On Nonparametric Estimation: With a Focus on Agriculture

Citation	Färe, R., Grosskopf, S., Pasurka, C., & Martins-Filho, C. (2013). On Nonparametric Estimation: With a Focus on Agriculture. Annual Reviews of Resource Economics, 5(1), 93-110. doi:10.1146/annurev-resource-110811-114534
DOI	doi:10.1146/annurev-resource-110811-114534
Publisher	Annual Reviews
Version	Version of Record
Citable Link	http://hdl.handle.net/1957/46916
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

On Nonparametric Estimation: With a Focus on Agriculture

Rolf Färe,[1,2] Shawna Grosskopf,[1,*] Carl Pasurka,[3] and Carlos Martins-Filho[4,5]

[1]Department of Economics and [2]Department of Agricultural and Resource Economics, Oregon State University, Corvallis, Oregon 97331; email: rolf.fare@orst.edu, shawna.grosskopf@orst.edu

[3]Office of Policy, US Environmental Protection Agency, Washington, DC 20460; email: Pasurka.Carl@epa.gov

[4]Department of Economics, University of Colorado, Boulder, Colorado 80309-0256; email: Carlos.Martins@colorado.edu

[5]Markets, Trade and Institutions Division, IFPRI, Washington, DC 20006

Annu. Rev. Resour. Econ. 2013. 5:93–110

First published online as a Review in Advance on June 13, 2013

The *Annual Review of Resource Economics* is online at resource.annualreviews.org

This article's doi: 10.1146/annurev-resource-110811-114534

JEL codes: Q1, Q5, C61, O4

*Corresponding author

Keywords

data envelopment analysis, DEA, efficiency, productivity, linear programming

Abstract

We review nonparametric estimation of efficiency and productivity, by which we mainly mean activity analysis, or data envelopment analysis (DEA). The review covers topics that we hope will be of special interest to those doing research in the realm of agriculture. We also include a brief appendix addressing nonparametric estimation from an econometric perspective.

1. INTRODUCTION

The concept of nonparametric estimation is used in at least two different areas. One is nonparametric econometrics, and another is activity analysis, or data envelopment analysis (DEA). Here we discuss the latter but include an appendix by Carlos Martins-Filho that provides a brief summary of nonparametric density and nonparametric regression from an econometrician's point of view.

The most obvious advantage of our nonparametric approach relative to traditional parametric approaches is obvious from the name—no specification of a specific parametric functional form is required, nor is specification of error structures, thus avoiding possible specification error. Another advantage is the ability to estimate frontier functions consistent with economic theory such as minimum cost and maximum profit, among others. Although stochastic (parametric) frontier approaches also clearly allow for frontier estimation, they require specification of functional form, as well as assumptions concerning error structure, both of which may again introduce specification error. Another advantage of DEA-type approaches is the ease with which technologies with multiple outputs and inputs, which are typical in agriculture, can be estimated. Furthermore, these technologies can be estimated even in the absence of prices or costs, which as we see below proves useful when we wish to model joint production with undesirable outputs and useful for analysis in environmental and natural resource applications.

Our approach is to review the main models and ideas employed in nonparametric estimation, using linear programming methods rather than comprehensively reviewing specific papers or applications. We do, however, provide examples from agricultural economics of how the various models have been applied. Again, rather than including an exhaustive inventory of this applied literature, we include only representative examples.

After introducing what we refer to as the basic reference technology, we discuss the different optimization models associated with activity analysis, or DEA, as our nonparametric estimators, with examples from agricultural economics. Some topics, such as risk, uncertainty, and shadow pricing, are better suited to parametric formulations or econometric nonparametric estimation as, for example, in Kumbhakar & Tsionas (2009, 2010) and are not included here.

Our first variation on the basic model is the area of capacity and capacity utilization as formulated by Johansen (1968). As our agricultural example shows, this model can be used to estimate the von Liebig law of the minimum.

Productivity, especially the nonparametric Malmquist productivity index, is the topic of Section 5. We show how it is defined and estimated by using nonparametric tools. A contribution by Tauer & Lordkipanidze (2000) provides an example of how productivity may be applied in agriculture.

Our next topic is the joint production of good and bad outputs. We show how to modify our basic technology to accommodate both good and bad outputs in an activity analysis setup.

Our final topic is network DEA. This is a family of models that allow us to go inside the black box and examine the interactions among subtechnologies while preserving the nonparametric character of the DEA technology. These subtechnologies may reflect the connections among firms in an industry, processes within a firm, or connections over time, which can be used to form a dynamic network.

To take full advantage of our review, some knowledge of linear programming is useful. For those who seek guidance concerning software to implement the linear programming problems presented here, we refer the reader to Barr (2004), who reviews some options.

2. BASIC TECHNOLOGY AND ITS REPRESENTATION

In this section we introduce the basic activity analysis formulation of technology, which is perhaps best known under the name data envelopment analysis (DEA). (This name was coined by

Charnes et al. 1978.) We begin with some basics: We assume that there are $k = 1, \ldots, K$ observations or decision-making units (DMUs), which can be farms, firms, etc. Each DMU uses $x = (x_1, \ldots, x_N) \in \mathfrak{R}_+^N$ inputs to produce $y = (y_1, \ldots, y_M) \in \mathfrak{R}_+^M$ outputs. We follow Kemeny et al. (1956) and impose the following conditions on our data:

$$\begin{aligned} &(i) \sum_{m=1}^{M} y_{km} > 0, k = 1, \ldots, K, \quad (ii) \sum\nolimits_{k=1}^{K} y_{km} > 0, m = 1, \ldots, M, \\ &(iii) \sum_{n=1}^{N} x_{kn} > 0, k = 1, \ldots, K, \quad (iv) \sum\nolimits_{k=1}^{K} x_{kn} > 0, n = 1, \ldots, N. \end{aligned} \tag{1}$$

The first two conditions state that each DMU produces at least one type of output and that each output is produced by at least one DMU. Similarly, the last two conditions require that each DMU use at least one input and that each input be employed by at least one DMU. These conditions relax those originally required by von Neumann (1945) and Charnes et al. (1978) that the data have no zeros. These conditions are easy to verify from simple inspection of the data.

To continue with formulation of our basic activity analysis technology, we introduce what are known as intensity variables,

$$z_k \geqq 0, k = 1, \ldots, K, \tag{2}$$

i.e., one for each DMU. Using the above assumptions and our data matrix, we can specify our activity analysis technology set as

$$\begin{aligned} T = \Big\{ (x, y) : &\sum\nolimits_{k=1}^{K} z_k x_{kn} \leqq x_n, \quad n = 1, \ldots, N, \\ &\sum\nolimits_{k=1}^{K} z_k y_{km} \geqq y_m, \quad m = 1, \ldots, M, \\ &z_k \geqq 0, \quad k = 1, \ldots, K \Big\}. \end{aligned} \tag{3}$$

Alternatively, we can equivalently specify the technology as

$$P(x) = \{(y) : (x, y) \in T\} \quad \text{the output set} \tag{4}$$

or

$$L(y) = \{(x) : (x, y) \in T\} \quad \text{the input set,} \tag{5}$$

where clearly

$$x \in L(y) \Leftrightarrow (x, y) \in T \Leftrightarrow y \in P(x). \tag{6}$$

Given the conditions specified in Kemeny et al. (1956), one can prove that the activity analysis technology satisfies the following conditions (see Shephard 1970 for a proof):

I. $P(0) = \{0\}$, implying inactivity;
II. $P(x)$ is bounded for each $x \in \mathfrak{R}_+^N$, implying scarcity;
III. T is a closed set, and thus $P(x)$ and $L(y)$ are closed;
IV. if $y \in P(x)$ and $y' \leqq y$, then $y' \in P(x)$, implying strong disposability of outputs;
V. $x \in L(y)$, $x' \geqq x$ imply $x' \in L(y)$, resulting in strong or free disposability of inputs;
VI. $T = \lambda T$, $\lambda > 0$, implying constant returns to scale.

Note that Conditions II and III together imply that $P(x)$ [but not T or $L(y)$] is a compact set.

If we change the M inequalities on our outputs in our activity analysis technology above to strict equalities, the technology satisfies weak disposability of outputs, i.e.,

$$y \in P(x) \text{ and } 0 \leqq \theta \leqq 1, \text{ then } \theta y \in P(x). \tag{7}$$

In the same way, if the N input inequalities in our activity analysis technology are changed to equalities, the technology satisfies weak disposability of inputs, i.e.,

$$x \in L(y) \text{ and } \lambda \geqq 1, \text{ then } \lambda x \in L(y). \tag{8}$$

This implies that we are allowing for backward-bending isoquants. Strong disposability implies weak disposability, but the converse is not true.

We can also vary the returns-to-scale property in our activity analysis technology by adding constraints to the intensity variables, i.e., the z's: If $z_k \geqq 0$, $k = 1, \ldots, K$, and $z_k \leqq 1$, then the technology exhibits nonincreasing returns to scale (NIRS), and if $z_k \geqq 0, k = 1, \ldots, K$, and $z_k = 1$, then it exhibits variable returns to scale (VRS). See Färe & Grosskopf (2009) for more details regarding returns to scale and disposability of outputs and inputs.

3. OBJECTIVE FUNCTIONS

The nonparametric literature often seeks to identify a best result; i.e., we seek to optimize some objective or objective function. These are generally of two types: value functions and distance functions. Such functions are often dual to each other; the distance functions are associated with the primal or quantity-based specification, and the value functions are associated with dual or price-based specifications. The value functions are more common in the economics branch of this literature, whereas the distance-type functions are typical in the operations research branch.

As mentioned in Section 2, the frontier or boundary of our activity analysis technology is piecewise linear, which implies that the isoquant (approximately the boundary of the set) may not be the same as the efficient subset, i.e., the part of the boundary whose elements are nondominated. Particularly in the operations research literature, considerable effort has been expended to adjust the distance-type functions to attain the efficient subset. A typical example of a technology in which the efficient subset and isoquant do not coincide is the Leontief type:

$$L(y) = \{(x_1, x_2) : y \leqq \min\{x_1, x_2\}\}. \tag{9}$$

If the goal is to measure technical efficiency relative to the efficient subset, i.e., where $x_1 = x_2$, rather than the isoquant that includes the vertical and horizontal extensions of the points where $x_1 = x_2$, the typical distance function must be modified to remove slack, which we address presently.

But we begin with what are generally considered to be the first nonparametric efficiency measures that are due to Farrell (1957). To introduce them, consider the following now-classic figure based on Farrell's paper (**Figure 1**).

In **Figure 1**, the input set $L(y)$ is bounded below by its isoquant, here labeled yy. The objective in this case is to identify the cost-minimizing input bundle for observation A, given the input prices represented by the slope of the isocost ww. The optimal bundle is x^*, with associated minimum cost wx^*; i.e., this is a typical value function optimization. Farrell (1957) defines overall cost-efficiency as the ratio of minimal cost to observed cost, which for DMU A is equivalent to

$$0C/0A.$$

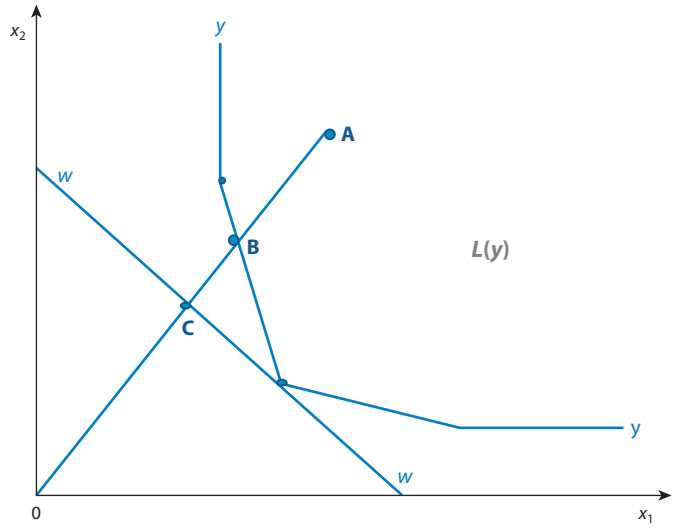

Figure 1

Farrell (1957) decomposition of cost-efficiency.

Farrell (1957) decomposes this (value-type) efficiency score into technical efficiency (0B/0A), which is the distance function or primal objective, and a residual allocative or price-related component (0C/0B), which altogether can be stated as

$$0C/0A = 0B/0A \cdot 0C/0B.$$

That is, overall cost-efficiency is the product of technical and allocative efficiency.

To formalize the Farrell measure and its decomposition, first define the cost function

$$C(y, w) = \min\{wx : x \in L(y)\}, \tag{10}$$

where $w \in \Re_+^N$ is a vector of input prices. This cost function can be estimated relative to our activity analysis technology as a linear programming problem, i.e.,

$$\begin{aligned} C(y^{k'}, w) = &\min_{x,z} wx \\ \text{subject to} \quad &\textstyle\sum_{k=1}^{K} z_k x_{kn} \leqq x_n, \quad n = 1, \ldots, N, \\ &\textstyle\sum_{k=1}^{K} z_k y_{km} \geqq y_{k'm}, \quad m = 1, \ldots, M, \\ &z_k \geqq 0, \quad k = 1, \ldots, K. \end{aligned} \tag{11}$$

This problem may be solved for each observation in the data set $k' = 1, \ldots, K$ and yields the minimum cost for each observation as well as the cost-minimizing input bundle. The dual to this cost function is Shephard's (1953) input distance function, which is defined as

$$D_i(y, x) = \max\{\lambda : x/\lambda \in L(y)\} \tag{12}$$

and may be estimated as the solution to the following linear programming problem:

$$\begin{aligned}\left(D_i(y^{k'}, x^{k'})\right)^{-1} &= \min \beta \\ \text{subject to} \quad &\sum_{k=1}^{K} z_k x_{kn} \leqq \beta x_{k'n}, \quad n = 1, \ldots, N, \\ &\sum_{k=1}^{K} z_k y_{km} \geqq y_{k'm}, \quad m = 1, \ldots, M, \\ &z_k \geqq 0, \quad k = 1, \ldots, K.\end{aligned} \tag{13}$$

The (dual) relation between the two functions can be captured by the following, which is based on the Mahler (1939) inequality, namely

$$\frac{C(y, w)}{wx} \leqq \frac{1}{D_i(y, x)}. \tag{14}$$

This inequality is derived as follows, first using the definition of the cost function as minimal cost,

$$C(y, w) \leqq wx \text{ for all } x \in L(y), \tag{15}$$

and because $x/D_i(y, x) \in L(y)$, we have

$$C(y, w) \leqq w\left(x/D_i(y, x)\right), \tag{16}$$

yielding our Mahler inequality from Equation 14 above. The Farrell decomposition may be derived from this inequality by introducing Farrell's allocative efficiency component, AE_i, and thus

$$\frac{C(y, w)}{wx} = \frac{1}{D_i(y, x)} \cdot AE_i, \tag{17}$$

where $1/D_i(y, x)$ is the Farrell measure of technical efficiency.

Farrell suggests that a similar decomposition can be obtained by taking an output-increasing orientation rather than an input-decreasing orientation. The associated value function in this case is the revenue function, for which we need to introduce output prices $p \in \Re_+^M$ and define

$$R(x, p) = \max\{py : y \in P(x)\}, \tag{18}$$

which has an associated dual distance function

$$D_o(x, y) = \min\{\theta : y/\theta \in P(x)\}. \tag{19}$$

From these functions, we can follow the derivation above to obtain a Farrell output-oriented decomposition as

$$\frac{R(x, p)}{py} = \frac{1}{D_o(x, y)} \cdot AE_o, \tag{20}$$

where $1/D_o(x, y)$ is the output-oriented technical efficiency component and AE_o the allocative efficiency component.

In addition to cost and revenue efficiency, we can also define profit or Nerlovian efficiency (introduced by Chambers et al. 1998), which is based on the profit function defined as

$$\Pi(p, w) = \max\{py - wx : (x, y) \in T\}. \tag{21}$$

Here, T is our basic activity analysis technology. From basic economics, we know that under constant returns to scale, $\Pi(p, w) = 0$ in equilibrium. To allow for nonzero profits, we need to

impose either NIRS or VRS. If NIRS are imposed, i.e., we add the restriction that $\sum_{k=1}^{K} z_k \leqq 1$, then $\Pi(p, w) \leqq 0$. If VRS are imposed, i.e., we add the restriction that $\sum_{k=1}^{K} z_k = 1$, then $\Pi(p, w) \in (-\infty, +\infty)$, which allows for losses as well as for nonnegative profits.

Dual to the profit function is the directional technology distance function, introduced by Luenberger (1995) as the shortage function and defined as

$$\overrightarrow{D}_T\left(x, y; g_x, g_y\right) = \sup\left\{\beta : \left(x - \beta g_x, y + \beta g_y\right) \in T\right\}. \tag{22}$$

Here (g_x, g_y) are the directional vectors, which give the direction in which (x, y) is projected onto the boundary of the technology set T.

From the definition of the profit function, we know that

$$\Pi(p, w) \geqq py - wx \quad \text{for all} \quad (x, y) \in T,$$

and if we take into account the fact that

$$\begin{aligned} &\left(x - \overrightarrow{D}_T(x, y; g_x, g_y) g_x, y + \overrightarrow{D}_T(x, y; g_x, g_y) g_y\right) \in T, \text{ then} \\ &\frac{\Pi(p, w) - (py - wx)}{pg_y + wg_x} \geqq \overrightarrow{D}_T\left(x, y; g_x, g_y\right). \end{aligned} \tag{23}$$

This inequality tells us that the normalized difference between maximum profit and observed profit (which we term the Nerlovian efficiency indicator[1]) is greater than or equal to the associated technical efficiency. As with our other value inequalities, we can arrive at a decomposition of profit efficiency by including a residual allocative component, i.e.,

$$NI = \frac{\Pi(p, w) - (py - wx)}{pg_y + wg_x} = \overrightarrow{D}_T\left(x, y; g_x, g_y\right) + \overrightarrow{AE}_T. \tag{24}$$

The directional distance function is in fact a generalization of the Shephard distance functions introduced in conjunction with the Farrell measures of technical efficiency. By the appropriate choice of the directional vectors, we can show these relationships. Thus, if $g_x = 0$, $g_y = y$, we have

$$\overrightarrow{D}_T(x, y; 0, y) = \left(1/D_o(x, y)\right) - 1, \tag{25}$$

and when $g_x = x$, $g_y = 0$, we have

$$\overrightarrow{D}_T(x, y; x, 0) = 1 - \left(1/D_i(y, x)\right). \tag{26}$$

Thus, the appropriate choice of directions allows us to move from indicators to indexes.

We now return to our Leontief technology,

$$L(y) = \{(x_1, x_2) : \min\{x_1, x_2\} \geqq y\},$$

which has L-shaped isoquants. The Farrell input technical efficiency measure equals one (i.e., signals technical efficiency) if and only if

[1]We follow Diewert (1998) and term additive measures indicators and refer to multiplicative measures like the Farrell measures as indexes.

$$\min\{x_1, x_2\} = y,$$

which can happen when an observation is on the horizontal or vertical extensions of the isoquant, rather than at the kink that is the efficient subset of this technology. The gap between the A and B in **Figure 2** represents what is known as the slack or, in this case, excess of input x_1 relative to the input usage at B. Clearly B uses fewer inputs yet produces the same output as observation A, so A is therefore not technically efficient—only observation B is efficient in **Figure 2**.

We can define both a multiplicative technical efficiency measure and an additive technical efficiency measure that will have the indication property that its value is one and zero, respectively, if and only if the observation is efficient. In **Figure 2**, efficiency occurs if the observation is at B.

The multiplicative index that identifies whether an observation is a member of the efficient subset of the technology (as at B) is introduced by Färe & Lovell (1978) as the Russell measure, which may be defined for **Figure 2** as

$$R(y,x) = \min\left\{\frac{\lambda_1 + \lambda_2}{2} : (\lambda_1 x_1, \lambda_2 x_2) \in L(y)\right\}. \tag{27}$$

The additive measure or indicator is a slack-based directional distance function introduced by Färe & Grosskopf (2010) and defined as

$$\max\{\beta_1 + \beta_2 : (x_1 - \beta_1 \cdot 1, x_2 - \beta_2 \cdot 1) \in L(y)\}. \tag{28}$$

Both of these may be estimated as linear programming problems by using our basic activity analysis technology from Section 2.

Agricultural applications of these concepts are too numerous to itemize here, but as an example we note that Weersink et al. (1990) employ the output-oriented Farrell measures using four variations on the basic technology to study the performance of Ontario dairy farms. They estimate the Farrell measure of technical efficiency under constant returns to scale and then decompose it into scale efficiency, congestion (due to weak disposability), and pure technical efficiency. This

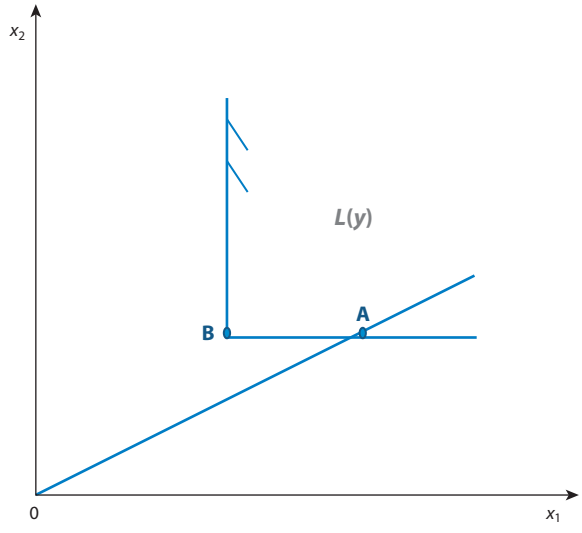

Figure 2

Slacks and efficiency.

analysis entails estimating efficiency relative to the VRS technology and technology that imposed weak input disposability. The decomposition reveals that the major source of inefficiency was due to deviation from the optimal scale.

4. CAPACITY AND THE LAW OF THE MINIMUM

The definition we employ for plant capacity is from Johansen (1968, p. 362): "...[C]apacity is the maximum amount that can be produced per unit of time with the existing plant and equipment, provided that the availability of variable factors of production is not restricted." When does this maximum exist? For a production unit to have this property, Färe (1984) shows that the fixed factor must satisfy so-called limitationality. This finding in turn leads us to the law of the minimum, which is associated with the name von Liebig. This law states that maximum production is controlled by the limiting factors, which intuitively are factors that are essential to production. Wang et al. (2006) provide an empirical example of how the capacity concept of Johansen can be used to study the role of micronutrients as potential limiting factors of production in a nonparametric framework.

To see how to identify limiting factors in the activity analysis framework, we first assume that output prices, $p \in \Re_+^M$, are known. (Other objective functions, such as output maximization, may also be used.) Then revenue maximization relative to our basic activity analysis technology may be estimated for DMU k' as

$$\begin{aligned} \max_{z,y} & \sum_{m=1}^{M} p_m y_m \\ \text{subject to } & \sum_{k=1}^{K} z_k x_{kn} \leqq \beta x_{k'n}, \quad n = 1, \ldots, N, \\ & \sum_{k=1}^{K} z_k y_{km} \geqq y_m, \quad m = 1, \ldots, M, \\ & z_k \geqq 0, \quad k = 1, \ldots, K. \end{aligned} \tag{29}$$

Now assume that x_1 is the unrestricted variable factor. Then we may compute maximum revenue as

$$\begin{aligned} \max_{z,y} & \sum_{m=1}^{M} p_m y_m \\ \text{subject to } & \sum_{k=1}^{K} z_k x_{kn} \leqq \beta x_{k'n}, \quad n = 2, \ldots, N, \\ & \sum_{k=1}^{K} z_k x_{k1} = x_1, \\ & \sum_{k=1}^{K} z_k y_{km} \geqq y_m, \quad m = 1, \ldots, M, \\ & z_k \geqq 0, \quad k = 1, \ldots, K. \end{aligned} \tag{30}$$

These revenue maximization problems differ in two ways:

- x_1 is free to vary in the second problem and is therefore an unrestricted variable input as in the Johansen (1968) definition, whereas it is restricted in the first problem.
- An equality rather than an inequality is used for the x_1 constraint in the second problem, which allows us to identify the optimal value of that input to be compared with the observed value. Thus, we can determine whether there is too much or too little of x_1 in terms of identifying capacity output.

As mentioned above, Wang et al. (2006) apply the Johansen (1968) capacity concept to measure the revenue efficiency of pear trees. They compute the optimal level of macronutrients, such as NO_3, as well as the optimal level of micronutrients such as zinc and iron.

5. PRODUCTIVITY

In this section we discuss productivity, in particular the Malmquist productivity index introduced by Caves et al. (1982). (For more detailed discussions of this topic, please refer to Färe et al. 2008.) In contrast to the more familiar Fisher (1922) and Törnqvist (1936) productivity indexes, this index does not require data on prices or shares to facilitate aggregation. The Malmquist index can be estimated by using our basic activity analysis model with optimization. The optimization provides the means of aggregation without additional price/value data.

To begin, we define the level of productivity very simply if we limit ourselves to the case of a single input and a single output; i.e., it is equivalent to the average product:

$$y/x. \tag{31}$$

If we wish to look at productivity change or growth between two periods, say $t = 0$, $t = 1$, we have

$$\frac{y^1/x^1}{y^0/x^0}. \tag{32}$$

Obviously the single-input, single-output case is unrealistic. Instead of aggregating by using prices or shares to get indexes of multiple inputs and outputs typical of the Fisher (1922) and Törnqvist (1936) approaches, we formulate productivity growth by using distance functions. We employ constant returns to scale to ensure consistency with our simple definition above. That assumption implies

$$P(\lambda x) = \lambda P(x),\ \lambda > 0 \Leftrightarrow D_o(\lambda x, y) = 1/\lambda D_o(y, x). \tag{33}$$

That is, under constant returns to scale, the output distance function is homogeneous of degree -1 in inputs. It is by definition homogeneous of degree $+1$ in outputs. Thus, we may now express ratios of productivity levels (average productivity) in terms of distance functions:

$$\frac{y^1/x^1}{y^0/x^0} = \frac{y^1/x^1 D_o(1,1)}{y^0/x^0 D_o(1,1)} = \frac{D_o(x^1, y^1)}{D_o(x^0, y^0)}. \tag{34}$$

Caves et al. (1982) define the $t = 0, 1$ Malmquist productivity change indexes as

$$\frac{D_o^0(x^1, y^1)}{D_o^0(x^0, y^0)} \quad \text{and} \quad \frac{D_o^1(x^1, y^1)}{D_o^1(x^0, y^0)}, \tag{35}$$

respectively. Here the superscript on D tells us the period of the reference technology, whereas the superscripts on x and y tell us from which period the data under evaluation are selected. One can show that the two indexes coincide if and only if the technology is Hicks neutral, i.e.,

$$D_o^t(x^t, y^t) = A(t) D_o(x^t, y^t). \tag{36}$$

This is a rather strong assumption, and hence one may follow Fisher (1922) and take the geometric mean of the two indexes. That is, here we define the Malmquist output productivity index as

$$M_0^1 = \left(\frac{D_o^0(x^1, y^1)}{D_o^0(x^0, y^0)} \frac{D_o^1(x^1, y^1)}{D_o^1(x^0, y^0)} \right)^{1/2}. \tag{37}$$

This productivity index can be multiplicatively decomposed into a catching-up component and a technical change component. The catching-up component tells us whether a DMU is getting

closer or farther from the best practice frontier over time, and it is also known as the efficiency change component:

$$EFFCH_0^1 = \frac{D_o^1(x^1, y^1)}{D_o^0(x^0, y^0)}. \tag{38}$$

The technical change component tells us whether and how far the frontier has shifted over time. It is estimated as

$$TECH_0^1 = \left(\frac{D_o^0(x^1, y^1)}{D_o^1(x^1, y^1)} \frac{D_o^0(x^0, y^0)}{D_o^1(x^0, y^0)} \right)^{1/2}. \tag{39}$$

In comparison, the classic Solow (1957) residual is equivalent to our technical change component because it does not explicitly allow for technical inefficiency.

The product of the two components forms the productivity index.[2] That is,

$$M_0^1 = EFFCH_0^1 \cdot TECH_0^1. \tag{40}$$

The Malmquist index and its components may be estimated by using our basic model. Note that we also have mixed-period distance functions

$$D_o^0(x^1, y^1) \text{ and } D_o^1(x^0, y^0). \tag{41}$$

Tauer & Lordkipanidze (2000, p. 24) apply the Malmquist productivity index defined above to determine whether "...productivity of a farmer may increase with age, reach some maximum level, and then decrease with further age." They find that productivity of farmers does follow this pattern (see p. 31). They also find that technical change is the most important component of productivity change for their sample.

Galanopoulos et al. (2004) apply the Malmquist index to evaluate productivity growth in the European Union during the 1990s. Their results indicate that productivity growth is attributed mainly to technical change over this period.

6. ENVIRONMENTAL DATA ENVELOPMENT ANALYSIS: GOOD AND BAD OUTPUTS

To create an environmental technology with good and bad outputs, our basic model must be extended. First we introduce a so-called bad output vector $b \in \Re_+^J$. Here by bad we mean that consumers are better off when less is produced or that (some other) firms are better off when less is produced. In contrast, so-called good outputs are those for which more is better. Second, we also need to know why and when bad outputs are produced. Here we see them as joint products with (or by-products of) the good outputs. As Baumgärtner et al. (2001, p. 365) state, "...the production of wanted goods gives rise to additional unwanted outputs ...." They base this conclusion on observations from thermodynamics. Our third consideration accounts for the fact that it may not be possible to dispose of bad outputs costlessly, whereas the traditional model does make that assumption for good outputs.

[2]Further decompositions are possible; see Färe & Grosskopf (1996).

More formally, we can now add axioms to our basic set to account for these three adjustments. Beginning with the fact that bad outputs are by-products of good production, we employ the axiom of null joint production, following Shephard & Färe (1974), which is defined as

$$\text{VII. if } (y,b) \in P(x) \text{ and } b = 0, \text{ then } y = 0,$$

where the output set is modified to read

$$P(x) = \{(y,b) : x \text{ can produce } (y,b)\}. \tag{42}$$

In words, if no bad output is produced, then no good output can be produced. Thus, the production of good output generates bads, just as there is no fire without smoke.

If we turn to disposability properties, the standard approach is to assume strong or free disposability of outputs. Thus, if we have good and bad outputs, this assumption would require that

$$\text{VIII. if } (y,b) \in P(x) \text{ and } (y',b') \leqq (y,b), \text{ then } (y',b') \in P(x).$$

As Førsund (2009, p. 10) points out, this yields the "...nonsensical results that zero bads can be achieved at no costs...."

Here we assume that only the good outputs are strongly disposable; i.e.,

$$\text{IX. } (y,b) \in P(x) \text{ and } y' \leqq y \text{ imply } (y',b) \in P(x).$$

We assume that the undesirable outputs—together with the good outputs—are jointly weakly disposable;[3] i.e.,

$$\text{X. if } (y,b) \in P(x) \text{ and } 0 \leqq \theta \leqq 1, \text{ then } (\theta y, \theta b) \in P(x).$$

This condition states that at the margin it is costly to dispose of bad outputs, given inputs, and that reduction in bads at the margin requires either diversion of some of the given inputs to abatement or reduction of overall production. In either case the effect is to also reduce good outputs, given inputs.

To generalize our basic model to allow for the above environmental aspects, we require that

$$\text{XI. } \sum_{j=1}^{J} b_{kj} > 0, k = 1, \ldots, K \text{ and } \sum_{k=1}^{K} b_{kj} > 0, j = 1, \ldots, J.$$

Condition XI states that each bad output must be produced by at least one DMU and that each DMU must produce at least some bad. This condition imposes null jointness, which can be verified by inspection of the data. We are now able to write our environmental nonparametric specification of technology as

$$\begin{aligned} P(x) = \Big\{ (y,b) : & \sum_{k=1}^{K} z_k y_{km} \geqq y_m, \quad m = 1, \ldots, M, \\ & \sum_{k=1}^{K} z_k b_{kj} = b_j, \quad j = 1, \ldots, J, \\ & \sum_{k=1}^{K} z_k x_{kn} \leqq x_n, \quad n = 1, \ldots, N, \\ & z_k \geqq 0, \quad k = 1, \ldots, K \Big\}. \end{aligned} \tag{43}$$

[3]Shephard (1970) introduces this property on the technology.

The $j = 1, \ldots, J$ constraints for the bad outputs are strict equalities, whereas those for the good outputs are inequalities. This scenario reflects the joint assumption of weak disposability of goods and bads and strong disposability of the goods.[4]

Ball et al. (2001) (see also Ball et al. 2004) apply this model in estimating productivity growth in the US agricultural sector. They apply their methods to state-level data recently made available by ERS, which include variables that proxy effects of pesticides and nitrates (found in fertilizers) on ground water and surface water. They employ a directional distance function as their objective with the technology constraints specified above. See Section 3 for a discussion of directional distance functions.

7. NETWORK MODELS

In the above sections we view the production process as essentially a black box, in which inputs enter the process and outputs emerge. In this section we focus on what may happen in the black box, especially in terms of interactions among possible subtechnologies that make up the process contained in the black box. These types of models do require more extensive data than our black-box models (for details, see Färe & Grosskopf 1996) but allow for greater flexibility—they allow us to explicitly include abatement technologies in our environmental case and allow for dynamic interactions over time, among many other applications.

We illustrate a network model in **Figure 3**. Our simple network technology consists of two subtechnologies, $P1$ and $P2$. This network has a source, which allocates the system exogenous input vector x to the subtechnologies in the box. There is also a sink, which collects the final outputs that exit the box. Our first subtechnology produces intermediate outputs y^{1i} as well as final outputs y^{1f}, which sum to

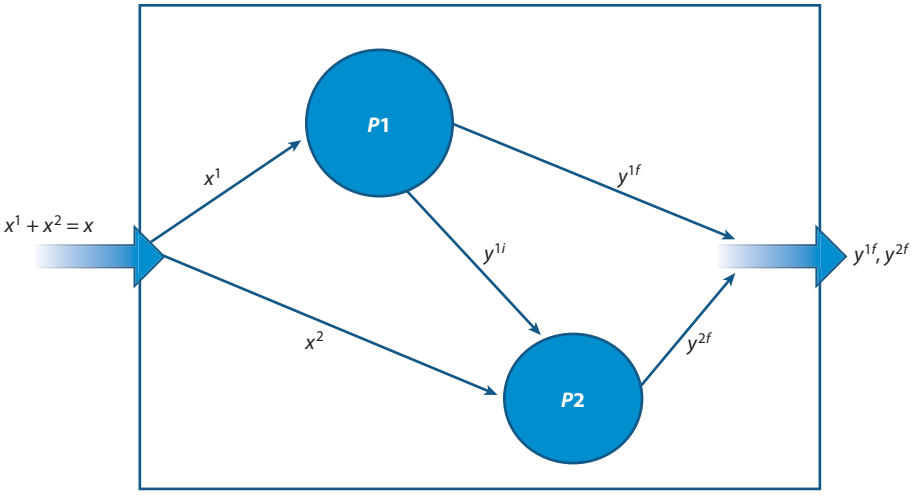

Figure 3

A network model.

[4]Førsund (2009), Rødseth (2011), and Murty et al. (2012) challenge this model.

$$y^1 = \left(y^{1i} + y^{1f}\right) \in P1\left(x^1\right). \tag{44}$$

The second subtechnology, $P2$, has inputs from two sources: the system exogenous allocation x^2 and the intermediate inputs from subtechnology $P1$, namely y^{1i}. Subtechnology $P2$ produces

$$y^2 \in P2\left(x^2, y^{1i}\right). \tag{45}$$

The two subtechnologies compete for the system exogenous inputs x such that $x = x^1 + x^2$. Thus, the network output set may be written as

$$\begin{aligned} P(x) = \{&\left(y^{1f}, y^{2f}\right) : \left(y^{1i} + y^{1f}\right) \in P1\left(x^1\right), \\ &y^{2f} \in P2\left(x^2, y^{1i}\right), \\ &x \geqq x^1 + x^2\}, \end{aligned} \tag{46}$$

with the interaction between the two technologies consisting of competing for system exogenous inputs and $P1$ delivering intermediate inputs to $P2$.

Suppose we wish to maximize revenue over the network model described above, i.e.,

$$\begin{aligned} &\max p^1 y^{1f} + p^2 y^{2f}: \ \left(y^{1i} + y^{1f}\right) \in P1\left(x^1\right), \\ &y^{2f} \in P2\left(x^2, y^{1i}\right), \\ &x \geqq x^1 + x^2. \end{aligned} \tag{47}$$

Then the solution yields, in addition to optimal final outputs (y^{1f^*}, y^{2f^*}),

1. optimal intermediate outputs y^{1i^*} and
2. optimal allocation of system exogenous inputs (x^{1^*}, x^{2^*}).

This network model allows us to study both the efficiency of the overall network as well as the efficiency of the subtechnologies. Of course, this model may be generalized to many subtechnologies.

To estimate our optimization over our network model, we can specify it as the following linear programming problem:

$$\begin{aligned} &\max p^1 y^{1f} + p^2 y^{2f} \\ &\text{subject to} \\ &\\ &\text{(subtechnology 1)} \\ &\textstyle\sum_{k=1}^K z_k^1 y_{km}^1 \geqq y_m^{1i} + y_m^{1f}, m = 1, \ldots, M, \\ &\textstyle\sum_{k=1}^K z_k^1 x_{kn}^1 \leqq x_n^1, n = 1, \ldots, N, \\ &z_k^1 \geqq 0, \ k = 1, \ldots, K, \\ &\\ &\text{(subtechnology 2)} \\ &\textstyle\sum_{k=1}^K z_k^2 y_{km}^{2f} \geqq y_m^{2f}, m = 1, \ldots, M, \\ &\textstyle\sum_{k=1}^K z_k^2 y_{km}^{1i} \leqq y_m^{1i}, m = 1, \ldots, M, \\ &\textstyle\sum_{k=1}^K z_k^2 x_{kn}^2 \leqq x_n^2, n = 1, \ldots, N, \\ &z_k^2 \geqq 0, \ k = 1, \ldots, K. \end{aligned} \tag{48}$$

Source:

$$x_n^1 + x_n^2 \leqq x_n, n = 1, \ldots, N.$$

Above, the objective is in the first line, the constraints for subtechnology $P1$ are in the top three inequalities, the constraints for subtechnology $P2$ are the next four constraints, and the overall allocation of system exogenous inputs is shown as the last constraint. Each subtechnology has its own intensity variables (z_k^1, z_k^2), and y^{1i} is an input to subtechnology $P2$. Finally, M, N, and K need not be the same for all technologies; i.e., they may have different numbers of outputs, inputs, and observations.

Jaenicke (2000) uses a variation of this model to study crop production that accounts for crop rotation. Specifically, his paper employs the network model as a dynamic DEA model. For another example of such a model, see Färe & Grosskopf (1996).

8. CONCLUSIONS

Our brief review of activity analysis, or DEA, as nonparametric estimators with connections to agricultural economics necessarily skims over or ignores several important aspects, including

1. the axiomatic underpinnings of the efficiency measures,
2. software available for the actual estimation of these measures, and
3. statistical underpinnings and hypothesis testing.

For the axiomatic underpinnings, we refer the reader to Russell, with a summary review in Russell & Schworm (2011). Regarding the last two points, we note that both freeware and commercial products are available, including the FEAR package for *R* developed by Wilson (2008). Simar & Wilson (2008) is also a good reference for statistical underpinnings and hypothesis testing.

9. APPENDIX: PARAMETRIC AND NONPARAMETRIC MODELS FROM AN ECONOMETRICIAN'S PERSPECTIVE, BY CARLOS MARTINS-FILHO

9.1. Density Estimation

Let $X(\omega): \Omega \to \mathfrak{R}$ be a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$ and $F(x) = P(\{\omega: X(\omega) \leq x\})$ for $x \in \mathfrak{R}$ be its distribution function. $F(x) \in \mathcal{F}$, where $\mathcal{F}$ is a class of functions with well-known properties (Jacod & Protter 2000). It is, of course, possible to restrict this class of functions to $\mathcal{F}_\theta \subset \mathcal{F}$, where $\mathcal{F}_\theta$ is a collection of distribution functions that is identified by a finite-dimensional parameter $\theta \in \Theta \subset \mathfrak{R}^K$. If one assumes that $P(\{\omega: X(\omega) \leq x\}) = F_\theta(x)$, then we speak of a parametric model. In these models, the estimation of $F_\theta(x)$ is equivalent to the estimation of the finite-dimensional parameter θ. The main advantage of parametric models is that, given a random sample $\{X_i\}_{i=1}^n$ of size n and the assumption that F_θ is absolutely continuous with density $f_\theta(x)$, one can construct a log-likelihood function

$$L(\theta) = \sum_{i=1}^{n} \log f_\theta(X_i)$$

and obtain a maximum likelihood estimator $\theta_m = \text{argmax}_\theta L(\theta)$ for θ. Under fairly general conditions, the asymptotic properties of θ_m as an estimator for θ are well developed (Cramér 1946, LeCam 1972). The most important such property is that $\sqrt{n}(\theta_m - \theta) \xrightarrow{d} N(0, v)$, where v is the

smallest variance attained in a class of estimators that converge uniformly in distribution to a normally distributed random variable at the $\sqrt{n}$ speed. This efficiency result, in essence, establishes that v attains the well-known Cramér-Rao lower bound. In summary, there is a well-established theory of estimation for parametric models.

The restriction that $F \in \mathcal{F}_\theta$ can often be incorrect. In addition, incorrect specification of the parametric class to which F belongs generally leads to estimators that are not $\sqrt{n}$ asymptotically normal and whose variances do not satisfy the Cramér-Rao lower bound. Hence, it is necessary to construct estimators for F when $F \in \mathcal{F}$, where $\mathcal{F}$ is a class of functions that cannot be described by a finite-dimensional parameter. The elements of $\mathcal{F}$ are normally restricted in other ways—e.g., they may be required to be absolutely continuous or smooth—but no assumption is made that they may be indexed by a finite-dimensional parameter θ. These types of models are termed nonparametric models. Under the assumption that $F \in \mathcal{F}$ are absolutely continuous, the estimation of their associated densities f, based on a random sample $\{X_i\}_{i=1}^n$, normally proceeds in one of two ways. The first assumes that f belongs to a class whose elements can be described by a collection of basis functions. These basis functions are then used to provide an estimator for f. Some of the estimators that result from these methods are spline, sieve, and series estimators (Efromovich 1999). The second way of estimating f involves local weighted approximations of f by polynomials of various orders. This type of estimators is termed kernel estimators (Tsybakov 2009).

The main advantage of nonparametric estimators is that they are by construction robust to the potential misspecification that $F \in \mathcal{F}_\theta$. Robustness to parametric misspecification comes, in general, at a cost. Mainly, nonparametric estimators of F or f do not converge in distribution at parametric rates, i.e., $\sqrt{n}$. Most importantly, the rate of convergence diminishes exponentially with the dimensionality of the estimation. That is, if $X(\omega)\colon \Omega \to \mathfrak{R}^K$, $K > 1$, the rate of convergence of nonparametric estimators diminishes exponentially with K.

9.2. Regression Estimation

In economics, there is great interest in the estimation of regression, i.e., $E(Y|X = x)$, where Y is a random variable and X is a K-dimensional random vector defined in a common probability space. If the conditional distribution of Y given $X = x$, denoted by $F_{Y|X=x}(y)$, is an element of $\mathcal{F}_\theta$, then the regression $E(Y|X = x) = \int y dF_{Y|X=x}(y)$ depends on θ, and we write $E(Y|X = x) := m(x;\theta) \in \mathcal{M}_\theta$, a parametrically indexed collection of measurable functions. This is a parametric model of regression, and estimation of m is equivalent to estimation of θ. Whenever $E(Y|X = x) := m(x) \in \mathcal{M}$, a class of functions that cannot be parametrically indexed by a finite-dimensional parameter θ, we speak of a nonparametric model of regression. The main advantages and disadvantages of estimating nonparametric models of regression mirror those described above in the case of density estimation. In particular, the exponentially decreasing rate of convergence with K alluded to above persists in the regression case. Further restrictions on the class $\mathcal{M}$, such as additivity of m, lead to nonparametric regression estimators that converge in distribution at nearly parametric ($n^{2/5}$) rates (Wang & Yang 2007). This type of restriction is especially relevant for empirical economics, in which large K are common.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Any errors, opinions, or conclusions are the authors' and should not be attributed to the US Environmental Protection Agency.

LITERATURE CITED

Ball VE, Färe R, Grosskopf S, Nehring R. 2001. Productivity of the U.S. agricultural sector: the case of undesirable outputs. In *New Developments in Productivity Analysis*, ed. CR Hulten, ER Dean, MJ Harper, pp. 541–86. Chicago: Chicago Univ. Press

Ball VE, Lovell CAK, Luu H, Nehring R. 2004. Incorporating environmental impacts in the measurement of agricultural productivity growth. *J. Agric. Res. Econ.* 29:436–60

Barr RS. 2004. DEA software tools and technology. In *Handbook on Data Envelopment Analysis*, ed. WW Cooper, LM Seiford, J Zhu, pp. 539–66. Boston: Kluwer

Baumgärtner S, Dykhoff H, Faber M, Proops J, Shiller J. 2001. The concept of joint production and ecological economics. *Ecol. Econ.* 36:365–72

Caves DW, Christensen LR, Diewert WE. 1982. The economic theory of index numbers and the measurement of input, output and productivity. *Econometrica* 50:1393–414

Chambers RG, Chung YH, Färe R. 1998. Profit, directional distance functions, and Nerlovian efficiency. *J. Optim. Theory Appl.* 98(2):351–64

Charnes A, Cooper WW, Rhodes E. 1978. Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* 2:429–44

Cramér H. 1946. *Mathematical Methods of Statistics*. Princeton: Princeton Univ. Press

Diewert WE. 1998. *Index number theory using differences rather than ratios*. Discuss. Pap. 98-10, Dep. Econ., Univ. B. C.

Efromovich S. 1999. *Nonparametric Curve Estimation*. New York: Springer

Färe R. 1984. The existence of plant capacity. *Int. Econ. Rev.* 25:209–13

Färe R, Grosskopf S. 1996. *Intertemporal Production Frontiers: With Dynamic DEA*. Boston: Kluwer

Färe R, Grosskopf S. 2009. A comment on weak disposability in nonparametric analysis. *Am. J. Agric. Econ.* 91:535–38

Färe R, Grosskopf S. 2010. Directional distance functions and slack-based measures of efficiency. *Eur. J. Oper. Res.* 200:320–32

Färe R, Grosskopf S, Margaritis D. 2008. Efficiency and productivity: Malmquist and more. In *The Measurement of Productive Efficiency and Efficiency Change*, ed. H Fried, CAK Lovell, S Schmidt, pp. 522–622. Oxford, UK: Oxford Univ. Press

Färe R, Lovell CAK. 1978. Measuring the technical efficiency of production. *J. Econ. Theory* 19:150–62

Farrell MJ. 1957. The measurement of productive efficiency. *J. R. Stat. Soc. A* 120:253–81

Fisher I. 1922. *The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability*. Boston/New York: Houghton Mifflin

Førsund F. 2009. Good modeling of bad outputs: pollution and multi-output production. *Int. Rev. Environ. Resour. Econ.* 3:1–38

Galanopoulos K, Karagiannis G, Koutroumanidis T. 2004. Malmquist productivity estimates for European agriculture in the 1990s. *Oper. Res.* 4:73–91

Jacod J, Protter P. 2000. *Probability Essentials*. Berlin: Springer

Jaenicke EC. 2000. Testing for intermediate outputs in dynamic DEA models: accounting for social capital in rotational crop production and productivity measures. *J. Prod. Anal.* 14:247–66

Johansen L. 1968. *Production Functions and the Concept of Capacity*. Oslo: Univ. Oslo

Kemeny JG, Morgenstern O, Thompson GL. 1956. A generalization of the von Neumann model of expanding economy. *Econometrica* 24:115–35

Kumbhakar S, Tsionas EG. 2009. Nonparametric estimation of production risk and risk preference functions. *Adv. Econom.* 25:223–62

Kumbhakar S, Tsionas EG. 2010. Estimation of production risk and risk preference function: a nonparametric approach. *Ann. Oper. Res.* 176:369–78
LeCam L. 1972. Limits of experiments. *Proc. Symp. Math. Stat. Probab., 6th*, Berkeley, 1:245–61. Berkeley: Univ. Calif. Press
Luenberger DG. 1995. *Microeconomic Theory*. Boston: McGraw Hill
Mahler K. 1939. Ein Übertragungsprinzip für konvexe Körpers. *Čas. Pěstování Mat. Fys.* 68:93–102
Murty S, Russell RR, Levkoff S. 2012. On modeling pollution-generating technologies. *J. Environ. Econ. Manag.* 64:117–35
Rødseth LV. 2011. *Treatment of undesirable outputs in production analysis: desirable modeling strategies and application*. PhD thesis. UMB Sch. Bus., Aas, Nor.
Russell RR, Schworm W. 2011. Properties of inefficiency indexes on <input,output> space. *J. Prod. Anal.* 36:143–56
Shephard RW. 1953. *Cost and Production Functions*. Princeton: Princeton Univ. Press
Shephard RW. 1970. *The Theory of Cost and Production Functions*. Princeton: Princeton Univ. Press
Shephard RW, Färe R. 1974. *The Law of Diminishing Returns*. Berlin/Heidelberg: Springer
Simar L, Wilson P. 2008. Statistical inference in nonparametric frontier models: recent developments and perspectives. In *The Measurement of Productive Efficiency and Efficiency Change*, ed. H Fried, CAK Lovell, S Schmidt, pp. 421–521. Oxford, UK: Oxford Univ. Press
Solow RM. 1957. Technical change and the aggregate production function. *Rev. Econ. Stat.* 39:312–20
Tauer L, Lordkipanidze N. 2000. Farmer efficiency and technology use with age. *Agric. Res. Econ. Rev.* 29:24–31
Törnqvist L. 1936. The Bank of Finland's consumption price index. *Bank Finland Mon. Bull.* 10:1–8
Tsybakov A. 2009. *Introduction to Nonparametric Estimation*. New York: Springer
von Neumann J. 1945. A model of general economic equilibrium. *Rev. Econ. Stud.* 13:1–9
Wang C, Färe R, Seavert CF. 2006. Revenue capacity efficiency of pear trees and its decomposition. *J. Am. Soc. Hortic. Sci.* 131:30–40
Wang L, Yang L. 2007. Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Ann. Stat.* 35:2474–503
Weersink A, Turvey CG, Godah A. 1990. Decomposition measures of technical efficiency for Ontario dairy farms. *Can. J. Agric. Econ.* 38:439–56
Wilson P. 2008. FEAR 1.0: a software package for frontier efficiency analysis with *R*. *Socioecon. Plan. Sci.* 42:247–54

Annual Review of
Resource Economics

Volume 5, 2013

Contents

Errata

An online log of corrections to *Annual Review of Resource Economics* articles may be found at http://resource.annualreviews.org