christensen & Greene (1976) who find mostly decreasing RTS and in some cases diseconomies of scale.[14]

We note that a first order stochastic dominance relationship exists between the estimated RTS for the SGP and SCP estimators, while the unrestricted and monotonically restricted nonparametric production function produce nearly identical distributions of RTS. The imposition of concavity on our frontier produces a flatter estimate of the production frontier which is what produces the noticeably left shifted distributions of RTS relative to the estimators that do not impose concavity. It appears that almost all firms possess estimated RTS less than 0.5 when using the SCP estimator whereas roughly 60% of firms have estimated RTS less than 0.5 when using the SGP estimator. This is in contrast to the approximately 20% of firms who have RTS less than 0.5 using either a standard nonparametric conditional mean model or a monotonically constrained conditional mean

[14]We do not obtain any firms with diseconomies of scale since we impose monotonicity in all three of our inputs.

model. This is to be expected as concave functions are more severely shaped constrained than monotonically restricted estimators.

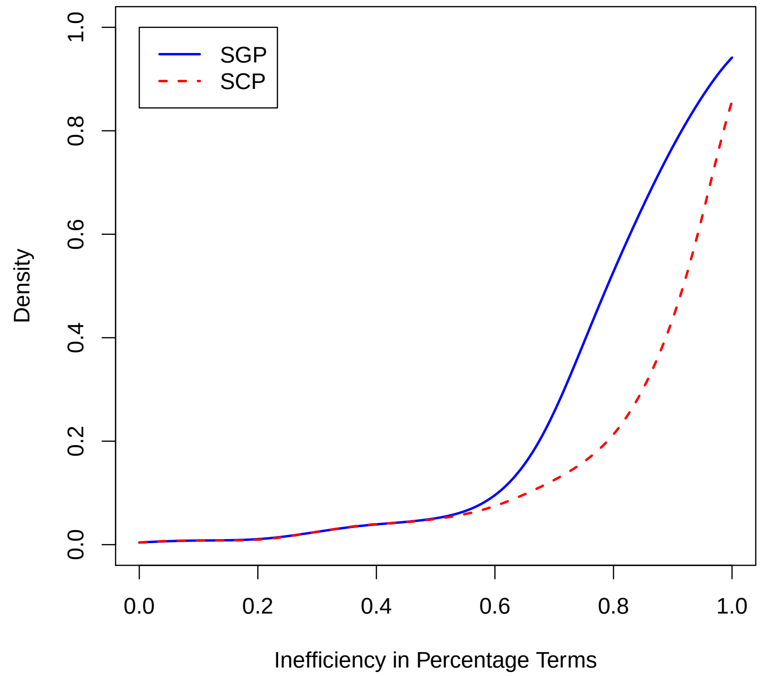

FIGURE 4. Distribution of estimates of inefficiency (% terms) for SGP and SCP.

The estimated CDFs of inefficiency for our SGP and SCP estimation routines presented in Figure 4 both tend to suggest that a majority of electricity plants are largely inefficient. Inefficiency estimates from the SCP estimator are stochastically dominated by those from the SGP estimator. What this dominance relationship suggests, along with our results from the distribution of estimated RTS, is that the SGP and SCP estimators provide different estimates of the frontier, which we highlighted earlier as a fundamental difference between direct estimation of the frontier and estimation of a conditional mean that is shifted to allow it to mimic a frontier. This is particularly noteworthy as both RTS and efficiency are measures routinely used by policy makers, thus estimator choice matters. Furthermore, SFA and COLS/MOLS, being conditional-mean based, are liable to the same critique as all such shifted estimators. Moreover, this result is to be expected in light of Theorem 3.2.

Both inefficiency distributions for our deterministic methods are plotted in Figure 4 and reveal that a large number of firms are highly inefficient. We report inefficiency measures in percentage terms, i.e. $(\hat{m} - y)/\hat{m}$, which implies the natural dominance relationship we observe and is consistent with our results from Theorem 3.2. What is missing from Figure 4 are the distributions of inefficiency for our stochastic methods, SSF and U-SSF. But for this application we obtain residuals with the wrong skewness (positive instead of negative) implying that estimation of σ^2 and λ is trivial as in these cases it is widely known (Olson, Schmidt & Waldman 1980) that $\hat{\lambda} = 0$.

Simar & Wilson (2009) show that even with one million observations, the probability of observing a random draw of composed errors with the wrong skewness is almost 50% when the variance ratio is set equal to 0.01. They also mention "...we know of no published papers reporting an estimate of zero for the variance parameter of the one-sided component in a stochastic frontier model." (Simar & Wilson (2009, p. 10)). A common response to this issue is to either select a different sample or to re-specify one's model, neither of which is attempted here. In our case re-specification is of no value as we are using methods robust to misspecfication. An alternative would be to resort to either local polynomial methods or ad hoc methods of bandwidth selection. However, this is not necessary. Simar & Wilson (2009) have proposed a simple bagging approach to handle samples that display the wrong skewness so that inference can still be conducted. We leave this as a topic for further study.

6. Concluding Remarks

Frontier methods provide powerful tools for assessing technical inefficiency among a sample of firms. The results from a frontier study have played important roles in guiding policy and assessing which firms are performing well (or poorly) in a given industry. Flexible estimation methodologies that impose as little structure as possible are axiomatically desirable, however, one can run the risk of not imposing sufficient structure. In this paper we propose a triad of methods for flexible frontier analysis that place minimal structure on the frontier while delivering a smooth continuously differential frontier that, in addition, satisfies requisite conditions dictated by basic axioms of production theory. We propose two deterministic frontier methods and one stochastic frontier method that each exploit recent developments in constrained kernel estimation techniques.

Our two deterministic alternatives have close links to the earlier goal programming literature as well as to the recent work on corrected concave nonparametric least squares.

A simulation study reveals that the methods perform remarkably well relative to their peers, while an empirical example illustrates the ease with which these methods can be employed. Additionally, the empirical example highlights the differences that can arise between the use of a shifted average production estimate versus an estimator that attempts to estimate the frontier directly. We find that, in a cross section of electricity generating plants, decreasing RTS is indicative of the entire sample while a majority of firms are inefficient.

We hope that these approaches are of interest to practitioners who worry about parametric misspecification and who are looking for smooth flexible alternatives that are consistent with basic production axioms.

References

Afriat, S. N. (1967), 'The construction of utility functions from expenditure data', *International Economic Review* **8**, 67–77.

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

Aigner, D. & Chu, S. (1968), 'On estimating the industry production function', *American Economic Review* **58**, 826–839.

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

Amsler, C., Lee, Y. H. & Schmidt, P. (2009), 'A survey of stochastic frontier models and likely future developments', *Seoul Journal of Economics* **22**, 5–27.

Arrow, K. J., Chenery, H. B., Minhas, B. S. & Solow, R. M. (1961), 'Capital-labor substitution and economic efficiency', *Review of Economics and Statistics* **63**(3), 225–250.

Banker, R. D. & Maindiratta, A. (1988), 'Nonparametric analysis of technical and allocative efficiencies in production', *Econometrica* **56**(6), 1315–1332.

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

Cazals, C., Florens, J. P. & Simar, L. (2002), 'Nonparametric frontier estimation: A robust approach', *Journal of Econometrics* **106**, 1–25.

Chambers, R. (1988), *Applied production analysis*, Cambridge University Press, Cambridge.

Christensen, L. R. & Greene, W. H. (1976), 'Economies of scale in u.s. electric power generation', *The Journal of Political Economy* **84**(1), 655–676.

Cobb, C. & Douglas, P. H. (1928), 'A theory of production', *American Economic Review* **18**, 139–165.

Croissant, Y. (2006), *Ecdat: Data sets for econometrics*. R package version 0.1-5.
URL: *http://www.r-project.org*

Daraio, S. & Simar, L. (2005), 'Introducing environmental variables in nonparametric frontier models: A probabilistic approach', *Journal of Productivity Analysis* **24**, 93–121.

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

Diaz-Martinez, Z. & Fernandez-Menendez., J. (2008), *DEA: Data Envelopment Analysis*. R package version 0.1-2.

Diewert, W. E. (1971), 'An application of the Shepard duality theorem: A generalized Leontief production function', *Journal of Political Economy* **79**(3), 481–507.

Duggal, V. G., Saltzman, C. & Klein, L. R. (1999), 'Infrastructure and productivity: a nonlinear approach', *Journal of Econometrics* **92**, 47–74.

Duggal, V. G., Saltzman, C. & Klein, L. R. (2007), 'Infrastructure and productivity: An extension to private infrastructure and it productivity', *Journal of Econometrics* **140**, 485–502.

Fan, J. & Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman & Hall/CRC.

Fan, Y., Li, Q. & Weersink, A. (1996), 'Semiparametric estimation of stochastic production frontier models', *Journal of Business & Economic Statistics* **14**(4), 460–468.

Farrell, M. J. (1957), 'The measurement of productive efficiency', *Journal of the Royal Statistical Society Series A, General* **120**(3), 253–281.

Førsund, F. R. & Hjalmarsson, L. (1979*a*), 'Frontier production functions and technical progress: A study of general mile processing in Swedish dairy plants', *Econometrica* **47**(4), 883–900.

Førsund, F. R. & Hjalmarsson, L. (1979*b*), 'Generalized Farrell measures of efficiency: An application to milk processing in Swedish dairy plants', *The Economic Journal* **89**(354), 294–315.

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

Hall, P. & Huang, H. (2001), 'Nonparametric kernel regression subject to monotonicity constraints', *The Annals of Statistics* **29**(3), 624–647.

Hanoch, G. & Rothschild, M. (1972), 'Testing the assumptions of production theory: a nonparametric approach', *Journal of Political Economy* **80**, 256–275.

Hardle, W. & Stoker, T. M. (1989), 'Investigating smooth multiple regression by the method of average derivatives', *Journal of the American Statistical Association* **84**(408), 986–995.

Hayfield, T. & Racine, J. S. (2008), 'Nonparametric econometrics: The np package', *Journal of Statistical Software* **27**(5).
URL: *http://www.jstatsoft.org/v27/i05/*

Henderson, D. J. & Parmeter, C. F. (2009), Imposing economic constraints in nonparametric regression: Survey, implementation and extension, *in* Q. Li & J. S. Racine, eds, 'Advances in Econometrics: Nonparametric Econometric Methods', Vol. 25, Emerald Group Publishing. Forthcoming.

Jondrow, J., Lovell, C. A. K., Materov, I. S. & Schmidt, P. (1982), 'On the estimation of technical efficiency in the stochastic frontier production function model', *Journal of Econometrics* **19**(2/3), 233–238.

Klein, L. R. (1947), The use of cross-section data in econometrics with application to the study of production of railroad services in the United States. Mimeo, National Bureau of Economic Research, New York.

Kumbhakar, S. C., Park, B. U., Simar, L. & Tsionas, E. G. (2007), 'Nonparametric stochastic frontiers: A local maximum likelihood approach', *Journal of Econometrics* **137**(1), 1–27.

Kumbhakar, S. C. & Tsionas, E. G. (2009), 'Stochastic error specification in primal and dual production systems', *Journal of Applied Econometrics* . Forthcoming.

Kuosmanen, T. & Johnson, A. (2009), 'Data envelopment analysis as nonparametric least squares regression', *Operations Research* . Forthcoming.

Kuosmanen, T. & Kortelainen, M. (2009), 'Stochastic nonparametric envelopment of data frontier estimation subject to shape constraints', *Journal of Productivity Analysis* . Forthcoming.

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

Leung, D. (2005), 'Cross-validation in nonparametric regression with outliers', *The Annals of Statistics* **33**, 2291–2310.

Li, Q. & Racine, J. S. (2004), 'Cross-validated local linear nonparametric regression', *Statistica Sinica* **14**, 485–512.

Li, Q. & Racine, J. S. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

Masry, E. (1996), 'Multivariate regression estimation: local polynomial fitting for time series', *Stochastic Processes and Their Applications* **65**, 81–101.

Matzkin, R. L. (1991), 'Semiparametric estimation of monotone and concave utility functions for polychotomous choice models', *Econometrica* **59**, 1315–1327.

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

Ng, P. T. & Maechler, M. (2009), *cobs: COBS – Constrained B-splines (Sparse matrix based)*. R package version 1.2-0.
URL: *http://wiki.r-project.org/rwiki/doku.php?id=packages:cran:cobs*

O'Donnell, C. J. & Coelli, T. J. (2005), 'A Bayesian approach to imposing curvature on distance functions', *Journal of Econometrics* **126**(2), 493–523.

Olson, J. A., Schmidt, P. & Waldman, D. A. (1980), 'A monte carlo study of estimators of stochastic frontier production functions', *Journal of Econometrics* **13**, 67–82.

original by Berwin A. Turlach R port by Andreas Weingessel ¡Andreas.Weingessel@ci.tuwien.ac.at¿, S. (2007), *quadprog: Functions to solve Quadratic Programming Problems.* R package version 1.4-11.

R Development Core Team (2009), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org*

Racine, J. S. & Li, Q. (2004), 'Nonparametric estimation of regression functions with both categorical and continuous data', *Journal of Econometrics* **119**(1), 99–130.

Racine, J. S., Parmeter, C. F. & Du, P. (2009), Constrained nonparametric kernel regression: Estimation and inference. Working Paper.

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

Ryan, D. L. & Wales, T. J. (2000), 'Imposing local concavity in the translog and generalized Leontief cost functions', *Economics Letters* **67**, 253–260.

Schmidt, P. (1985), 'Frontier production functions', *Econometric Reviews* **4**(2), 289–328.

Shepard, R. W. (1953), *Cost and production functions*, Princeton University Press.

Shepard, R. W. (1970), *The theory of cost and production functions*, Princeton University Press.

Simar, L. & Wilson, P. W. (2008), Statistical inference in nonparametric frontier models: recent developments and perspectives, *in* H. Fried, C. A. K. Lovell & S. S. Schmidt, eds, 'The Measurement of Productive Efficiency', 2 edn, Oxford University Press, Inc., chapter 4, pp. 421–521.

Simar, L. & Wilson, P. W. (2009), 'Estimation and inference in stochastic frontier models', *Econometric Reviews* . Forthcoming.

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

Winsten, C. B. (1957), 'Discussion on Mr. Farrell's paper', *Journal of the Royal Statistical Society Series A, General* **120**(3), 282–284.

Yatchew, A. J. (1985), 'Comment on 'Frontier production functions'', *Econometric Reviews* **4**(2), 345–352.