Journal of Productivity Analysis, 20, 361–390, 2003

Testing Independence in Models of Productive Efficiency

PAUL W. WILSON wilson@eco.utexas.edu
Department of Economics, The University of Texas, Austin, Texas 78712 USA

Abstract

Bootstrap methods for inference in nonparametric models of productive efficiency can be simplified, reducing computational burden, when the independence condition implicitly assumed in Simar and Wilson (1998) holds. This paper surveys nonparametric tests of independence that might be useful in this context.

JEL Classification: C1, C4, C6

Keywords: dependence, data envelopment analysis, productive efficiency

1. Introduction

Nonparametric frontier models have been widely applied and are used to estimate various types of productive efficiency among firms within an industry. One typically begins with the notion of a production set defined by

$$\mathscr{P} = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \text{ can produce } \mathbf{y}\}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}_+^p$ denotes a vector of p inputs and $\mathbf{y} \in \mathbb{R}_+^q$ denotes a vector of q outputs. The boundary of $\mathscr{P}$,

$$\mathscr{P}^{\partial} = \{(\mathbf{x}, \mathbf{y}) \mid (\mathbf{x}, \mathbf{y}) \in \mathscr{P}, (\theta\mathbf{x}, \theta^{-1}\mathbf{y}) \notin \mathscr{P} \ \forall\ 0 < \theta < 1\}, \tag{2}$$

is frequently referred to as the technology or the production frontier, and is given by the intersection of $\mathscr{P}$ and the closure of its complement.

Given an arbitrary point $(\mathbf{x}_0, \mathbf{y}_0) \in \mathbb{R}_+^{p+q}$, researchers frequently are interested in some measure of distance from $(\mathbf{x}_0, \mathbf{y}_0)$ to $\mathscr{P}^{\partial}$ along some path. For example, if the path from $(\mathbf{x}_0, \mathbf{y}_0)$ to $\mathscr{P}^{\partial}$ is radial in the plane that (i) contains $(\mathbf{x}_0, \mathbf{y}_0)$ and (ii) is parallel to the axes along which $\mathbf{x}$ is measured, a measure of input technical efficiency is obtained. The metric used to measure distance along such a path is usually a normalized Euclidean distance measure, such as the Farrell (1957) input efficiency measure or the Shephard (1970) input distance function. For $(\mathbf{x}_0, \mathbf{y}_0) \in \mathscr{P}$, these measures provide an indication of whether, and how much, the input vector $\mathbf{x}_0$ could be scaled back without reducing the quantities of outputs $\mathbf{y}_0$; the amount by which it is technically feasible to reduce inputs without reducing output quantities is constrained by the production set $\mathscr{P}$.[1]

Given that $\mathscr{P}$ is not observed and therefore must be estimated, any efficiency measure based on distance from a point in $\mathbb{R}_+^{p+q}$ to some point in $\mathscr{P}^{\partial}$ must necessarily also be estimated. As in every empirical setting, point estimates of efficiency are largely useless—one must use statistical inference to discover what has been learned about the underlying true efficiency that is estimated. Simar and Wilson (2000b) discussed the current "state of the art" of inference in nonparametric frontier models; to date, bootstrap methods must be used in all but a few special cases.

Simar and Wilson (1998) provided one such method, which incorporates an implicit assumption of independence. In the case of input technical efficiency, the Simar and Wilson (1998) method assumes independence between technical inefficiency and output levels as well as the mix of inputs that are produced, which determines the direction of the radial path along which input inefficiency is estimated.[2] In the case of output technical efficiency, the Simar and Wilson (1998) method assumes independence between technical inefficiency and input levels as well as the the mix of outputs (i.e., the direction of the radial path along which output inefficiency is estimated.[3]

From an economic viewpoint, the independence assumption in the Simar and Wilson (1998) method represents a possibly severe restriction that may not be plausible. Large firms may have access to better managers, perhaps because they can offer higher salaries and other compensation, are better positioned to compete for managers in the labor market, offer more prestige and visibility to talented managers who wish to advance their careers, etc. Since large firms typically produce more output than small firms, one would therefore expect dependence between input technical efficiency and output levels, rather than independence. Analogously, to the extent that large firms use more inputs than small firms, one would also expect dependence between output technical efficiency and input levels.

Unsophisticated managers may not only create technical inefficiency, but perhaps cost inefficiency as well by ignoring input prices and choosing sub-optimal input combinations more frequently than sophisticated managers who create less technical inefficiency. Such behavior would induce dependence between input technical efficiency and the mix of inputs. Consequently, the assumption of independence between input technical inefficiency and the mix of inputs may also not be warranted, especially if all firms face the same vector of input prices. A similar argument can be made for the case of output technical inefficiency; less savvy managers may choose sub-optimal mixes of outputs by ignoring output prices more often than more savvy managers, inducing dependence between output technical efficiency and output mix.

The independence assumptions maintained in Simar and Wilson (1998) are avoided in the generalized approach of Simar and Wilson (2000a), but at the cost of an increase in complexity and computational burden.[4] For the applied researcher wanting to make inferences about efficiencies of firms, using the Simar and Wilson (1998) method without first testing the independence assumption incurs the risk of invalid inference if the assumption does not hold. On the other hand, using the Simar and Wilson (2000a) method when independence holds involves increased effort, and perhaps statistical inefficiency (i.e., increased variance in the estimators).[5]

Additional reasons to consider independence arise in two-stage models where efficiency is estimated nonparametrically in the first stage, and then estimated efficiencies are regressed on a vector $\mathbf{z} \in \mathbb{R}^r$ of environmental variables in a second stage using standard parametric (usually, maximum likelihood) techniques. Simar and Wilson (2002a) showed that conventional methods for inference in the second stage are invalid, due in part to slow convergence rates in the first stage. The solution is to employ another bootstrap procedure. Of course, before doing so, it would be useful to test independence between efficiency and the environmental variables before incurring the cost of bootstrapping in a complicated setting; if independence cannot be rejected, there may be no need for the second-stage estimation. Alternatively, in cases where independence between efficiency and the environmental variables is rejected and second-stage estimates are obtained, one might also want to test independence between the second-stage regressors and residuals for the usual reasons. Rejection of independence here could point to a misspecification of the second-stage regression model, or perhaps a misspecification in the first-stage as well.

Statistical independence is a fundamental concept, but is rarely considered in efficiency studies. The idea of testing for independence in regression contexts has received considerable attention in the statistics and econometrics literature, however. Various tests exist for testing independence between residuals and regressors in a regression context; many of these, however, exploit the structure of the regression model and are not suitable for present purposes. There are a number of nonparametric tests for independence, dating from Hoeffding (1948), including those based on ranks (e.g., Hoeffding, 1948; Kallenberg and Ledwina, 1999), those based on empirical distribution functions (e.g., Blum et al., 1961; Skaug and Tjøstheim, 1993), those based on correlation integrals (e.g., Brock et al., 1991; Johnson and McClelland, 1998), and those based on kernel estimates (e.g., Robinson, 1991; Ahmad and Li, 1997; Zheng, 1997; Dette and Neumeyer, 2000). This paper provides a brief survey of independence tests that might be useful in the context of efficiency estimation, and provides some empirical examples to illustrate their use. Given the large number of available tests for independence, my treatment is necessarily limited. Moreover, while the bootstrap methods discussed in Simar and Wilson (1998, 2000a) could be adapted to the problem of testing for independence in efficiency models, I focus on relatively simple tests that might be used before bootstrapping efficiency estimates.

The next section presents a simple model of production and a particular efficiency estimator to establish notation. In addition, a brief discussion of statistical independence in this setting is given. The third section describes existing tests that rely on asymptotic results, while the fourth section examines a graphical procedure. Section five provides some empirical illustrations as well as Monte Carlo evidence, while conclusions are drawn in the final section.

2. The Statistical Model

A statistical paradigm for efficiency estimation has been discussed in Simar and Wilson (2000b); the discussion here is intended only to establish notation, in order to demonstrate the differences between the bootstrap approaches proposed by Simar and Wilson (1998, 2000a).

As noted in the introduction, there are various measures of efficiency that one might estimate. For purposes of illustration, I consider the Shephard (1970) input distance function defined by

$$\delta(\mathbf{x}_0, \mathbf{y}_0 \mid \mathscr{P}) \equiv \sup\{\theta > 0 \mid (\theta^{-1}\mathbf{x}_0, \mathbf{y}_0) \in \mathscr{P}\}. \tag{3}$$

Clearly, $\delta(\mathbf{x}_0, \mathbf{y}_0 \mid \mathscr{P}) \geq 1$ for all $(\mathbf{x}_0, \mathbf{y}_0) \in \mathscr{P}$. Since $\mathscr{P}$ is unobserved, $\delta(\mathbf{x}_0, \mathbf{y}_0 \mid \mathscr{P})$ is also unobserved, and consequently must be estimated. Before anything can be estimated, however, some assumptions are needed to define a data-generating process (DGP).

The production set $\mathscr{P}$ defined in (1) is sometimes described in terms of its sections

$$\mathscr{Y}(\mathbf{x}) \equiv \{\mathbf{y} \mid (\mathbf{x}, \mathbf{y}) \in \mathscr{P}\}, \tag{4}$$

and

$$\mathscr{X}(\mathbf{y}) \equiv \{\mathbf{x} \mid (\mathbf{x}, \mathbf{y}) \in \mathscr{P}\}, \tag{5}$$

which form the output feasibility and input requirement sets, respectively. Knowledge of either $\mathscr{Y}(\mathbf{x})$ for all $\mathbf{x}$ or $\mathscr{X}(\mathbf{y})$ for all $\mathbf{y}$ is equivalent to knowledge of $\mathscr{P}$; $\mathscr{P}$ implies (and is implied by) both $\mathscr{Y}(\mathbf{x})$ and $\mathscr{X}(\mathbf{y})$. Thus, both $\mathscr{Y}(\mathbf{x})$ and $\mathscr{X}(\mathbf{y})$ inherit the properties of $\mathscr{P}$.

Various assumptions regarding $\mathscr{P}$ are possible; here, I adopt those of Shephard (1970). Other assumptions are possible, but might require small modifications in the notation that follows.

ASSUMPTION A1 *$\mathscr{P}$ is closed and convex; $\mathscr{Y}(\mathbf{x})$ is closed, convex, and bounded for all $\mathbf{x} \in \mathbb{R}^p_+$; and $\mathscr{X}(\mathbf{y})$ is closed and convex for all $\mathbf{y} \in \mathbb{R}^q_+$.*

ASSUMPTION A2 $(\mathbf{x}, \mathbf{y}) \notin \mathscr{P}$ *if* $\mathbf{x} = 0, \mathbf{y} \geq 0, \mathbf{y} \neq 0$, *i.e., all production requires use of some inputs.*[6]

ASSUMPTION A3 *For $\tilde{\mathbf{x}} \geq \mathbf{x}$, $\tilde{\mathbf{y}} \leq \mathbf{y}$, if $(\mathbf{x}, \mathbf{y}) \in \mathscr{P}$ then $(\tilde{\mathbf{x}}, \mathbf{y}) \in \mathscr{P}$ and $(\mathbf{x}, \tilde{\mathbf{y}}) \in \mathscr{P}$, i.e., both inputs and outputs are strongly disposable.*

Assumption A2 merely says that there are no free lunches. Assumption A3 is sometimes called free disposability and is equivalent to an assumption of monotonicity of the technology $\mathscr{P}^{\partial}$. Assumptions A1–A3 are standard economic assumptions in the microeconomic theory of the firm. Some additional assumptions beyond those of Shephard in A1–A3 are needed, however, to define the DGP.

The applied researcher is confronted with a set $\mathscr{S}_n = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$ of observed inputs and outputs for n firms.

ASSUMPTION A4 *The sample observations in $\mathscr{S}$ are realizations of identically, independently distributed (i.i.d.) random variables with probability density function $f(\mathbf{x}, \mathbf{y})$ with support over $\mathscr{P}$.*

The convex hull of the free disposal hull of the sample observations in $\mathscr{S}$, denoted $\widehat{\mathscr{P}}$, has frequently been used to estimate the production set $\mathscr{P}$. In order for this estimator of $\mathscr{P}$—and the corresponding estimator of $\delta(\mathbf{x}_0, \mathbf{y}_0|\mathscr{P})$—to be consistent, the probability of observing firms on $\mathscr{P}^\partial$ must approach unity as the sample size increases:

ASSUMPTION A5 *For all $(\mathbf{x}, \mathbf{y}) \in \mathscr{P}^\partial$, $f(\mathbf{x}, \mathbf{y})$ is strictly positive, and $f(\mathbf{x}, \mathbf{y})$ is continuous in any direction toward the interior of $\mathscr{P}$.*

In addition, an assumption about the smoothness of the frontier is needed:

ASSUMPTION A6 *For all $(\mathbf{x}, \mathbf{y})$ in the interior of $\mathscr{P}$, $\delta(\mathbf{x}, \mathbf{y} \mid \mathscr{P})$ is differentiable in both its arguments.*[7]

Assumptions A4–A6 are based on those of Kneip et al. (1998); together, assumptions A1–A6 define the DGP $\mathscr{F}$ which yields the data in $\mathscr{S}_n$.

Korostelev et al. (1995) proved that $\widehat{\mathscr{P}}$ is a consistent estimator of $\mathscr{P}$ under conditions met by assumptions A1–A6 above. An estimator of the distance function $\delta(\mathbf{x}_0, \mathbf{y}_0 \mid \mathscr{P})$ can be constructed by replacing $\mathscr{P}$ on the right-hand side of (3) with $\widehat{\mathscr{P}}$. For practical purposes, the estimator can be written as the linear program

$$\delta(\mathbf{x}_0, \mathbf{y}_0 \mid \widehat{\mathscr{P}}) = \left[\min\left\{\theta > 0 \mid \mathbf{Yq} \geq \mathbf{y}_0, \mathbf{Xz} \leq \theta \mathbf{x}_0, \mathbf{i}'\mathbf{q} = 1, \mathbf{q} \in \mathbb{R}_+^n\right\}\right]^{-1}, \tag{6}$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n]$, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, $\mathbf{i}$ denotes an $(n \times 1)$ vector of ones, and $\mathbf{q}$ is an $(n \times 1)$ vector of intensity variables whose values are determined by solution of the linear programs in each case.

The estimator $\delta(\mathbf{x}_0, \mathbf{y}_0 \mid \widehat{\mathscr{P}})$ measures normalized distance from a point $(\mathbf{x}_0, \mathbf{y}_0)$ to the boundary of $\widehat{\mathscr{P}}$, and hence gives an estimate of the distances from $(\mathbf{x}_0, \mathbf{y}_0)$ to $\mathscr{P}^\partial$. Kneip et al. (1998) proved consistency for the input distance function estimator in (6) and established its rate of convergence given assumptions A1–A6; in particular,

$$\delta(\mathbf{x}_0, \mathbf{y}_0 \mid \widehat{\mathscr{P}}) = \delta(\mathbf{x}_0, \mathbf{y}_0 \mid \mathscr{P}) + O_p(n^{-2/(p+q+1)}). \tag{7}$$

The rate of convergence is low, as is typical in nonparametric estimation, and the rate slows as $p + q$ is increased—this is the well-known curse of dimensionality that plagues most nonparametric estimators.[8] Moreover, since $\widehat{\mathscr{P}} \subseteq \mathscr{P}$, $\delta(\mathbf{x}_0, \mathbf{y}_0 \mid \widehat{\mathscr{P}}) \leq \delta(\mathbf{x}_0, \mathbf{y}_0 \mid \mathscr{P})$, and therefore $\delta(\mathbf{x}_0, \mathbf{y}_0|\widehat{\mathscr{P}})$ is biased downward.

Few results exist on the sampling distribution of the distance function estimator in (6). Gijbels et al. (1999) derived the asymptotic distribution of the corresponding output distance function estimator for the special case of one input and one output ($p = q = 1$), along with an analytic expression for its large sample bias and variance; these results easily extend to the input distance function estimator in (6). Unfortunately, in the more general multivariate setting where $p + q > 2$, the radial nature of the distance functions and the complexity of the estimated frontier complicates the derivations. So far, the bootstrap methods described by Simar and Wilson (1998, 2000a) appear to offer the only way to approximate the asymptotic distribution of distance function estimators in multivariate settings.

Note that a point $(\mathbf{x}_0, \mathbf{y}_0) \in \mathbb{R}_+^{p+q}$ represented by Cartesian coordinates can also be represented by cylindrical coordinates $(\tau_0, \boldsymbol{\eta}_0, \mathbf{y}_0)$, where $(\tau_0, \boldsymbol{\eta}_0)$ are the polar coordinates of $\mathbf{x}_0 \in \mathbb{R}_+^p$. The modulus is $\tau_0 = \tau(\mathbf{x}_0) = \sqrt{\mathbf{x}_0'\mathbf{x}_0} \in \mathbb{R}_+^1$ and the jth element of the corresponding angle $\boldsymbol{\eta}_0 = \boldsymbol{\eta}(\mathbf{x}_0) \in [0, \pi/2]^{p-1}$ of $\mathbf{x}_0$ is given by $\arctan(x_{0,j+1}/x_{01})$ for $x_{01} \neq 0$ (where x_{0j} represents the jth element of $\mathbf{x}_0$); if $x_{01} = 0$, then all elements of $\eta(\mathbf{x}_0)$ equal zero.

Let $\delta_0 = \delta(\mathbf{x}_0, \mathbf{y}_0 \mid \mathscr{P})$, and let $\widehat{\delta}_0 = \delta(\mathbf{x}_0, \mathbf{y}_0 \mid \widehat{\mathscr{P}})$. Now consider a point $(\mathbf{x}_0, \mathbf{y}_0) \in \mathscr{P}$, and its projection $(\mathbf{x}_0/\delta_0, \mathbf{y}_0)$ onto $\mathscr{P}^\partial$ in the direction orthogonal to $\mathbf{y}$. The moduli of these points are related to the input distance function via

$$0 \leq \delta(\mathbf{x}_0, \mathbf{y}_0 \mid \mathscr{P}) = \frac{\tau(\mathbf{x}_0)}{\tau(\mathbf{x}_0/\delta_0)} \leq 1. \tag{8}$$

If $\mathscr{P}$ were known, the point $(\tau_0, \eta_0, \mathbf{y}_0)$ could be represented by $(\delta_0, \eta_0, \mathbf{y}_0)$. Alas, $\mathscr{P}$ is not known; but its estimate, $\widehat{\mathscr{P}}$, is known. The projection of $(\mathbf{x}_0, \mathbf{y}_0) \in \mathscr{P}$ onto $\widehat{\mathscr{P}}$ in the direction orthogonal to $\mathbf{y}$ is the point $(\mathbf{x}_0/\widehat{\delta}_0, \mathbf{y}_0)$, and analogous to (8),

$$0 \leq \delta(\mathbf{x}_0, \mathbf{y}_0 | \widehat{\mathscr{P}}) = \frac{\tau(\mathbf{x}_0)}{\tau(\mathbf{x}_0/\widehat{\delta}_0)} \leq 1. \tag{9}$$

Conditionally on $\widehat{\mathscr{P}}$, all the terms in (9) are known, and hence the point $(\mathbf{x}_0, \mathbf{y}_0) = (\tau_0, \eta_0, \mathbf{y}_0)$ is fully represented by $(\widehat{\delta}_0, \boldsymbol{\eta}_0, \mathbf{y}_0)$ given the estimate $\widehat{\delta}_0$ that results when conditioning on $\widehat{\mathscr{P}}$.

Writing $f(\mathbf{x}, \mathbf{y})$ in terms of cylindrical coordinates, the density can be decomposed by writing

$$f(\mathbf{x}, \mathbf{y}) = f(\tau, \boldsymbol{\eta}, \mathbf{y}) = f(\tau \mid \boldsymbol{\eta}, \mathbf{y}) f(\boldsymbol{\eta} \mid \mathbf{y}) f(\mathbf{y}), \tag{10}$$

where $f(\mathbf{y})$is defined on $\mathbb{R}_+^1$, $f(\boldsymbol{\eta} \mid \mathbf{y})$ is defined on $[0, \pi/2]^{q-1}$, $f(\tau \mid \boldsymbol{\eta}, \mathbf{y})$ is defined on $\mathbb{R}_+^1$, and all the conditional densities exist. The density $f(\tau | \boldsymbol{\eta}, \mathbf{y})$ on $[0, \tau(\mathbf{x}/\delta)]$ and (8) imply a density $f(\delta \mid \boldsymbol{\eta}, \mathbf{y})$ on the interval $[0, 1]$. Similarly, the density $f(\tau \mid \boldsymbol{\eta}, \mathbf{y})$ and (9) imply a density $f(\widehat{\delta} \mid \boldsymbol{\eta}, \mathbf{y})$ on the interval $[0, 1]$. Hence, (10) can be replaced by

$$f(\mathbf{x}, \mathbf{y}) = f(\delta, \boldsymbol{\eta}, \mathbf{y}) = f(\delta \mid \boldsymbol{\eta}, \mathbf{y}) f(\boldsymbol{\eta} \mid \mathbf{y}) f(\mathbf{y}). \tag{11}$$

The homogeneous bootstrap method described by Simar and Wilson (1998) implicitly assumes

$$f(\delta \mid \boldsymbol{\eta}, \mathbf{y}) = f(\delta), \tag{12}$$

i.e., that δ is independent of $(\boldsymbol{\eta}, \mathbf{y})$. On the other hand, the heterogeneous bootstrap method described by Simar and Wilson (2000a) allows for the possibility of dependence between δ and $(\boldsymbol{\eta}, \mathbf{y})$, i.e.,

$$f(\delta \mid \boldsymbol{\eta}, \mathbf{y}) \neq f(\delta). \tag{13}$$

As discussed in the introduction, one might expect that large firms have access to better management. If so, to the extent that large firms produce more output than smaller firms, efficiency and output will be dependent, and (12) will not hold. To the extent that technically inefficient managers also choose sub-optimal input combinations (as determined by input prices), and to the extent that firms face similar input prices, input technical inefficiency and the mix of inputs as reflected by η may also be dependent.

Suppose that in either case, B bootstrap replications are employed to estimate confidence intervals for the distance function $\delta_0 = \delta(\mathbf{x}_0, \mathbf{y}_0 \mid \mathscr{P})$ corresponding to a single point $(\mathbf{x}_0, \mathbf{y}_0) \in \mathbb{R}_+^{p+q}$. With a sample size of n, the homogeneous bootstrap requires solving $(n + B)$ linear programs, while the heterogeneous bootstrap requires solving a minimum of $(n + B + Bn)$ linear programs.[9] Since B will typically be at least 1,000, the difference in computational burden is substantial, particularly when n is large. If one wishes to evaluate the accuracy of the estimated confidence intervals by iterating the bootstrap as discussed in Simar and Wilson (2001), the difference in computational burden between the two approaches becomes even greater. For estimation of a single or only a few confidence intervals, the difference in computing times between the two methods is not likely to be important on modern machinery; however, the difference becomes substantial when confidence intervals corresponding to each of n observations are to be estimated.

Unfortunately, if one blindly employs the homogeneous bootstrap and $\delta$is not independent of $(\boldsymbol{\eta}, \mathbf{y})$, the resulting confidence intervals may have poor coverage, and are unlikely to improve as sample size increases. Therefore, a reliable test of the independence assumption is needed.

3. Tests Based on Analytic, Asymptotic Results

To simplify notation, I restate the problem at the end of the last section as one of testing independence between a scalar random variable $\delta$and a vector of random variables $\mathbf{w} \in \mathbb{R}^m$that have been rescaled to have unit variance; $\mathbf{w}$ might equal $(\boldsymbol{\eta}, \mathbf{y})$ in the notation of the last section (in which case $m = p + q - 1$), or $\mathbf{w}$ might comprise a set of environmental variables in a second-stage regression as described in Simar and Wilson (2002a).[10] Let $\mathbf{w}_i$ denote the ith observed value of $\mathbf{w}$, and let $\mathbf{w}_{ik}$ denote the kth element of the ith observed value $\mathbf{w}_i, k = 1, \ldots, m$; let $\mathbf{w}_{\cdot k}$ denote the kth element of the random vector $\mathbf{w}$. Assume all elements of $\mathbf{w}$ are continuous. Finally, consistent with the notation of the previous section, let δ_i denote the (un)observed input distance function corresponding to the ith observed input and output vectors; i.e., $\delta_i = \delta(\mathbf{x}_i, \mathbf{y}_i \mid \mathscr{P})$. Similarly, let $\widehat{\delta}_i$ denote the (observed) estimate of δ_i computed from (6); i.e., $\widehat{\delta}_i = \delta(\mathbf{x}_i, \mathbf{y}_i \mid \widehat{\mathscr{P}})$.

With only a small abuse of notation, let $f(\delta)$and $f(\mathbf{w})$ denote the marginal probability density functions (PDFs) of δ and $\mathbf{w}$, respectively, and let $f(\delta, \mathbf{w})$ denote their joint PDF; the PDFs are distinguished by their arguments. Analogously, let $F(\delta)$ and $F(\mathbf{w})$ denote the marginal distribution functions (DFs) of δ and $\mathbf{w}$, and let $F(\delta, \mathbf{w})$ denote their joint DF. In cases where there is ambiguity, I will use subscripts to distinguish the PDFs and DFs (e.g., $f_\delta(\cdot)$, and $F_\delta(\cdot)$ denote the PDF and DF for δ).

The problem of testing the null hypothesis of independence between δ and $\mathbf{w}$, denoted $\delta \perp\!\!\!\perp \mathbf{w}$, is not unlike the problem of testing independence between errors and regressors in the parametric, classical linear regression model. Here, δ is unobserved, as are the errors in the standard regression setting. However, (6) provides a consistent estimator of δ, and hence the estimates $\widehat{\delta}_i$ may be substituted for the unobserved values δ_i in tests of $\delta \perp\!\!\!\perp \mathbf{w}$. Given the slow convergence rate shown in (7), using the $\widehat{\delta}_i$ instead of the δ_i can be expected to reduce the power of any tests of $\delta \perp\!\!\!\perp \mathbf{w}$, but there is no alternative since the δ_i cannot be observed. Throughout this section, I describe tests of $\delta \perp\!\!\!\perp \mathbf{w}$ in terms of δ_i and $\mathbf{w}_i$, but of course the δ_i must be replaced by the estimated values $\widehat{\delta}_i$ to implement the tests.[11]

If δ and $\mathbf{w}$ were multivariate normal, then $\delta \perp\!\!\!\perp \mathbf{w} \Longleftrightarrow \mathrm{COV}(\delta, w_{\cdot k}) = 0 \; \forall\, k = 1, \ldots, m$; i.e., δ and $\mathbf{w}$ are independent if and only if δ and each of the w_k have zero covariance. In this case, testing independence would be a matter of simply testing for zero linear correlation. However, δ is certainly not normal, and the elements of $\mathbf{w}$ may not be as well. In general,

$$\delta \perp\!\!\!\perp \mathbf{w} \Longleftrightarrow E[g_1(\delta) g_2(\mathbf{w})] = E[g_1(\delta)] E[g_2(\mathbf{w})], \tag{14}$$

for every Borel function $g_1(\cdot): \mathbb{R}^1 \to \mathbb{R}^1$and $g_2(\cdot): \mathbb{R}^m \to \mathbb{R}^1$. Any dependence between δ and $\mathbf{w}$ may be of a complicated form; linear correlation is only one aspect of dependence in the general case, and even if δ and w_k have zero linear correlation, they may still be dependent if δ or $\mathbf{w}$ is not normally distributed.

It is always possible to write $f(\delta, \mathbf{w}) = f(\delta \mid \mathbf{w}) f(\mathbf{w})$, and (14) implies that if $\delta \perp\!\!\!\perp \mathbf{w}$, then $\mathbf{w}$ contains no information about δ; in other words, the conditioning event in $f(\delta \mid \mathbf{w})$ is not informative. Consequently,

$$\delta \perp\!\!\!\perp \mathbf{w} \Longleftrightarrow f(\delta \mid \mathbf{w}) = f(\delta) \tag{15}$$

or, equivalently,

$$\delta \perp\!\!\!\perp \mathbf{w} \Longleftrightarrow f(\delta, \mathbf{w}) = f(\delta) f(\mathbf{w}). \tag{16}$$

This forms the basis for a number of tests of independence.

3.1 Tests Based on Smoothing

Ahmad and Li (1997) test the null hypothesis $H_0: f(\delta, \mathbf{w}) = f(\delta) f(\mathbf{w})$ versus the alternative hypothesis $H_1: f(\delta, \mathbf{w}) \neq f(\delta) f(\mathbf{w})$ by considering the integrated square difference between $f(\delta, \mathbf{w})$ and $f(\delta)f(\mathbf{w})$:

$$
\begin{aligned}
I &= \int\int [f(\delta, \mathbf{w}) - f(\delta)f(\mathbf{w})]^2 \, d\delta \; d\mathbf{w} \\
&= \int\int f(\delta, \mathbf{w}) \; dF(\delta, \mathbf{w}) + \int f(\delta) \, dF(\delta) \int f(\mathbf{w}) \, dF(\mathbf{w}) \\
&\quad - 2 \int\int f(\delta)f(\mathbf{w}) dF(\delta, \mathbf{w}).
\end{aligned} \tag{17}
$$

This can be estimated by the statistics

$$
\begin{aligned}
\widehat{I}_n = \left(n^2 h_\delta h_{\mathbf{w}}^m\right)^{-1} \sum_i \sum_j K_{ij}^\delta K_{ij}^{\mathbf{w}} + \left(n^4 h_\delta h_{\mathbf{w}}^m\right)^{-1} \left(\sum_i \sum_j K_{ij}^\delta\right) \left(\sum_\ell \sum_r K_{\ell r}^{\mathbf{w}}\right) \\
- 2\left(n^3 h_\delta h_{\mathbf{w}}^m\right)^{-1} \sum_i \sum_j \sum_\ell K_{ij}^\delta K_{j\ell}^{\mathbf{w}},
\end{aligned} \tag{18}
$$

and

$$
\begin{aligned}
\widetilde{I}_n = \left(n^2 h_\delta h_{\mathbf{w}}^m\right)^{-1} \sum_i \sum_{j \neq i} K_{ij}^\delta K_{ij}^{\mathbf{w}} + \left(n^4 h_\delta h_{\mathbf{w}}^m\right)^{-1} \left(\sum_i \sum_{j \neq i} K_{ij}^\delta\right) \left(\sum_\ell \sum_{r \neq \ell} K_{\ell r}^{\mathbf{w}}\right) \\
- 2\left(n^3 h_\delta h_{\mathbf{w}}^m\right)^{-1} \sum_i \sum_{j \neq i} \sum_{\ell \neq j} K_{ij}^\delta \, K_{j\ell}^{\mathbf{w}},
\end{aligned} \tag{19}
$$

where $K_{ij}^\delta = K_1\left(h_\delta^{-1}(\delta_i - \delta_j)\right)$, $K_{ij}^{\mathbf{w}} = K_2\left(h_{\mathbf{w}}^{-m}(\mathbf{w}_i - \mathbf{w}_j)\right)$, h_δ and $h_{\mathbf{w}}$ are bandwidth parameters, $K_1(\cdot)$is a univariate Gaussian PDF with zero mean and variance equal to the sample variance of the $\widehat{\delta}_i$, and $K_2(\cdot)$ is an m-variate Gaussian PDF with zero mean and covariance matrix equal to the sample covariance matrix of the observed $\mathbf{w}_i$.[12] $K_1(\cdot)$ and $K_2(\cdot)$ are kernel functions.

Note that $\widetilde{I}_n$ equals $\widehat{I}_n$less the terms where $j = i$, etc. Ahmad and Li (1997) show that $\widehat{I}_n$ contains a degenerate U-statistic, which leads to a bias term that must be estimated when $\widehat{I}_n$is used. In particular,

$$
\widehat{T}_{1n} \equiv \widehat{\sigma}_0^{-1} n h_\delta^{1/2} h_{\mathbf{w}}^{m/2} \left(\widehat{I}_n - \widehat{b}(n)\right) \xrightarrow{d} N(0, 1), \tag{20}
$$

and

$$
\widehat{T}_{2n} \equiv \widehat{\sigma}_0^{-1} n h_\delta^{1/2} h_{\mathbf{w}}^{m/2} \widetilde{I}_n \xrightarrow{d} N(0, 1), \tag{21}
$$

where

$$
\begin{aligned}
\widehat{b}(n) = \left(n h_\delta h_{\mathbf{w}}^m\right)^{-1} K^\delta(0) K^{\mathbf{w}}(0) - \left(n h_\delta^{1/2} h_{\mathbf{w}}^{1/m}\right)^{-1} \\
\times \left\{K^\delta(0) \widehat{E}[f(\mathbf{w})] + K^{\mathbf{w}}(0) \widehat{E}[f(\delta)]\right\}^3,
\end{aligned} \tag{22}
$$

$\widehat{E}[f(\delta)] = n^{-1}\sum_i f(\delta_i)$ and similarly for $\widehat{E}[f(\mathbf{w})]$, and

$$\widehat{\sigma}_0^2 = n^{-2}\Big[\Big(\sum \widehat{f}(\delta_i)\Big)\Big(\sum \widehat{f}(\mathbf{w}_i)\Big)\Big]\int \left(K^{\delta}(\delta)\right)^2 \, d\delta \int (K^{\mathbf{w}}(\mathbf{w}))^2 d\mathbf{w}. \tag{23}$$

Knowledge of the asymptotic distributions in (20)–(21) permit tests of the null hypothesis of independence. Ahmad and Li's proof of these results does not depend on $f(\delta)$having support over $\mathbb{R}$. Kernel density estimators are, of course, biased near boundaries, but this bias disappears as $n \to \infty$since the bandwidth tends to zero as $n \to \infty$. One could modify the statistics $\widehat{I}_n$ and $\widetilde{I}_n$ to account for the bounds on δ, but I do not consider this since the distributional results in (20)–(21) are only asymptotic.[13]

Zheng (1997) takes a slightly different approach, and bases his test on the difference between the conditional and the marginal DFs of δ, i.e.,

$$\Delta(\delta|\mathbf{w}) = F(\delta|\mathbf{w}) - F(\delta). \tag{24}$$

Under the null hypothesis of independence, $\Delta(\delta \,|\, \mathbf{w}) = 0$, while under the alternative hypothesis of dependence, $\Delta(\delta \,|\, \mathbf{w}) \neq 0$. Zheng's test requires an estimator of the integral

$$W = \int E[\Delta(\delta \,|\, \mathbf{w}) f(\mathbf{w})]\omega(\delta) \, d\delta, \tag{25}$$

where $\omega(\delta)$ is a weighting function. Zheng suggests the estimator

$$\begin{aligned}\widehat{W}_n = n^{-1}(n-1)^{-1}h_{\mathbf{w}}^{-m}\sum_{i\neq j}\Bigg\{K_1\left(\frac{\mathbf{w}_i - \mathbf{w}_j}{h_{\mathbf{w}}}\right)\\ \times \int \Big[I(\delta_i \leq \delta) - \widehat{F}_n(\delta)\Big]\Big[I(\delta_j \leq \delta) - \widehat{F}_n(\delta)\Big]\omega(\delta)\, d\delta\Bigg\},\end{aligned} \tag{26}$$

where $I(\cdot)$ is the indicator function, $K_1(\cdot)$ is defined as before, and

$$\widehat{F}_n(\delta) = n^{-1}\sum_{\ell=1}^{n} I(\delta_\ell \leq \delta), \tag{27}$$

is the empirical distribution function of the δ_i. Zheng establishes that

$$\widehat{T}_{3n} \equiv n\sigma_2^{-1}h_{\mathbf{w}}^{m/2}\widehat{W}_n \xrightarrow{d} N(0,1), \tag{28}$$

where

$$\begin{aligned}\sigma_2^2 = 2\int [K_1(u)]^2 \, du \int\int E[f(\mathbf{w})]\\ \times [F_\delta(u \wedge v) - F_\delta(u)F_\delta(v)]^2\omega(u)\,\omega(v)\, du\, dv,\end{aligned} \tag{29}$$

$F_\delta(\cdot)$ denotes the DF of the δ_i, and $u \wedge v$ denotes the minimum of the scalars u and v.

The tests of Ahmad and Li (1997) and Zheng (1997) incorporate nonparametric kernel density estimators, and consequently require selection of bandwidth parameters. There is a substantial recent literature on this problem; see, for example, Silverman (1986), Staniswalis (1989), Hall et al. (1991), Sheather and Jones (1991), Hall et al. (1992), Scott (1992), Kim et al. (1994), and Jones et al. (1996). Most of the available methods fall into two groups: those that make (perhaps unwarranted) assumptions about the underlying true density, such as the well-known normal reference rule, and those that make no assumptions and are fully non-parametric. Those in the former group are easy to apply, but do not perform well when the assumptions on the underlying density do not hold. Those in the latter group involve substantial computational burdens, which are worse in the present case due to two features of the problem: (i) data are discretized, due to the sometimes large number of estimated efficiencies equal to unity; and (ii) the underlying density in the efficiency estimation context has bounded support.[14]

The problems of bandwidth selection in the efficiency estimation context have been discussed in by Simar and Wilson (2000a), where minimization of a weighted cross-validation function for selecting bandwidths based on features of the data in a particular sample is discussed. One of the points of testing the independence assumption in (12), however, is that if independence cannot be rejected, one might be willing to use the homogeneous bootstrap method of Simar and Wilson (1998), thereby avoiding the increased computational burden of using the heterogeneous bootstrap method of Simar and Wilson (2000a). The increased computational burden of the heterogeneous bootstrap method lies primarily in the cross-validation exercise required to find a bandwidth for kernel smoothing. The increased computational burden of the heterogeneous bootstrap method is not formidable—and CPU cycles become cheaper each month—the pointis simply that the computational advantage of the homogeneous bootstrap method will be lost if one uses a test of independence along the lines of Ahmad and Li (1997) or Zheng (1997), and other approaches to testing independence are available that do not require selection of bandwidths for kernel density estimation.

As an obvious alternative to the tests of Ahmad and Li (1997) and Zheng (1997), one can easily construct a Kolmogorov-type test and use ordinary bootstrap methods to obtain a critical value. Let

$$\widehat{F}_n(\mathbf{w}) = n^{-1} \sum_{\ell=1}^{n} I(\mathbf{w}_\ell \leq \mathbf{w}), \tag{30}$$

denote the empirical distribution function for $\mathbf{w}$, analogous to the definition of $\widehat{F}_n(\delta)$, the empirical distribution function of δ, in (27). Similarly, let

$$\widehat{F}_n(\delta, \mathbf{w}) = n^{-1} \sum_{\ell=1}^{n} I(\delta_\ell \leq \delta) \times I(\mathbf{w}_\ell \leq \mathbf{w}), \tag{31}$$

denote the empirical joint distribution of $(\delta, \mathbf{w})$. Then the statistic

$$\widehat{T}_{4n} \equiv \sum_{i=1}^{n} \left[\widehat{F}_n(\widehat{\delta}_i, \mathbf{w}_i) - \widehat{F}_n(\widehat{\delta}_i)\widehat{F}_n(\mathbf{w}_i) \right]^2, \tag{32}$$

can be used to test the null hypothesis $H_0 : \delta \perp\!\!\!\perp \mathbf{w}$. The statistic $\widehat{T}_{4n}$ estimates the integrated square difference between $F(\delta, \mathbf{w})$ and $F(\delta)F(\mathbf{w})$, and thus is similar in spirit to the statistic proposed by Härdle and Mammen (1993) for testing parametric regression surfaces against nonparametric alternatives.[15]

Since interest lies in whether $F(\delta, \mathbf{w}) = F(\delta)F(\mathbf{w})$, one can use bootstrap methods to resample δ and $\mathbf{w}$, and then re-apply (27) and (30)–(31) to obtain bootstrap values T^*_{4n} of the test statistic in (32). These, in turn, can be used to obtain a critical value. If $\widehat{T}_{4n}$ exceeds the chosen critical value, then one would reject the null hypothesis of independence. Alternatively, the bootstrap values can be used to estimate p-values as described in Simar and Wilson (2002b).

Resampling of the $\mathbf{w}$ can be accomplished by drawing from the empirical distribution defined in (30). This approach does not work well for the δ, however, due to the fact that some (perhaps many) of the realized estimates $\widehat{\delta}_i$ will equal unity (see Simar and Wilson, 2000b, for discussion). Alternatively, one can choose an appropriate bandwidth for a univariate kernel density estimate of the density of the $\widehat{\delta}_i$, and then draw from this density estimate; see Simar and Wilson (1998) or Silverman (1986) for details.[16]

The entire bootstrap procedure can be performed in several steps:

1. For each observation in $\mathscr{S}_n$, estimate the input distance function using (6), compute the input angles $\eta_{ij}, j = 1, \ldots, p-1$ as described in Section 2, and define $\mathbf{w}_i = [\eta_{i1}, \ldots, \eta_{i,p-1}\; y_{i1}, \ldots, y_{iq}]$ for each $i = 1, \ldots, n$.
2. Compute $\widehat{T}_{4n}$ defined in (32).
3. Choose an appropriate bandwidth for a univariate kernel density estimate of the density of the estimated distance function values $\widehat{\delta}_i$.
4. Draw n times, independently, from the kernel density estimate in step 3 to form the set $\{\delta^*_i\}_{i=1}^n$.
5. Draw n times, uniformly and with replacement, from the set $\{\mathbf{w}_i\}_{i=1}^n$ to form the set $\{\mathbf{w}^*_i\}_{i=1}^n$.
6. Compute T^*_{4n} by letting δ^*_i and $\mathbf{w}^*_i$ replace $\widehat{\delta}_i$ and $\mathbf{w}_i$ in (32).
7. Repeat steps 4–6 B times to obtain a set of bootstrap values $\{T^*_{4n,b}\}_{b=1}^B$.
8. Compute the bootstrap critical value $\widehat{p} = \#\{\widehat{T}^*_{4n,b} \geq \widehat{T}_{4n}\}/B$.

Note that the test described here is conditional on the initial distance function estimates, since resampling is from the empirical distribution of the $\widehat{\delta}_i$.

A slight variation on this approach is obtained by defining

$$\widehat{T}_{5n} \equiv \max_i \mid \widehat{F}_n(\widehat{\delta}_i, \mathbf{w}_i) - \widehat{F}_n(\widehat{\delta}_i)\widehat{F}_n(\mathbf{w}_i) \mid. \tag{33}$$

This statistic is simply an estimate of the maximum L_1 difference between $F(\delta, \mathbf{w})$ and $F(\delta)F(\mathbf{w})$, and is analogous to the well-known Kolmogorov–Smirnov statistic. Critical values can be obtained by straightforward modification of the bootstrap algorithm discussed above. Monte Carlo evidence on the performances of $\widehat{T}_{4n}$ and $\widehat{T}_{5n}$ is presented in Section 5.

3.2 Tests Based on Correlation Integrals

Johnson and McClelland (1998) base their test on the idea that if $\delta \perp\!\!\!\perp \mathbf{w}$, then

$$\Pr\big[\, \| (\delta_i, \mathbf{w}_i) - (\delta_j, \mathbf{w}_j) \|_s < \varepsilon\big] = \Pr\big[\, | \delta_i - \delta_j | < \varepsilon\big] \times \Pr\big[\, \| \mathbf{w}_i - \mathbf{w}_j \|_s < \varepsilon\big], \tag{34}$$

for $i \neq j$ and all $\varepsilon > 0$, where $\| \cdot \|_s$ denotes the sup-norm.[17]

The probability on the left in (34) is given by the correlation integral

$$\begin{aligned} C((\delta, \mathbf{w}), \varepsilon) &\equiv \Pr\big[\, \| (\delta_i, \mathbf{w}_i) - (\delta_j, \mathbf{w}_j) \|_s < \varepsilon\big] \\ &= E\big\{\mathscr{I}\big[(\delta_i, \mathbf{w}_i) - (\delta_j, \mathbf{w}_j), \varepsilon\big]\big\}, \end{aligned} \tag{35}$$

where $\mathscr{I}(\cdot, \cdot)$ is an indicator function defined by

$$\mathscr{I}\big[(\delta_i, \mathbf{w}_i) - (\delta_j, \mathbf{w}_j), \varepsilon\big] = \begin{cases} 1, & \text{if } \| (\delta_i, \mathbf{w}_i) - (\delta_j, \mathbf{w}_j) \|_s < \varepsilon, \\ 0, & \text{otherwise.} \end{cases} \tag{36}$$

Thus, (34) is equivalent to $C((\delta, \mathbf{w}), \varepsilon) = C(\delta, \varepsilon) \times C(\mathbf{w}, \varepsilon)$.

Johnson and McClelland suggest U-statistics as estimators of the correlation integrals:

$$\widehat{C}_n((\delta, \mathbf{w}), \varepsilon) = \frac{2}{n(n-1)} \sum_{1 \le j < k \le n} \mathscr{I}(\delta_j - \delta_k, \varepsilon) \times \left[\prod_{\ell=1}^{m} \mathscr{I}(w_{j\ell} - w_{k\ell}, \varepsilon)\right], \tag{37}$$

$$\widehat{C}_n(\mathbf{w}, \varepsilon) = \frac{2}{n(n-1)} \sum_{1 \le j < k \le n} \left[\prod_{\ell=1}^{m} \mathscr{I}(w_{j\ell} - w_{k\ell}, \varepsilon)\right], \tag{38}$$

and with $\widehat{C}_n(\delta, \varepsilon)$ defined analogously to $\widehat{C}_n(\mathbf{w}, \varepsilon)$ in (38). Then, for fixed ε,

$$\widehat{T}_{6n} \equiv n^{1/2} \sigma_4^{-1} \left[\widehat{C}_n((\delta, \mathbf{w}), \varepsilon) - \widehat{C}_n(\delta, \varepsilon)\widehat{C}_n(\mathbf{w}, \varepsilon)\right] \xrightarrow{d} N(0, 1), \tag{39}$$

where

$$\sigma_4^2 = 4\Big[\psi(\delta,\varepsilon)\psi(\mathbf{w},\varepsilon) - \psi(\delta,\varepsilon)C(\mathbf{w},\varepsilon) - \psi(\mathbf{w},\varepsilon)C(\delta,\varepsilon) + C(\delta,\varepsilon)^2 C(\mathbf{w},\varepsilon)^2\Big], \tag{40}$$

$$\psi(\mathbf{w},\varepsilon) = E\big\{E\big[\mathscr{I}(\mathbf{w}_j - \mathbf{w}_k,\varepsilon) \mid \mathbf{w}_j = \mathbf{w}_i\big] \times E\big[\mathscr{I}(\mathbf{w}_j - \mathbf{w}_k,\varepsilon) \mid \mathbf{w}_k = \mathbf{w}_{j+i}\big]\big\}, \tag{41}$$

$C(\mathbf{w},\varepsilon) = E\big[\mathscr{I}(\mathbf{w}_j - \mathbf{w}_k,\varepsilon)\big]$, and similarly for $\psi(\delta,\varepsilon)$and $C(\delta,\varepsilon)$. To implement the test, $C(\mathbf{w},\varepsilon)$ and $C(\delta,\varepsilon)$ can be replaced by their estimators defined by (38), $\psi(\mathbf{w},\varepsilon)$ can be replaced by its estimator

$$\widehat{\psi}(\mathbf{w},\varepsilon) = \frac{2}{n(n-1)^2}\sum_{i=1}^{n}\left[\sum_{\ell=1}^{n-1}\mathscr{I}(\mathbf{w}_i - \mathbf{w}_\ell,\varepsilon)\sum_{j=\ell+1}^{n}\mathscr{I}(\mathbf{w}_i - \mathbf{w}_j,\varepsilon)\right], \tag{42}$$

and similarly for $\psi(\delta,\varepsilon)$ to obtain an estimator of σ_4^2.

The test based on $\widehat{T}_{6n}$ has been generalized for time-series contexts (Johnson and McClelland, 1998), and is similar to the test discussed by Brock et al. (1991).

While the asymptotic properties of the Johnson and McClelland test under the null hypothesis of independence are not affected by choice of ε, Johnson and McClelland (1997) note that the power of the test in finite-samples might be affected. If the data are standardized so that δ and each element of $\mathbf{w}$ have unit variance, then results from Brock et al. (1991) suggest that power of the similar Brock et al. (1991) test is maximized for values of ε in the range $(0.5, 1.4)$. In their Monte Carlo simulations, Johnson and McClelland (1997) set ε equal to one standard deviation of the data, after normalizing all variables to have the same standard deviation.

3.3. *Tests Based on Ranks*

As noted at the beginning of this section, linear correlation is only one aspect of dependence. For two random variables with symmetric densities, Pearson's correlation coefficient, Spearman's rho, and Kendall's tau statistics are frequently used as measures of dependence, and a number of tests are based on these (e.g., Hoefding, 1948; Kruskal, 1958; Lehman, 1966). Kallenberg and Ledwina (1999) generalize these ideas through the use of copulas to test independence between pairs of random variables. Their results extend to the problem here, where independence between the scalar δ and the vector $\mathbf{w}$ is to be tested.[18]

The random variables δ and $\mathbf{w}$ may be transformed by setting $\delta^* = F_\delta(\delta)$, $\mathbf{w}^* = F_{\mathbf{w}}(\mathbf{w})$. The joint distribution of $(\delta^*, \mathbf{w}^*)$ is the grade representation of the joint distribution $F(\delta, \mathbf{w})$, and is a particular copula function as defined by Sklar (1958). Introduction of a copula transforms the problem from one of integrals over a hyperplane into integrals over the unit square in $\mathbb{R}^2$.

It is well known that both density functions and conditional mean functions in regression problems can be estimated by (i) first taking a Fourier expansion of the density function or conditional mean function, then (ii) choosing a specification for the resulting set of basis functions (often a system of orthogonal polynomials), and finally (iii) estimating the unknown Fourier coefficients after truncating the expansion in (i). In regression, this approach has been called orthogonal series estimation, as well as semi-non-parametric estimation.

These ideas can be extended to the problem of testing independence. Under mild assumptions that are typically satisfied in the problems considered here,

$$\rho_{jk} = \mathrm{COV}\left[\lambda_j(\delta^*), \lambda_k(\mathbf{w}^*)\right] = 0 \ \forall \ j,k \geq 1, \tag{43}$$

implies $\delta \perp\!\!\!\perp \mathbf{w}$, where $\{\lambda_\ell(\cdot)\}$ is an appropriate set of orthonormal polynomials. The idea is to characterize any dependence between the underlyingrandom variables δ and $\mathbf{w}$ in terms of linear correlations between functions of their grade representations. If the underlying variables are in fact independent, then all covariances in (43) will be identically zero. Hence, testing the condition in (43) for all $j,k \geq 1$ is equivalent to testing $\delta \perp\!\!\!\perp \mathbf{w}$. As in the case of orthogonal series estimation of conditional mean functions in regression problems, it is not possible to test (43) for all values of j and k greater than one, so truncation is needed here as in the regression case.

Testing the condition in (43) involves testing a sequence of models; correlation among higher-order polynomials in (43) reflects increasingly complicated forms of dependence. Kallenberg and Ledwina (1993) note that a classical approach to the problem is to define exponential families with growing dimension and which yield sufficient statistics. They define two families:

$$h_1(\delta^*, \mathbf{w}^*) = c_1(\boldsymbol{\theta}_1) \exp\left[\sum_{j=1}^{K} \theta_{1j} \lambda_j(\delta^*) \lambda_j(\mathbf{w}^*)\right], \tag{44}$$

and

$$h_2(\delta^*, \mathbf{w}^*) = c_2(\boldsymbol{\theta}_2) \exp\left[\sum_{(j,k)\,\in\,\mathscr{A}} \theta_{2\ell} \lambda_j(\delta^*) \lambda_k(\mathbf{w}^*)\right], \tag{45}$$

where $\{\lambda_\ell(\cdot)\}$ is a set of orthonormal polynomials, $c_1(\cdot), c_2(\cdot)$ are normalizing constants, θ_1 is a K-vector of constants, $\mathscr{A}$ is a set of M pairs of indices that always includes $(1, 1)$, θ_2 is an M-vector of constants, and $\ell = 1, \ldots, M$ indexes the terms under the summation sign in (45). Under the null hypothes is $\delta \perp\!\!\!\perp \mathbf{w}$, the distribution of $(\delta^*, \mathbf{w}^*)$ is the Lebesgue measure on $[0, 1] \times [0, 1]$, and hence corresponds to $\theta_1 = 0$ and $\theta_2 = 0$ in (44)–(45).

The family in (44) imposes a certain type of symmetry in the possible dependence between $\delta^*$and $\mathbf{w}^*$ since the orthonormal polynomials in each term under the summation share the same subscript. However, note that the polynomials $\lambda_j(\cdot)$ involve powers of $\delta^*$and $\mathbf{w}^*$; so the products under the summation are not restricted to terms with the same power for δ^* and $\mathbf{w}^*$. Kallenberg and Ledwina note, however, that there exist distributions for which $E[\lambda_j(\delta^*)\lambda_j(\mathbf{w}^*)] = 0$ for all j, yet δ and $\mathbf{w}$ are dependent. But they also note that many distributions used in modeling incorporate the type of symmetry employed in (44). As an alternative, (45) relaxes this symmetry restriction.

Kallenberg and Ledwina (1993) define the set $\{\lambda_\ell(\cdot)\}$ of orthonormal polynomials as the set of orthonormal Legendre polynomials on $[0, 1]$, although other systems of orthogonal polynomials described by Szegö (1959) could be used as well. The set of orthonormal Legendre polynomials on $[0, 1]$ is given by

$$
\begin{aligned}
&\lambda_1(t) = \sqrt{3}(2t - 1)\\
&\lambda_2(t) = \sqrt{5}(6t^2 - 6t + 1)\\
&\lambda_3(t) = \sqrt{7}(20t^3 - 30t^2 + 12t - 1)\\
&\lambda_4(t) = 3(70t^4 - 140t^3 + 90t^2 - 20t + 1) \qquad (46)\\
&\qquad\vdots\\
&\lambda_\ell(t) = \ell^{-1}\sqrt{2\ell + 1}\{[(-2\ell + 1) + (4\ell - 2)t]\lambda_{\ell-1}(t) - (\ell - 1)\lambda_{\ell-2}(t)\}\\
&\qquad\vdots
\end{aligned}
$$

A smooth statistic to test $\boldsymbol{\theta}_1 = 0$ in (44) is given by

$$\widehat{T}_{7n}(K) = n^{-1/2}\sum_{j=1}^{K}\left\{\sum_{i=1}^{n}\lambda_j\left[\widehat{F}_n(\delta_i) - \frac{1}{2n}\right]\lambda_j\left[\widehat{F}_n(\mathbf{w}_i) - \frac{1}{2n}\right]\right\}^2, \qquad (47)$$

where $\widehat{F}_n(\delta_i)$ is the empirical distribution function of the δ_i defined in (27) and evaluated at δ_i, and $\widehat{F}_n(\mathbf{w}_i)$ is the empirical distribution function for the $\mathbf{w}_i$ defined in (30) and evaluated at $\mathbf{w}_i$.

To implement the test, a value must be chosen for K, which will determine the number of orthonormal polynomials that are included under the summation in (47). Kallenberg and Ledwina suggest a modified form of Schwarz' (1978) rule for model selection, where K is set to

$$
\begin{aligned}
\widehat{K} = \ &\min_K \{1 \leq K \leq d(n) | \widehat{T}_{7n}(K) - K\log n \geq \widehat{T}_{7n}(j) - j\log n,\\
&j = 1, \ldots, d(n)\}, \qquad (48)
\end{aligned}
$$

and $d(n)$ is a sequence of numbers tending to infinity as $n \to \infty$. Kallenberg and Ledwina prove that for $\widehat{K}$ defined as in (48) and $d(n) = o((n/\log n)^{1/10})$, $\widehat{K} \xrightarrow{p} 1$ and $\widehat{T}_{7n}(\widehat{K}) \xrightarrow{d} \chi^2$ with one degree of freedom. This result amounts to a first-order approximation when used in finite samples.

Unfortunately, the simulation results presented by Kallenberg and Ledwina suggest that setting $\widehat{K} = 1$ results in over-estimation of $\Pr(\widehat{T}_{7n}(\widehat{K}) \leq t)$ in finite samples. Consequently, the authors recommend an alternative, second-order approximation given by

$$\Pr\Big(\widehat{T}_{7n}(1) \le t\Big) \approx \begin{cases} \big[2\Phi\big(\sqrt{t}\big) - 1\big]\big[2\Phi(\sqrt{\log n}) - 1\big], & \text{if } t \le \log n \\ \big[2\Phi\big(\sqrt{t}\big) - 1\big]\big[2\Phi(\sqrt{\log n}) - 1\big], & \\ \quad + \Big(\frac{t - \log n}{\log n}\Big) 2\Phi(-\sqrt{\log n}), & \text{if } \log n < t < 2\log n, \\ \big[2\Phi\big(\sqrt{t}\big) - 1\big]\big[2\Phi(\sqrt{\log n}) - 1\big] & \\ \quad + 2\Phi(-\sqrt{\log n}), & \text{if } t \ge 2\log n, \end{cases} \tag{49}$$

where $\Phi(\cdot)$ denotes the standard Gaussian DF.

Recall that the statistic $\widehat{T}_{7n}(\widehat{K})$ is based on the exponential family in (44). The summand in this family is restricted to the diagonal elements of the matrix $[\lambda_j(\delta^*)\lambda_k(\mathbf{w}^*)]$, and as such is symmetric in $\delta^*$and $\mathbf{w}^*$. Consequently, tests based on $\widehat{T}_{7n}(\widehat{K})$are likely to do well at detecting situations where the correlation between $(\delta^*)^s$and $(\mathbf{w}^*)^t$ is the same as the correlation between $(\delta^*)^t$ and $(\mathbf{w}^*)^s, s, t \in \{1, 2, \ldots\}$, but may perform less well in other cases. Consequently, Kallenberg and Ledwina propose a second statistic based on the exponential family in (45), where the symmetry restriction is relaxed.

To derive the second statistic, consider sets of unique pairs of indices $\{(1,1)\}, \{(1,1),(i,j)\}, \{(1,1),(i,j),(k,\ell)\}, \ldots$, with $i,j,k,\ell,\ldots = 1,\ldots,d(n)$. Now let $\mathscr{K}$ be one of these sets of indices, and define the statistic

$$\widehat{T}_{8n}(\mathscr{K}) = \sum_{(r,s)\in\mathscr{K}} \widehat{V}(r,s), \tag{50}$$

where

$$\widehat{V}(r,s) = n^{-1}\left\{\sum_{i=1}^{n} \lambda_r\left[\widehat{F}_\delta(\delta_i) - \frac{1}{2n}\right]\lambda_s\left[\widehat{F}_{\mathbf{w}}(\mathbf{w}_i) - \frac{1}{2n}\right]\right\}^2. \tag{51}$$

To implement the test based on the statistic in (50), a particular set of indices $\widehat{\mathscr{K}}$ must be found for which $\widehat{T}_{8n}(\widehat{\mathscr{K}}) - \#(\widehat{\mathscr{K}})\log n$ is maximal, where $\#(\mathscr{K})$ denotes the number of elements in the set $\mathscr{K}$. This problem is the analog of choosing K in (47). Define $\mathscr{C}_m = \{\widehat{V}(r,s) | (r,s) \ne (1,1), r = 1,\ldots,m, s = 1,\ldots,m\}$. Let $\max_1 \mathscr{C} = \max \mathscr{C}$ denote the maximum element of a set $\mathscr{C}$ as usual, and define $\max_2 \mathscr{C} = \max(\mathscr{C} \setminus \{\max_1 \mathscr{C}\}), \max_3 \mathscr{C} = \max(\mathscr{C} \setminus \{\max_1 \mathscr{C}, \max_2 \mathscr{C}\})$, etc., where the operator $\setminus$ is defined by $\mathscr{C} \setminus \{\max_1 \mathscr{C}\} = \overline{\mathscr{C} \cap \{\max_1 \mathscr{C}\}}$. The operator $\max_j, j > 3$, can be defined recursively.

For $d(n) = 2$, it is clear that

$$\widehat{T}_{8n}(\widehat{\mathscr{K}}) = \begin{cases} \widehat{V}(1,1), & \text{if } \max_1 \mathscr{C}_2 < \log n, \\ \widehat{V}(1,1) + \max_1 \mathscr{C}_2, & \text{otherwise.} \end{cases} \tag{52}$$

For $d(n) = J, J > 2$ an integer,

$$\widehat{T}_{8n}(\widehat{\mathscr{K}}) = \begin{cases} \widehat{V}(1,1), & \text{if } \max_1 \mathscr{C}_J < \log n, \\ \widehat{V}(1,1) + \max_1 \mathscr{C}_J, & \text{if } \max_2 \mathscr{C}_J < \log n, \\ \vdots & \\ \widehat{V}(1,1) + \sum_{j=1}^{J-2} \max_j \mathscr{C}_J, & \text{if } \max_{J-1} \mathscr{C}_J < \log n, \\ \widehat{V}(1,1) + \sum_{j=1}^{J-1} \max_j \mathscr{C}_J, & \text{otherwise.} \end{cases} \tag{53}$$

Kallenberg and Ledwina (1999) prove that if $d(n) = o(\log n / \log \log n)$, then $\widehat{\mathscr{K}} \xrightarrow{p} \{(1,1)\}$and $\widehat{T}_{8n}(\widehat{\mathscr{K}}) \xrightarrow{d} \chi^2$ with one degree of freedom, but they find that this first-order approximation is even more inaccurate in finite samples than the first-order approximation for their first statistic. Alternatively, they suggest the approximation

$$\Pr\Big(\widehat{T}_{8n}((1,1)) \le t\Big) \approx \begin{cases} \big[2\Phi\big(\sqrt{t}\big) - 1\big]\big[2\Phi(\sqrt{\log n}) - 1\big]^3, & \text{if } t \le \log n, \\ \Big(\frac{t - \log n}{\log n}\Big)\Big\{\big[2\Phi(\sqrt{\log n}) - 1\big]^3\Big\} & \\ \quad + \big[2\Phi\big(\sqrt{t}\big) - 1\big]\big[2\Phi(\sqrt{\log n}) - 1\big]^3, & \text{if } \log n < t < 2\log n, \\ 1 + 2\big[\Phi\big(\sqrt{t}\big) - 1\big]\big[2\Phi(\sqrt{\log n}) - 1\big]^3, & \text{if } t \ge 2\log n, \end{cases} \tag{54}$$

based on the observation that the terms in $\widehat{T}_{8n}(\widehat{\mathscr{K}})$ are approximately chi-square, and using Fisher's normal approximation for the chi-square distribution.

4. A Graphical Method

Instead of using formal statistical tests, one can use graphical methods to assess whether δ and $\mathbf{w}$ might be dependent. A naive approach would be to examine m scatter plots for δ_i plotted against each $w_k, k = 1, \ldots, m$, but this leads to at least two problems. First, the human eye is not adept at finding complicated forms of dependence in simple scatter plots. Second, for $m > 1$, one would have to deal with multiple comparisons, which is often problematic in more formal contexts. Better approaches are based on plotting the data after a transformation to make dependence more apparent, and to eliminate the problem of multiple comparisons.

Fisher and Switzer (1985) propose a method for bivariate data that easily generalizes to the multivariate problem faced here. Their chi-plots transform the data so that rather than examining a simple scatter plot, the researcher views a plot that has distinct patterns depending on whether the data are independent, have some degree of monotone relationship, or have a more complex dependence structure.

For each $(\delta_i, \mathbf{w}_i)$, let

$$H_i = (n-1)^{-1} \sum_{j \neq i} I(\delta_j \leq \delta_i, \mathbf{w}_j \leq \mathbf{w}_i), \tag{55}$$

$$F_i = (n-1)^{-1} \sum_{j \neq i} I(\delta_j \leq \delta_i), \tag{56}$$

$$G_i = (n-1)^{-1} \sum_{j \neq i} I(\mathbf{w}_j \leq \mathbf{w}_i), \tag{57}$$

and

$$S_i = \text{sign}\left[\left(F_i - \frac{1}{2}\right)\left(G_i - \frac{1}{2}\right)\right]. \tag{58}$$

Then calculate

$$\chi_i = (H_i - F_i G_i)[F_i(1 - F_i)G_i(1 - G_i)]^{-1/2}, \tag{59}$$

and

$$\lambda_i = 4S_i \max\left\{\left(F_i - \frac{1}{2}\right)^2, \left(G_i - \frac{1}{2}\right)^2\right\}. \tag{60}$$

The value λ_i is a measure of the distance between the point $(\delta_i, \mathbf{w}_i)$ and the center of the observed data; χ_i measures the degree to which the multivariate empirical distribution function fails to factorize into a product of marginal distribution functions at the point $(\delta_i, \mathbf{w}_i)$. Each χ_i thus provides a measure of association between δ and $\mathbf{w}$ separately, at each data point. Fisher and Switzer note that the behavior of the chi-transformation in (59) will be erratic for points at the edge of the sample distribution. Thus, their chi-plot is constructed by plotting the pairs (λ_i, χ_i) for all $|\lambda_i| < 4[1/(n-1) - (1/2)]^2$. If δ and $\mathbf{w}$ are independent, then approximately $100p\%$ of the χ_i should fall between $-c_p n^{-1/2}$ and $c_p n^{-1/2}$, where $c_p = 1.54$, 1.78, and 2.18 for $p = 0.90$, 0.95, and 0.99, respectively.

Rather than constructing a single chi-plot based on δ and $\mathbf{w}$, one can also construct m separate chi-plots to examine dependence between δ and each element of $\mathbf{w}$ separately.

5. Empirical Illustrations and Monte Carlo Evidence

Charnes et al. (1981, Tables 1–4) listed 70 observations on five inputs and three outputs for an experimental public education program. These data have been used in various empirical illustrations; in particular, Simar and Wilson (2000a, Table 1) reported estimates of the Shephard (1970) input distance function obtained from the Charnes et al. data, as well as bootstrapped confidence intervals. Summary statistics for the input distance function estimates, the four input angles η_j, and the three

Table 1. Summary statistics for Charnes et al. (1981) data.

	Mean	Median	Std. Dev.	Skewness	Kurtosis
$\widehat{\delta}$	1.0522	1.0376	0.0616	1.2089	0.9613
η_1	0.3356	0.2958	0.1556	1.9688	4.0856
η_2	0.8559	0.8458	0.2103	0.4611	0.0937
η_3	0.8537	0.8638	0.2118	0.4077	− 0.0276
η_4	0.3093	0.2390	0.2158	1.5388	1.6293
y_1	25.3604	27.7400	17.4317	2.6353	11.7901
y_2	29.8299	26.2150	20.5719	2.7967	13.2793
y_3	22.4934	19.5900	13.6180	1.9933	7.4835

Table 2. Sample Pearson correlation coefficients for Charnes et al. (1981) data.

	$\widehat{\delta}$	η_1	η_2	η_3	η_4	y_1	y_2	y_3
$\widehat{\delta}$	1.0000	− 0.1945	− 0.1024	− 0.0983	− 0.1166	− 0.1984	− 0.1912	− 0.0758
	(0.0000)	(0.1067)	(0.3990)	(0.4188)	(0.3365)	(0.0997)	(0.1128)	(0.5327)
η_1	− 0.1945	1.0000	0.9270	0.9211	0.7121	− 0.2223	− 0.1755	− 0.0909
	(0.1067)	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0644)	(0.1462)	(0.4545)
η_2	− 0.1024	0.9270	1.0000	0.9850	0.6336	− 0.2439	− 0.1826	− 0.0692
	(0.3990)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0418)	(0.1303)	(0.5691)
η_3	− 0.0982	0.9211	0.9850	1.0000	0.6253	− 0.2421	− 0.1778	− 0.0714
	(0.4188)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0435)	(0.1409)	(0.5572)
η_4	− 0.1166	0.7121	0.6336	0.6253	1.0000	− 0.4215	− 0.4033	− 0.3617
	(0.3365)	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0003)	(0.0005)	(0.0021)
y_1	− 0.1984	− 0.2223	− 0.2439	− 0.2421	− 0.4215	1.0000	0.9876	0.9381
	(0.0997)	(0.0644)	(0.0418)	(0.0435)	(0.0003)	(0.0000)	(0.0001)	(0.0001)
y_2	− 0.1912	− 0.1755	− 0.1826	− 0.1778	− 0.4033	0.9876	1.0000	0.9444
	(0.1128)	(0.1462)	(0.1303)	(0.1409)	(0.0005)	(0.0001)	(0.0000)	(0.0001)
y_3	− 0.0758	− 0.0909	− 0.0692	− 0.0714	− 0.3617	0.9381	0.9444	1.0000
	(0.5327)	(0.4545)	(0.5691)	(0.5572)	(0.0021)	(0.0001)	(0.0001)	(0.0000)

Note: For each correlation coefficient ρ, the corresponding number in parentheses gives an estimate of the probability that the coefficient exceeds the estimated value in absolute value under the null hypothesis that the true correlation is zero.

Table 3. Statistics for tests of independence for Charnes et al. (1981) data.

Statistic	Estimate	p-value
$\widehat{T}_{4n}$	0.01388	0.5650
$\widehat{T}_{5n}$	0.03714	0.7325
$\widehat{T}_{7n}(1)$	3.01593	0.0825
$\widehat{T}_{8n}(1,1)$	9.39993	0.0109

Table 4. Summary statistics for Aly et al. (1990) data.

	Mean	Median	Std. Dev.	Skewness	Kurtosis
$\widehat{\delta}$	1.2605	1.0999	0.3420	1.6342	2.9355
η_1	1.4976	1.5233	0.1111	− 6.2155	48.3471
η_2	1.5698	1.5699	0.0005	− 3.2489	15.7207
y_1	10,769.9565	5,465.5000	18,783.5410	5.0210	31.1068
y_2	5,984.1832	2,112.5000	17,137.6489	8.2487	81.3536
y_3	5,530.2857	2,821.5000	16,708.3803	14.6295	241.7228
y_4	2,631.8230	1,124.5000	5,960.2010	7.2922	64.6027
y_5	7,700.1273	3,856.5000	17,292.1486	8.1386	82.3131

outputs y_k in the Charnes et al. (1981) data are shown in Table 1. Examination of the mean, median, and coefficients of skewness and kurtosis reveals that the data are clearly non-normal. This is typical of data used in productivity studies. In particular, each variable is skewed and kurtotic.

The sample Pearson correlation coefficients for the Charnes et al. (1981) data are displayed in Table 2. The sample correlation coefficients indicate a mild degree of correlation between the $\delta_i$and the first angle (η_1) and the first two outputs (y_1 and y_2). The correlation coefficients also suggest that the three outputs are almost collinear. Software packages such as SAS often report p-values for each estimated correlation coefficient $\widehat{\rho}$ for a two-sided test of the null hypothesis $H_0 : \rho = 0$. These probabilities are also reported in Table 2, and are computed by SAS software by treating the quantity $[(n-2)^{1/2}(1-\widehat{\rho}^2)^{-1/2}\widehat{\rho}]$ as a Student-t random variable with $(n-2)$ degrees of freedom. Implicitly, this involves an assumption of normality, which is not appropriate for these data as revealed by the summary statistics in Table 1. The estimated probabilities in Table 2 would suggest rejecting the null hypothesis of zero correlation between δ and $\omega_{\cdot 1}$ at 0.1 significance, but only barely; however, such tests are invalid due to the failure of the underlying normality assumption.

The situation faced at this point is not unlike what is likely to happen in many empirical studies of productive efficiency. Estimated Pearson correlation coefficients suggest no or perhaps only very mild dependence between δ and $\mathbf{w}$, but the data are clearly non-normal—zero correlation by itself is not evidence of lack of independence.

Estimates and corresponding p-values for the test statistics examined in Section 3 are shown in Table 3. The bootstrapped p-values for $\widehat{T}_{4n}$ and $\widehat{T}_{5n}$ are insignificant.[19] The statistic based on correlation integrals, $\widehat{T}_{6n}$, could not be computed; the estimated variance term in (40) was negative.[20] The rank-based tests based on $\widehat{T}_{7n}$ and $\widehat{T}_{8n}$ result in significant p-values, leading to rejection of the null hypothesis of independence.

Figure 1 shows a chi-plot for the Charnes et al. (1981) data constructed using the method described in Section 4. The horizontal dashed lines in Figure 1 are drawn at values ± 0.1841, corresponding to $c_p = 1.54$ and $p = 0.9$ as discussed in Section 4. Examination of the plot reveals 12 points outside the area between the dashed lines,

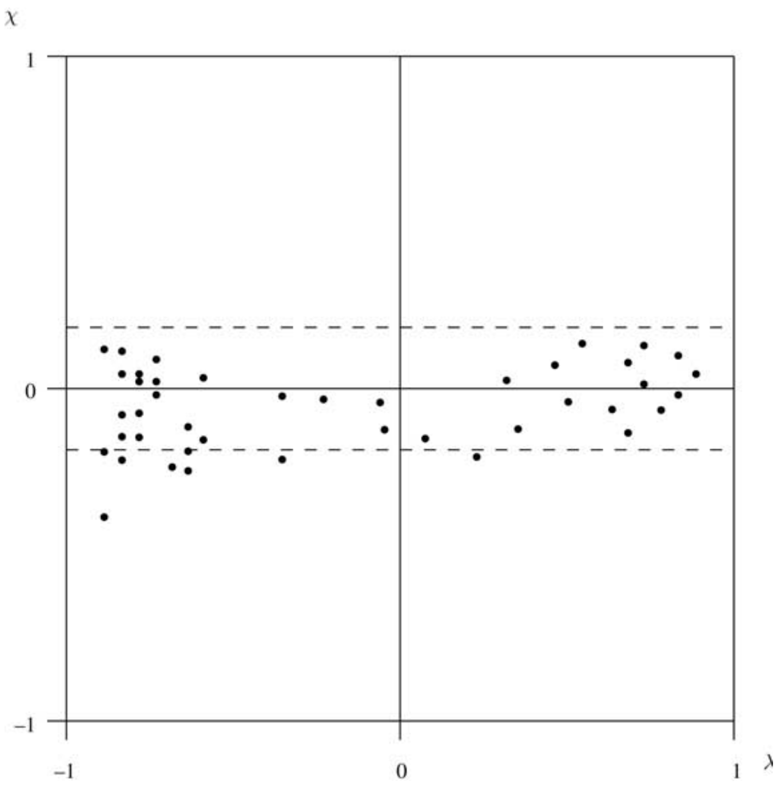

Figure 1. Chi-Plot for Charnes et al. (1981) Data.

or 17% of the observations. The chi-plot therefore suggests perhaps a mild degree of association dependence between δ and **w**.

The same tests were performed on 322 observations on commercial banks used by Aly et al. (1990). Summary statistics for these data and the corresponding input distance function estimates are given in Table 4. As with the previous data, the summary statistics suggest that the data are not normally distributed. Pearson correlation coefficients are given in Table 5, and are suggestive of rather strong correlation between δ and each of the outputs y_k as well as the first input angle η_1. As before, however, the probabilities shown below each correlation coefficient in Table 5 are based on a normality assumption that is clearly invalid here.

The tests discussed in Section 3 easily reject the null hypothesis of independence for the Aly et al. (1990) data, as shown in Table 6. The correlation-integral statistic $\widehat{T}_{6n}$ could not be computed again due to a negative variance estimate in (40). The chi-plot for the Aly et al. (1990) data appears in Figure 2, and is constructed as before. The plot reveals many observations outside the window defined by the dashed lines, with observations falling below the window on the left, and above on the right. The chi-plot also suggests dependence in the Aly et al. (1990) data.

Table 5. Sample Pearson correlation coefficients for Aly et al. (1990) data.

	$\widehat{\delta}$	η_1	η_1	y_1	y_2	y_3	y_4	y_5
$\widehat{\delta}$	1.0000	0.1279	0.0432	− 0.1759	− 0.1461	− 0.1194	− 0.1289	− 0.1518
	(0.0000)	(0.0217)	(0.4401)	(0.0015)	(0.0087)	(0.0323)	(0.0207)	(0.0063)
η_1	0.1279	1.0000	0.2186	0.1250	0.0843	0.0591	0.0634	0.0773
	(0.0217)	(0.0000)	(0.0001)	(0.0249)	(0.1312)	(0.2905)	(0.2566)	(0.1666)
η_2	0.0432	0.2186	1.0000	0.0843	− 0.0059	0.0040	0.1192	− 0.0186
	(0.4401)	(0.0001)	(0.0000)	(0.1314)	(0.9166)	(0.9426)	(0.0324)	(0.7399)
y_1	− 0.1759	0.1250	0.0843	1.0000	0.8108	0.6355	0.7336	0.8507
	(0.0015)	(0.0249)	(0.1314)	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
y_2	− 0.1461	0.0843	− 0.0059	0.8108	1.0000	0.6032	0.8201	0.8883
	(0.0087)	(0.1312)	(0.9166)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
y_3	− 0.1194	0.0591	0.0040	0.6355	0.6032	1.0000	0.6610	0.8216
	(0.0323)	(0.2905)	(0.9426)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
y_4	− 0.1289	0.0634	0.1192	0.7336	0.8201	0.6610	1.0000	0.8374
	(0.0207)	(0.2566)	(0.0324)	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0001)
y_5	− 0.1518	0.0773	− 0.0186	0.8507	0.8883	0.8216	0.8375	1.0000
	(0.0063)	(0.1666)	(0.7399)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0000)

Note: For each correlation coefficient ρ, the corresponding number in parentheses gives an estimate of the probability that the coefficient exceeds the estimated value in absolute value under the null hypothesis that the true correlation is zero.

To provide further insight, I report results from some simple Monte Carlo experiments. In each experiment, for each Monte Carlo trial, data are generated from an 8-variate normal distribution with a covariance matrix containing ones in the diagonal elements, and ρ in all the off-diagonal elements. The value of ρ was varied across experiments, with $\rho \in \{0.0, 0.1, \ldots, 0.5\}$. On each of 1,000 Monte Carlo trials, the test statistics discussed in Section 3 were computed along with their corresponding p-values. In applying $\widehat{T}_{4n}$ and $\widehat{T}_{5n}$, a univariate Epanechnikov kernel was used, with bandwidths chosen using the normal reference rule described by Silverman (1986) and Scott (1992).[21]

Table 7 reports rejection rates for each statistic for the case of $n = 70$ observations, as in the Charnes et al. (1981) data. Rejection rates are given by the number of rejections over all Monte Carlo trials of the null hypothesis of independence, divided by the total number of Monte Carlo trials, 1,000. Tests were performed at sizes $\alpha = 0.10,\ 0.05$, and 0.01.[22] With $n = 70$ observations, the results suggest that when dependence takes the form of linear correlation, the tests proposed by Kallenberg

Table 6. Statistics for tests of independence for Aly et al. (1990) data.

Statistic	Estimate	p-value
$\widehat{T}_{4n}$	0.02693	0.0060
$\widehat{T}_{5n}$	0.04271	0.0065
$\widehat{T}_{7n}(1)$	41.62840	1.104×10^{-10}
$\widehat{T}_{8n}(1,1)$	50.85242	6.198×10^{-11}

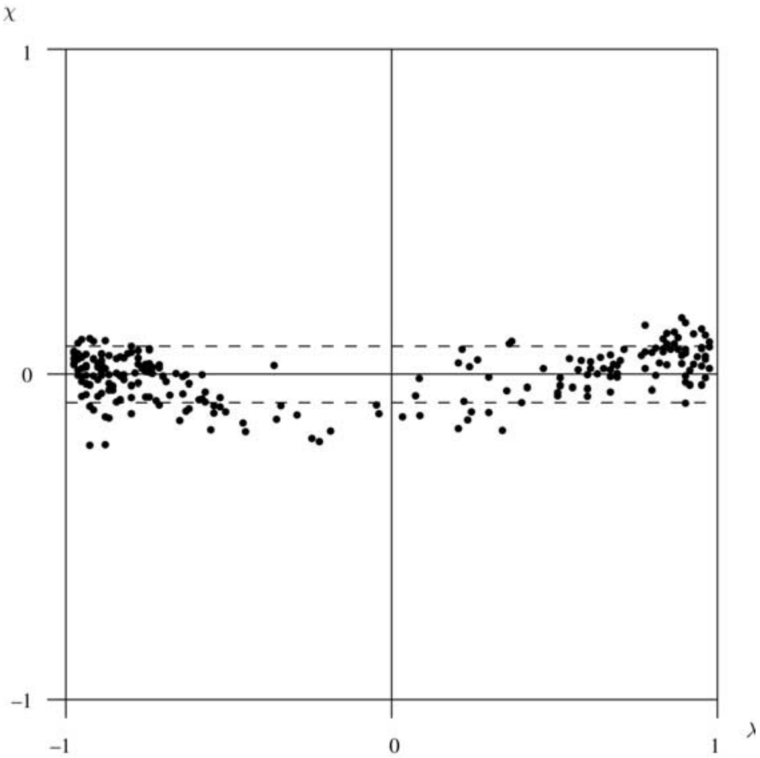

Figure 2. Chi-Plot for Aly et al. (1990) Data.

Table 7. Estimated rejection rates for independence tests (multivariate normal data, $n = 70$).

	α			α		
Statistic	0.10	0.05	0.01	0.10	0.05	0.01
		$\rho = 0.0$			$\rho = 0.3$	
$\widehat{T}_{4n}$	0.007	0.001	0.000	0.860	0.769	0.499
$\widehat{T}_{5n}$	0.058	0.029	0.004	0.739	0.584	0.301
$\widehat{T}_{7n}$	0.000	0.000	0.000	0.001	0.000	0.000
$\widehat{T}_{8n}$	0.000	0.000	0.000	0.020	0.017	0.017
		$\rho = 0.1$			$\rho = 0.4$	
$\widehat{T}_{4n}$	0.117	0.074	0.010	0.989	0.974	0.861
$\widehat{T}_{5n}$	0.186	0.112	0.029	0.930	0.856	0.594
$\widehat{T}_{7n}$	0.000	0.000	0.000	0.078	0.022	0.001
$\widehat{T}_{8n}$	0.000	0.000	0.000	0.307	0.299	0.271
		$\rho = 0.2$			$\rho = 0.5$	
$\widehat{T}_{4n}$	0.515	0.364	0.118	1.000	1.000	0.987
$\widehat{T}_{5n}$	0.431	0.301	0.102	0.991	0.969	0.871
$\widehat{T}_{7n}$	0.000	0.000	0.000	0.558	0.355	0.088
$\widehat{T}_{8n}$	0.000	0.000	0.000	0.828	0.824	0.707

and Ledwina (1999) have low power. As ρ increases from 0.0 to 0.3, the tests based on $\widehat{T}_{7n}$ and $\widehat{T}_{8n}$ reject the null hypothesis of independence in far few cases than the bootstrap tests based on $\widehat{T}_{4n}$ and $\widehat{T}_{5n}$. Even in the cases where $\rho = 0.4$ and $\rho = 0.5$, the bootstrap tests have larger power than Kallenberg and Ledwina's tests.

It should be stressed, however, that the qualitative results from this simple Monte Carlo experiment do not necessarily extend to other types of dependence. The experiment reported here merely obviates the point that the various tests of independence may perform differently. This is to be expected with other types of dependence as well; indeed, it seems unlikely that a single test of independence would perform well in all cases. It may well be the case that the tests proposed by Kallenberg and Ledwina would have better power if the departure from independence took a form different from the one in these experiments.

The chi-plot technique discussed in Section 4 does not lead to a formal test of independence, but the method was applied in the Monte Carlo experiments described above. On each Monte Carlo trial, the pairs (λ_i, χ_i) were computed, and the null hypothesis of independence was rejected if fewer than $100p\%$ of the χ_i (for which the corresponding λ_i was less than $4[1/(n-1) - 1/2]^2$ in absolute value) fell between $-c_p n^{-1/2}$ and $c_p n^{-1/2}$, where $c_p = 1.54$, 1.78, and 2.18 for $p = 0.90$, 0.95, and 0.99 as discussed in Section 4.

Rejection rates based on this approach are shown in Table 8. The results indicate that (at least) when data are multivariate normal, the chi-plot over-rejects when the null hypothesis of independence is true.[23]

Tables 9–10 report results from similar experiments with $n = 322$ observations, consistent with the Aly et al. (1990) data. Table 9 indicates that Kallenberg and Ledwina's (1999) tests perform much better than was the case in Table 7 with only 70 observations. Although the tests based on $\widehat{T}_{7n}$ and $\widehat{T}_{8n}$ under-reject in the case where $\rho = 0.0$, the tests have power comparable to the bootstrap tests for $\rho \geq 0.2$. Performance of the bootstrap tests improves also as the number of observations increases from 70 to 322. Results for the chi-plots shown in Table 10 indicate that this approach still leads to over-rejection; moreover, the problem has become worse in Table 10 than in Table 8 with only 70 observations. Consequently, I do not recommend use of the chi-plot, even though it is perhaps the easiest of the various approaches to apply.

To summarize, the lesson to be drawn from the simple Monte Carlo experiments reported in Tables 7–10 is that dependence can take many forms, and no single test

Table 8. Estimated rejection rates for chi-plots (multivariate normal data, $n = 70$).

ρ	$p = 0.90$	$p = 0.95$	$p = 0.99$
0.0	0.428	0.356	0.182
0.1	0.560	0.620	0.416
0.2	0.796	0.798	0.730
0.3	0.957	0.958	0.942
0.4	0.994	0.995	0.992
0.5	1.000	1.000	0.999

Table 9. Estimated rejection rates for independence tests (multivariate normal data, $n = 322$).

	α			α		
Statistic	0.10	0.05	0.01	0.10	0.05	0.01
		$\rho = 0.0$			$\rho = 0.3$	
$\widehat{T}_{4n}$	0.092	0.045	0.011	0.999	0.999	0.999
$\widehat{T}_{5n}$	0.097	0.045	0.006	0.999	0.998	0.987
$\widehat{T}_{7n}$	0.042	0.018	0.001	0.999	0.999	0.997
$\widehat{T}_{8n}$	0.016	0.001	0.001	0.999	0.999	0.999
		$\rho = 0.1$			$\rho = 0.4$	
$\widehat{T}_{4n}$	0.687	0.585	0.335	1.000	1.000	1.000
$\widehat{T}_{5n}$	0.517	0.377	0.154	1.000	1.000	1.000
$\widehat{T}_{7n}$	0.421	0.264	0.066	1.000	1.000	1.000
$\widehat{T}_{8n}$	0.384	0.153	0.151	1.000	1.000	1.000
		$\rho = 0.2$			$\rho = 0.5$	
$\widehat{T}_{4n}$	0.992	0.984	0.940	1.000	1.000	1.000
$\widehat{T}_{5n}$	0.964	0.922	0.744	1.000	1.000	1.000
$\widehat{T}_{7n}$	0.973	0.928	0.770	1.000	1.000	1.000
$\widehat{T}_{8n}$	0.944	0.865	0.864	1.000	1.000	1.000

statistic is likely to be robust with respect to all types of dependence in very small samples. Even in the relatively straightforward case of linear correlation among normal variates, the tests that have been considered here differ substantially in terms of their rejection rates when there are only 70 observations. Of course, the curse of dimensionality discussed in Section 2 implies that attempts at inference with only 70 observations is probably fruitless in any case with $p + q = 8$.

6. Conclusions

For both the Charnes et al. (1981) and Aly et al. (1990) data, sample correlation coefficients indicate a weak degree of dependence between δ and $\boldsymbol{w}$. However, one might reasonably question the validity of the estimated probabilities associated with the correlation coefficients shown in Tables 4 and 7, since these are based on an assumption of normality that is clearly violated by the data.

Table 10. Estimated rejection rates for chi-plots (multivariate normal data, $n = 322$).

ρ	$p = 0.90$	$p = 0.95$	$p = 0.99$
0.0	0.557	0.595	0.724
0.1	0.910	0.921	0.953
0.2	1.000	1.000	1.000
0.3	1.000	1.000	1.000
0.4	1.000	1.000	1.000
0.5	1.000	1.000	1.000

The tests described in Section 3 are fully nonparametric, and thus avoid distributional assumptions about the data. The bootstrap alternatives to the Ahmad and Li (1997) and Zheng (1997) tests appear to outperform Kallenberg and Ledwina's (1999) tests in small samples, and performances are not very different in larger samples.

Given that the homogeneous bootstrap proposed by Simar and Wilson (1998) is a restricted version of the heterogeneous bootstrap of Simar and Wilson (2000a), one should proceed cautiously in testing independence. Rejection the null hypothesis of independence clearly indicates that the heterogeneous bootstrap should be used instead of the homogeneous bootstrap. But of course, failure to reject the null hypothesis is different from acceptance of the null, and might merely reflect a lack of sufficient data to clearly indicate whether dependence is present. It is also important to remember that nonparametric efficiency estimators suffer from the well-known curse of dimensionality; in the case of the data used for illustrative purposes here, the 70 observations in the Charnes et al. (1981) data as well as the 322 observations in the Aly et al. (1990) data are probably far too few to obtain meaningful estimates of efficiency and confidence intervals with $p+1=8$ dimensions. With larger sample sizes, the nonparametric test statistics for independence would likely have increased power as well.

Acknowledgments

I thank Knox Lovell, Léopold Simar, and participants at the Workshop on DEA at The University of Southern Denmark, September 21–22, 2001, Odense, Denmark, for comments on an earlier version. Richard Grabowski graciously provided the data for one of my empirical examples some years ago. Any remaining errors are of course my fault.

Notes

1. Although the discussion throughout this paper is in terms of the input orientation, one can also use the output orientation after a straightforward change in notation.
2. For example, when $p=2$ and the input technical efficiency of an arbitrary point $(\mathbf{x}_0, \mathbf{y}_0) \in \mathscr{P}$ is estimated, the direction of the radial path along which inefficiency is estimated is defined by $\arctan(x_{01}/x_{02})$, where $\mathbf{x}_0 = [x_{01} \quad x_{02}]$. This is merely the angle that results when the Cartesian coordinates $\mathbf{x}_0$ are translated to polar coordinates.
3. The choice of direction in which to estimate technical efficiency is largely a red herring unless one has made additional, behavioral assumptions. One could also estimate technical efficiency along a hyperbolic path (see the discussion of graph efficiency measures in Färe et al., 1985); in this case, one could use the Simar and Wilson (1998) method in a straightforward way with only a small modification in notation.
4. The exposition in Simar and Wilson (2000a) is in terms of the input orientation, but the output orientation can be accommodated by straightforward changes in notation. For the case of graph efficiency, one can use the Simar and Wilson (2000a) method as usual with either an input or output orientation to generate bootstrap samples, and then apply the graph efficiency estimator

to measure distance from an original observation to the convex hull of the resulting bootstrap sample.

5. The last point cannot be shown generally since the distributions of the nonparametric input and output technical inefficiency estimators remain unknown, but is based on the intuitive notion that when independence holds, the Simar and Wilson (1998) method incorporates information that is ignored in the more general Simar and Wilson (2000a) method. But, when independence does not hold, the Simar and Wilson (1998) method involves a mis-specification of the data-generating process. Note that in parametric, stochastic frontier production function models, researchers necessarily assume that the one-sided inefficiency process is independent with respect to the left-hand side output quantity—otherwise, an endogeneity problem results—but also often assume independence with respect to inputs as well. The independence with respect to inputs is easily relaxed, however, by parameterizing the mean of the one-sided process in terms of input variables and perhaps other, environmental variables.
6. Here and throughout, inequalities involving vectors are defined on an element-by-element basis; e.g., for $\widetilde{\mathbf{x}}, \mathbf{x} \in \mathbb{R}^p_+$, $\widetilde{\mathbf{x}} \geq \mathbf{x}$ means that some, but perhaps not all or none, of the corresponding elements of $\widetilde{\mathbf{x}}$ and $\mathbf{x}$ may be equal, while some (but perhaps not all or none) of the elements of $\widetilde{\mathbf{x}}$ may be greater than corresponding elements of $\mathbf{x}$.
7. Assumption A6 is stronger than required; Kneip et al. (1998) require only Lipschitz continuity for the distance function, which is implied by the simpler, but stronger requirement in A6.
8. The term curse of dimensionality was apparently first used in this context by Bellman (1961).
9. The precise number of linear programs that must be solved with the heterogeneous bootstrap is uncertain due to the reasons given in footnote 12 of Simar and Wilson (2000a).
10. Rescaling the elements of $\mathbf{w}$ to have unit variance allows use of a single bandwidth parameter, rather than m bandwidths, in the test statistics defined below.
11. The $\widehat{\delta}_i$ are also correlated with each other in finite samples, as are estimated residuals in the regression case. In both cases, the autocorrelation disappears asymptotically, although at a slow rate in the case of the $\widehat{\delta}_i$.
12. Alternatively, $K_2(\cdot)$ could be defined to have shape give by an M-estimator of the covariance of the $\mathbf{w}_i$, as in Simar and Wilson (2000a).
13. Simar and Wilson (1998, 2000a) describe reflection methods designed to reduce the bias of kernel density estimators near the boundary at one for δ. Their purpose—drawing deviates from kernel density estimates—is different from the one here. Here, the goal is to estimate integrals; presumably, the any bias near the boundary at one will have a relatively small effect on the overall integral over the range of support of δ.
14. Hjort (1999) discusses ideas based on using Hermite expansions of mean integrated square error to select bandwidths. Such rules offer the computational ease of the normal reference rule, but allow more flexibility. Their performance remains to be discovered, however.
15. In their problem, Härdle and Mammen (1993) demonstrate that the conventional, naive bootstrap is inconsistent. The inconsistency results from the fact that the alternative, nonparametric estimator has asymptotic bias while the parametric estimator does not (under the null). Here, both terms inside the square brackets of (32) are nonparametric, and so the problem does not arise.
16. Various methods exist for choosing an appropriate bandwidth for univariate kernel density estimates; see Silverman (1986) or Scott (1992).
17. Given a vector $\mathbf{v}$ with elements $v_k, k = 1, \ldots, K, \|\mathbf{v}\|_s < \varepsilon$ if and only if $|v_k| < \varepsilon \, \forall \, k = 1, \ldots, K$.
18. Kallenberg and Ledwina (1999) provide examples which show that their method is capable of detecting forms of dependence that are not detected by Spearman's rank correlation test. Their tests can detect alternatives that are dependent but that have low or even zero linear grade correlation, while classical tests such as those based on Spearman's rank correlation coefficient do not.
19. In the bootstrap procedure for $\widehat{T}_{4n}$ and $\widehat{T}_{5n}$, I used an Epanechnikov kernel. The bandwidth was chosen using the modified normal reference rule of Hjort and Jones (1996) discussed by Fan and Gijbels (1996); the idea is based on an Edgeworth expansion of the density to be estimated around a Gaussian density, and uses terms computed from third and fourth sample moments of the data.

20. The modified version of $\widehat{T}_{6n}$ discussed by Johnson and McClelland (1998) also could not be computed for the same reason, using Gauss code provided by David S. Johnson.
21. Here, I use the normal reference rule without the adjustment suggested by Hjort and Jones (1996) since the generated data are normal. In actual applications, the data are typically not normal.
22. The null hypothesis of independence is rejected by a particular test whenever its p-value is less than the test size α.
23. As noted earlier, the chi-plot does not lead to a formal statistical test; consequently, rejection rates do not decline monotonically as one moves left to right along a particular row in Table 8; rejection on each trial is not based on an estimated p-value as was the case in Table 7.

References

Ahmad, I. A. and Q. Li. (1997). "Testing Independence by Nonparametric Kernel Method." *Statistics and Probability Letters* 34, 201–210.

Aly, H. Y., R. G. Grabowski, C. Pasurka and N. Rangan. (1990). "Technical, Scale, and Allocative Efficiencies in U.S. Banking: An Empirical Investigation." *Review of Economics and Statistics* 72, 211–218.

Bellman, R. J. (1961). *Adaptive Control Processes*. Princeton, NJ: Princeton University Press.

Blum, J. R., J. Kiefer and M. Rosenblatt. (1961). "Distribution Free Tests of Independence Based on the Sample Distribution Function." *Annals of Mathematical Statistics* 32, 485–498.

Brock, W., D. Hsieh and B. LeBaron. (1991). *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*. Cambridge, MA: MIT Press.

Charnes, A., W. W. Cooper and E. Rhodes. (1981). "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through." *Management Science* 27, 668–697.

Dette, H. and N. Neumeyer. (2000). "A Note on a Specification Test of Independence." *Metrika* 51, 133–144.

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

Färe, R., S. Grosskopf and C. A. K. Lovell. (1985). *The Measurement of Efficiency of Production*. Boston: Kluwer-Nijhoff Publishing.

Farrell, M. J. (1957). "The Measurement of Productive Efficiency." *Journal of the Royal Statistical Society, Series A* 120, 253–281.

Fisher, N. I. and P. Switzer. (1985). "Chi-plots for Assessing Dependence." *Biometrika* 72, 253–265.

Gijbels, I., E. Mammen, B. U. Park and L. Simar. (1999). "On Estimation of Monotone and Concave Frontier Functions." *Journal of the American Statistical Association* 94, 220–228.

Härdle, W. and E. Mammen. (1926–1947). "Comparing Nonparametric versus Parametric Regression Fits." *Annals of Statistics* 21.

Hall, P., J. S. Marron and B. U. Park. (1992). "Smoothed Cross Validation." *Probability Theory and Related Fields* 92, 1–20.

Hall, P., S. J. Sheather, M. C. Jones and J. S. Marron. (1991). "On Optimal Data-based Bandwidth Selection in Kernel Density Estimation." *Biometrika* 78, 263–269.

Hjort, N. L. (1999). "Towards Semiparametric Bandwidth Selectors for Kernel Density Estimators." Unpublished working paper. Department of Mathematics, Oslo, Norway: Universitetet i Oslo.

Hjort, N. L. and M. C. Jones. (1996). "Better Rules of Thumb for Choosing Bandwidth in Density Estimation." Unpublished working paper. Department of Mathematics, Oslo, Norway: Universitetet i Oslo.

Hoeffding, W. (1948). "A Non-Parametric Test of Independence." *Annals of Mathematical Statistics* 58, 546–557.

Johnson, D. and R. McClelland. (1997). "Nonparametric Tests for the Independence of Regressors and Disturbances as Specification Tests." *Review of Economics and Statistics* 79, 335–340.

Johnson, D. and R. McClelland. (1998). "A General Dependence Test and Applications." *Journal of Applied Econometrics* 13, 627–644.

Jones, M. C., J. S. Marron and S. J. Sheather. (1996). "Progress in Data-based Bandwidth Selection for Kernel Density Estimation." *Computational Statistics* 11, 337–381.

Kallenberg, W. C. M. and T. Ledwina. (1999). "Data-driven Rank Tests for Independence." *Journal of the American Statistical Association* 94, 285–301.

Kim, W. C., B. U. Park and J. S. Marron. (1994). "Asymptotically Best Bandwidth Selectors in Kernel Density Estimation." *Statistics and Probability Letters* 19, 119–127.

Kneip, A., B. U. Park and L. Simar. (1998). "A Note on the Convergence of Nonparametric DEA Estimators for Production Efficiency Scores." *Econometric Theory* 14, 783–793.

Kruskal, W. H. (1958). "Ordinal Measures of Association." *Journal of the American Statistical Association* 53, 814–861.

Lehman, E. L. (1966). "Some Concepts of Dependence." *Annals of Mathematical Statistics* 37, 1137–1153.

Robinson, P. M. (1991). "Consistent Nonparametric Entropy-Based Testing." *Review of Economic Studies* 58, 437–453.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York: John Wiley and Sons, Inc.

Schwarz, G. (1978). "Estimating the Dimension of a Model." *Annals of Statistics* 6, 461–464.

Sheather, S. J. and M. C. Jones. (1991). "A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation." *Journal of the Royal Statistical Society, Series B* 53, 683–690.

Shephard, R. W. (1970). *Theory of Cost and Production Function*. Princeton: Princeton University Press.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Simar, L. and P. W. Wilson. (1998). "Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models." *Management Science* 44(11), 49–61.

Simar, L. and P. W. Wilson. (2000a). "A General Methodology for Bootstrapping in Nonparametric Frontier Models." *Journal of Applied Statistics* 27, 779–802.

Simar, L. and P. W. Wilson. (2000b). "Statistical Inference in Nonparametric Frontier Models: The State of the Art." *Journal of Productivity Analysis* 13, 49–78.

Simar, L. and P. W. Wilson. (2001). "Aplicación del 'Bootstrap' Para Estimadores D.E.A." In A. Alvarez *La Medición de la Eficiencia y la Productividad*. Madrid: Pirámide. Translation of "Performance of the Bootstrap for DEA Estimators and Iterating the Principle." Discussion Paper No. 0002. Institut de Statistique, Louvain-la-Neuve, Belgium: Université Catholique de Louvain.

Simar, L. and P. W. Wilson. (2002a). "Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes." discussion paper no. 0307, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Simar, L. and P. W. Wilson. (2002b). "Nonparametric Tests of Returns to Scale." *European Journal of Operational Research* 139, 115–132.

Skaug, H. J. and D. Tjøstheim. (1993). "A Nonparametric Test of Serial Independence based on the Empirical Distribution Function." *Biometrika* 80, 591–602.

Sklar, A. (1959). "Fonctions de répartition à *n* dimensions et leurs marges." *Publications de l'Institut de Statistique de l'Université de Paris* 8, 229–231.

Staniswalis, J. (1989). "Local Bandwidth Selection for Kernel Estimates." *Journal of the American Statistical Association* 84, 284–288.

Szegö, G. (1959). *Orthogonal Polynomials*. New York: American Mathematical Society.

Zheng, J. X. (1997). "A Consistent Specification Test of Independence." *Nonparametric Statistics* 7, 297–306.