I N S T I T U T D E S T A T I S T I Q U E
B I O S T A T I S T I Q U E E T
S C I E N C E S A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2013/48

Testing Hypotheses in Nonparametric Models of Production

KNEIP A., SIMAR, L. and P. WILSON

Testing Hypotheses in Nonparametric Models of Production

Alois Kneip Léopold Simar Paul W. Wilson*

October 2013

Abstract

Data envelopment analysis (DEA) and free disposal hull (FDH) estimators are widely used to estimate efficiencies of production units. In applications, practitioners use DEA estimators far more frequently than FDH estimators, and thereby assume, at least implicitly, that production sets are convex. Moreover, use of the constant returns to scale (CRS) version of the DEA estimator requires an assumption of CRS. While several bootstrap methods have been developed for making inference about the efficiencies of individual units, to date no methods have existed for making consistent inference about differences in mean efficiency across groups of producers or for testing hypotheses about model structure such as returns to scale or convexity of the production set. This paper builds on central limit theorem results of Kneip et al. (2013) to develop additional theoretical results permitting consistent tests of model structure. Monte Carlo results illustrating the performance of the tests in terms of size and power are also presented. In addition, the variable returns to scale version of the DEA estimator is proved to attain the faster convergence rate of the CRS-DEA estimator under CRS.

*Kneip: Institute für Gessellschafts- und Wirtschaftswissenschaften, Statistische Abteilung, Universität Bonn, Bonn, Germany; email: akneip@uni-bonn.de.

Simar: Institut de Statistique, Biostatistique, et Sciences Actuarielles, Université Catholique de Louvain-la-Neuve, Louvain-la-Neuve, Belgium; email: leopold.simar@uclouvain.be.

Wilson: Department of Economics and School of Computing, Clemson University, Clemson, South Carolina, USA; email: pww@clemson.edu.

Research support from the "Interuniversity Attraction Pole", Phase VII (No. P7/06) of the Belgian Science Policy, from U.S. National Science Foundation (grant no. SMA-1243436), and from the Cyber Infrastructure Technology Integration group at Clemson University is gratefully acknowledged. Any remaining errors are solely our responsibility.

1 Introduction

Nonparametric efficiency estimators are widely used to benchmark producers' performance by estimating distance from a producer's location in input-output space to the boundary of the set of feasible combinations of inputs and outputs—i.e., the production set—in one of several possible directions. Estimators that involve enveloping the observed set of input-output vectors with convex sets are known in the literature as data envelopment analysis (DEA) estimators, and can be traced to the ideas of Kantorovich (1939), Koopmans (1951), and Debreu (1951). The early work by Kantorovich was suppressed by Soviet authorities; in the West, Farrell (1957) is typically credited with the first empirical application of DEA estimators. The methods were subsequently popularized by Charnes et al. (1978), Banker et al. (1984), and others. Alternatively, the free disposal hull (FDH) estimator proposed by Deprins et al. (1984) envelops observed input-output vectors with a non-convex set.

Many hundreds of examples of nonparametric efficiency estimation can be found in the literature; Gattoufi et al. (2004) list over 1,800 published studies, and internet searches find many more.[1] The statistical properties of these estimators were unknown until recently; consequently, most studies have not employed statistical inference. In recent years, however, many results have been obtained, permitting inference about the efficiency of individual producers; a recent survey of these results is provided by Simar and Wilson (2013). In addition, Kneip et al. (2013) develop new central limit theorems for means of nonparametric efficiency estimators, permitting inference about mean efficiency and convenient summarization of results.

While it is useful to make inference about the efficiency of individual producers, as well as mean efficiency among groups of producers, more is needed. This paper extends the results of Kneip et al. (2013) to develop methods for testing differences in mean efficiency across groups of producers, as well as model features such as returns to scale or convexity of the production set. Regarding tests of differences in mean efficiency across groups, one might wonder if this is not a trivial problem. However, the results of Kneip et al. (2013) make clear that standard central limit theorems do not apply (except when the number of inputs and outputs are implausibly small); in addition, it is well-known that nonparametric efficiency estimators are correlated, which introduces additional complication. As will be seen, the problem is not

[1] A search on google.com on 29 October 2013 using the keywords "dea," "efficiency," and "production" yielded approximately 5,480,000 results.

straightforward.

One might similarly wonder whether constant returns to scale (CRS) might be tested against the alternative hypothesis of non-constant, variable returns to scale (VRS) by simply comparing means of DEA estimators that impose CRS (CRS-DEA) and means of DEA estimators that permit VRS (VRS-DEA), or whether convexity versus non-convexity might be tested by comparing means of FDH and VRS-DEA estimators. However, here too the problem is complicated for the same reasons that testing whether mean efficiency is the same for two groups of producers. Testing CRS versus VRS also involves an additional complication—Park et al. (2010) prove that the CRS-DEA rate converges at rate $n^{2/(p+q)}$ under CRS, where n is the sample size and p and q give the numbers of inputs and outputs, respectively, while Kneip et al. (1998) prove that the VRS-DEA estimator converges at rate $n^{2/(p+q+1)}$ under VRS. Careful reading of Kneip et al. (2013) reveals that convergence rates play an important role in the new central limit theorems obtained there; for purposes of testing CRS versus VRS, one needs the convergence rate of the VRS-DEA estimator *under CRS,* but until now this has been unknown. Below, we present a theorem (and a proof) establishing that the VRS-DEA estimator attains the same convergence rate as the CRS-DEA estimator under CRS.

The ability to test whether the production set is convex or non-convex is crucially important in applications; DEA estimators impose convexity, and are statistically consistent only if the production set is convex. FDH estimators, on the other hand, remain consistent regardless of whether the production set is convex, but their convergence rate is slower than that of DEA estimators for a given number of inputs and outputs. If the production set is convex, it is similarly important to be able to test whether returns to scale are constant or variable. Although the theorem given below indicates that the variance of the VRS-DEA estimator is of the same *order* as the variance of the CRS-DEA estimator when CRS holds, one should expect the variance of the CRS-DEA estimator to be smaller than that of the VRS-DEA estimator under CRS, since the CRS-DEA estimator imposes CRS whereas the VRS-DEA estimator does not (i.e., the CRS-DEA estimator exploits the information that the frontier is CRS while the VRS-DEA estimator does not). Until now, the choice between FDH, CRS-DEA, and VRS-DEA estimators in many empirical studies has been largely ad-hoc. The results we provide here will allow researchers to choose the appropriate estimator for a given situation.

The paper unfolds as follows. In the next section, a statistical model is established, with requisite assumptions, and the nonparametric efficiency estimators are briefly described. In Section 3, various issues surrounding tests of differences in mean or model features are dealt with to propose specific test statistics and to derive results required for implementing the tests and making appropriate inference. The performance of the tests in finite samples is examined in a series of Monte Carlo experiments described in Section 4, and conclusions are given in the final section.

2 A Statistical Model

Denote a vector of p input quantities by $x \in \mathbb{R}_+^p$, and a vector of q output quantities by $y \in \mathbb{R}_+^q$. The production set

$$\Psi = \{(x, y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y\}, \tag{2.1}$$

gives the set of combinations of inputs and outputs that are feasible. The *technology*, or *efficient frontier* of Ψ, is given by

$$\Psi^\partial = \left\{(x, y) \in \Psi \mid \left(\gamma^{-1}x, \gamma y\right) \notin \Psi \text{ for all } \gamma > 1\right\}. \tag{2.2}$$

The Farrell (1957) input-oriented measure of technical efficiency,

$$\theta(x, y) = \inf\{\theta > 0 \mid (\theta x, y) \in \Psi\}, \tag{2.3}$$

gives the minimum feasible, proportionate reduction in input levels, holding output levels constant, for a firm operating at $(x, y) \in \Psi$. Clearly, $\theta(x, y) \in (0, 1]\ \forall\ (x, y) \in \Psi$; if $\theta(x, y) = 1$, the firm is said to be technically efficient in the input direction. Alternatively, if $\theta(x, y) < 1$, the firm is said to be technically *in*efficient.

Technical efficiency can be also be measured in output, hyperbolic, or arbitrary, linear directions toward the frontier as discussed by Simar and Wilson, (2000, 2013), Wilson (2011), Simar and Vanhems (2012), and Simar et al. (2012). To conserve space, the analysis below is presented in terms of the input-oriented measure in (2.3); it is trivial (but perhaps tedious) to extend all of the results that follow to the other directions by simply adapting the notation.

Various assumptions on the production set Ψ can be made, but typical assumptions (e.g., Shephard, 1970; Färe, 1988; Simar and Wilson (2000); etc.) include the following.

Assumption 2.1. Ψ *is closed, and* Ψ^∂ *exists.*

Assumption 2.2. *Both inputs and outputs are strongly disposable; i.e., for* $\widetilde{x} \geq x,\ 0 \leq \widetilde{y} \leq y$, *if* $(x, y) \in \Psi$ *then* $(\widetilde{x}, y) \in \Psi$ *and* $(x, \widetilde{y}) \in \Psi$.[2]

Strong disposability in Assumption 2.2 implies weak monotonicity for the frontier, and is standard in micro-economic theory of the firm. Additional assumptions about the structure of Ψ or Ψ^∂ are often made. For example, in studies where DEA estimators are employed, Ψ is assumed (often implicitly) to be convex. Where CRS-DEA estimators are used, Ψ^∂ is assumed to be characterized by constant returns to scale everywhere (e.g., Charnes et al., 1978; etc.). As noted in the introduction, these assumptions have typically been ad-hoc. Such assumptions should be tested.

Of course, the set Ψ is unobserved, and hence must be estimated from a sample $\mathcal{X}_n = \{(X_i, Y_i)\}_{i=1}^n$ of observed input-output pairs $X_i \in \mathbb{R}_+^p$, $Y_i \in \mathbb{R}_+^q$. The free-disposal hull of the sample observations in $\mathcal{X}_n$, i.e.,

$$\widehat{\Psi}_{\text{FDH}}(\mathcal{X}_n) = \bigcup_{(X_i,Y_i)\in\mathcal{X}_n} \left\{(x, y) \in \mathbb{R}^{p+q} \mid y \leq Y_i,\ x \geq X_i\right\}, \tag{2.4}$$

was proposed by Deprins et al. (1984) to estimate Ψ. Replacing Ψ with $\widehat{\Psi}_{\text{FDH}}(\mathcal{X}_n)$ on the right-hand side (RHS) of (2.3) yields the FDH estimator $\widehat{\theta}_{\text{FDH}}(x, y \mid \mathcal{X}_n)$ of $\theta(x, y)$.[3]

Alternatively, if Ψ is convex, then Ψ can be estimated by

$$\widehat{\Psi}_{\text{VRS}}(\mathcal{X}_n) = \left\{(x, y) \in \mathbb{R}^{p+q} \mid y \leq \boldsymbol{Y}\boldsymbol{\omega},\ x \geq \boldsymbol{X}\boldsymbol{\omega},\ \boldsymbol{i}_n'\boldsymbol{\omega} = 1,\ \boldsymbol{\omega} \in \mathbb{R}_+^n\right\}, \tag{2.5}$$

where $\boldsymbol{X} = \big(X_1,\ \ldots,\ X_n\big)$ and $\boldsymbol{Y} = \big(Y_1,\ \ldots,\ Y_n\big)$ are $(p \times n)$ and $(q \times n)$ matrices of input and output vectors, respectively; $\boldsymbol{i}_n$ is an $(n \times 1)$ vector of ones, and $\boldsymbol{\omega}$ is a $(n \times 1)$ vector of weights. This is the convex hull of $\widehat{\Psi}_{\text{FDH}}(\mathcal{X}_n)$, and is called the VRS-DEA estimator of Ψ. Replacing Ψ on the RHS of (2.3) with $\widehat{\Psi}_{\text{VRS}}(\mathcal{X}_n)$ yields the VRS-DEA estimator of $\theta(x, y)$.

If Ψ^∂ exhibits globally constant returns to scale (CRS), i.e, if $(ax, ay) \in \Psi$ for all $(x, y) \in \Psi$ and $a \in [0, \infty)$, then Ψ can be estimated by the conical hull of $\widehat{\Psi}_{\text{FDH}}(\mathcal{X}_n)$ obtained by

[2] Note that as usual, inequalities involving vectors are defined on an element-by-element basis.

[3] Afriat (1972, Theorem 1.1) defines a left- (but not right-) continuous function similar to the FDH estimator $\widehat{\Psi}_{\text{FDH}}(\mathcal{X}_n)$ for the case $p \geq 1,\ q = 1$. Note, however, that $\widehat{\Psi}_{\text{FDH}}(\mathcal{X}_n)$ is not a function, and is defined for arbitrary $p \geq 1$ as well as $q \geq 1$. Moreover, Afriat's function does not permit measurement of efficiency in the input direction, nor (in general) in hyperbolic or directional orientations.

dropping the constraint $\boldsymbol{i}_n'\boldsymbol{\omega} = 1$ from the RHS of (2.5); denote this estimator by $\widehat{\Psi}_{\text{CRS}}(\mathcal{X}_n)$. Again using the plug-in principle, replacing Ψ on the RHS of (2.3) with $\widehat{\Psi}_{\text{CRS}}(\mathcal{X}_n)$ yields the CRS-DEA estimator $\widehat{\theta}_{\text{CRS}}(x, y \mid \mathcal{X}_n)$ of $\theta(x, y)$. By construction, for a given sample $\mathcal{X}_n$, $\widehat{\Psi}_{\text{FDH}}(\mathcal{X}_n) \subseteq \widehat{\Psi}_{\text{VRS}}(\mathcal{X}_n) \subseteq \widehat{\Psi}_{\text{CRS}}(\mathcal{X}_n)$.

Computation of the FDH and DEA efficiency estimators is straightforward. FDH efficiency estimates can be computed as

$$\widehat{\theta}_{\text{FDH}}(x, y) = \min_{i \in \mathcal{I}(y)} \left(\max_{j=1,\, \ldots,\, p} \left(\frac{X_i^j}{x^j} \right) \right), \tag{2.6}$$

where $\mathcal{I}(y) = \{i \mid y_i \geq y,\ i = 1,\ \ldots,\ n\}$ and X_i^j, x^j are the jth elements of X_i and x, respectively (throughout, subscripts will be used to index different vectors, while superscripts will be used to index elements of vectors). DEA efficiency estimates are typically computed by solving linear programs; for the VRS-DEA estimator, one can compute

$$\widehat{\theta}_{\text{VRS}}(\boldsymbol{x}, \boldsymbol{y}) = \min_{\theta, \boldsymbol{\omega}} \left\{ \theta \mid y \leq \boldsymbol{Y}\boldsymbol{\omega},\ \theta x \geq \boldsymbol{X}\boldsymbol{\omega},\ \boldsymbol{i}_n'\boldsymbol{\omega} = 1,\ \boldsymbol{\omega} \in \mathbb{R}_+^n \right\}. \tag{2.7}$$

The CRS-DEA estimator $\widehat{\theta}_{\text{CRS}}(x, y \mid \mathcal{X}_n)$ can be computed similarly by dropping the constraint $\boldsymbol{i}_n'\boldsymbol{\omega} = 1$ on the RHS of (2.7).

Asymptotic properties of FDH efficiency estimators are given by by Park et al. (2000) and Daouia et al. (2013). Asymptotic properties of VRS-DEA efficiency estimators are investigated in Kneip et al. (1998), Jeong (2004), Jeong and Park (2006), Kneip et al. (2008), while asymptotic properties of CRS-DEA efficiency estimators are examined in Park et al. (2010). Under appropriate assumptions, each estimator is consistent and converges at rate n^{κ}, where $\kappa = 1/(p + q)$, $2/(p + q + 1)$, or $2/(p + q)$ for the FDH, VRS-DEA, and CRS-DEA cases, respectively. In addition, each estimator has a non-degenerate limiting distribution. These results have been extended to the hyperbolic and directional orientations by Wilson (2011), Simar and Vanhems (2012), and Simar et al. (2012), with similar rates of convergence and limiting distributions.

Additional, technical assumptions required for each central limit theorem result (for means of FDH, VRS-DEA, and CRS-DEA estimates) established by Kneip et al. (2013) and used below are given in the Appendix, in Section A.1:.

3 Testing Issues in Nonparametric Frontier Models

3.1 Testing the equality of the mean of 2 groups of firms

Suppose the researcher is confronted with two independent samples of sizes n_1 and n_2 of firms belonging to groups labeled G_1 and G_2. In such situations, it is natural to test whether $\mu_{1,\theta} = E(\theta(X, Y) \mid (X, Y) \in G_1)$ and $\mu_{2,\theta} = E(\theta(X, Y) \mid (X, Y) \in G_2)$ are equal against the alternative that, say, Group 1 is more efficient than Group 2, i.e. $\mu_{1,\theta} > \mu_{2,\theta}$.[4] More formally, one might test the null hypothesis $H_0\colon \mu_{1,\theta} = \mu_{2,\theta}$ versus the alternative hypothesis $H_1\colon \mu_{1,\theta} > \mu_{2,\theta}$.

Testing for mean efficiency across two groups was suggested—but not implemented—in the pioneering application of Charnes et al. (1981), who considered two groups of schools, one receiving a treatment effect and the other not receiving the treatment. To give additional examples where such a test might be useful, one might test whether mean efficiency among for-profit producers is greater than mean efficiency of non-profit producers in studies of hospitals, banks and credit unions, or perhaps other industries. One might similarly be interested in comparing average performance of publicly-traded versus privately-held firms, or in regional differences that might reflect variation in state-level regulation or other industry features.

Suppose iid samples $\mathcal{X}_{1,n_1} = \{(X_i, Y_i)\}_{i=1}^{n_1}$ and $\mathcal{X}_{2,n_2} = \{(X_i, Y_i)\}_{i=1}^{n_2}$ of input-output pairs from groups 1 and 2 (respectively) are available. In addition, assume these samples are independent of each other. The two samples yield independent estimators

$$\widehat{\mu}_{1,n_1} = n_1^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}_{1,n_1}} \widehat{\theta}(X_i, Y_i \mid \mathcal{X}_{1,n_1}) \tag{3.1}$$

and

$$\widehat{\mu}_{2,n_2} = n_2^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}_{2,n_2}} \widehat{\theta}(X_i, Y_i \mid \mathcal{X}_{2,n_2}) \tag{3.2}$$

of $\mu_{1,\theta}$ and $\mu_{2,\theta}$, respectively; the conditioning indicates the sample used to compute the efficiency estimates under the summation signs. In addition, the subscripts on $\widehat{\theta}(\cdot)$ have been dropped; either the FDH, VRS-DEA, or CRS-DEA estimators with corresponding convergence rates n^κ could be used, although the same estimator would be used for both groups. Theorem 4.1 of Kneip et al. (2013) establishes (under appropriate regularity conditions; see Section A.1:

[4] *Mutatis mutandis*, alternative tests with a two sided alternative or with other measures of efficiency would follow the same procedure.

for details) consistency and other properties of these estimators. The same theorem, however, makes clear that standard, conventional central limit theorems can be used to make inference about the population means $\mu_{1,\theta}$ and $\mu_{2,\theta}$ only when the dimensionality $(p+q)$ is small enough so that $\kappa > 1/2$ due to the bias of the estimators $\widehat{\mu}_{1,n_1}$ and $\widehat{\mu}_{2,n_2}$. The assumptions required for consistency of $\widehat{\mu}_{1,n_1}$ and $\widehat{\mu}_{2,n_2}$ are decreasingly restrictive as one moves from the CRS-DEA case to the VRS-DEA case, and finally to the FDH case. The presentation in the remainder of this sub-section is in terms of the VRS-DEA case; the results extend easily to the other cases with appropriate changes in assumptions and notation.

Kneip et al. (2013) use a bias estimate to develop new central limit theorems for making inference about mean efficiency. First, divide the sample for group $\ell \in \{1,\ 2\}$ by setting $m_{\ell,1} = [n_\ell/2]$ and $m_{\ell,2} = n_\ell - [n_\ell/2]$, where $[a]$ denotes the integer part of a. Set $k = 1$. Then let $\mathcal{X}^{(1)}_{\ell,m_{\ell,1},k}$ denote a random subset of size $m_{\ell,1}$ of observed input-output pairs in $\mathcal{X}_{\ell,n_\ell}$, and let $\mathcal{X}^{(2)}_{\ell,m_{\ell,2},k}$ be the set of remaining input-output pairs in $\mathcal{X}_{\ell,n_\ell}$ so that $\mathcal{X}^{(1)}_{\ell,m_{\ell,1},k} \cap \mathcal{X}^{(2)}_{\ell,m_{\ell,2},k} = \emptyset$ and $\mathcal{X}^{(1)}_{\ell,m_{\ell,1},k} \cup \mathcal{X}^{(2)}_{\ell,m_{\ell,2},k} = \mathcal{X}_{\ell,n_\ell}$. Hence the samples $\mathcal{X}_{\ell,n_\ell}$ are split evenly where n_ℓ is even, or almost evenly (with a difference of one observation) where n_ℓ is odd. Now let

$$\widehat{\mu}^{(j)}_{\ell,m_{\ell,j},k} = (m_{\ell,j})^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}^{(j)}_{\ell,m_{\ell,j},k}} \widehat{\theta}\left(X_i, Y_i \mid \mathcal{X}^{(j)}_{\ell,m_{\ell,j},k}\right) \tag{3.3}$$

for $j \in \{1, 2\}$. Define

$$\widetilde{\mu}^{*}_{\ell,n_\ell,k} = 0.5\left(\widehat{\mu}^{(1)}_{\ell,m_{\ell,1},k} + \widehat{\mu}^{(2)}_{\ell,m_{\ell,2},k}\right) \tag{3.4}$$

and

$$\widetilde{B}_{\ell,\kappa,n_\ell,k} = (2^\kappa - 1)^{-1}\left(\widetilde{\mu}^{*}_{\ell,n_\ell,k} - \widehat{\mu}_{\ell,n_\ell}\right). \tag{3.5}$$

Of course, for group $\ell \in \{1,\ 2\}$ with n_ℓ observations, there are $\binom{n_\ell}{n_\ell/2}$ possible splits of the sample. To reduce the variation of the bias estimate in (3.5), the above steps can be repeated, shuffling the observations before each split of the two samples, for $k = 1,\ \ldots,\ K$ with $K << \left(\binom{n_1}{n_1/2} \wedge \binom{n_2}{n_2/2}\right)$. Then set

$$\widehat{B}_{\ell,\kappa,n_\ell} = K^{-1}\sum_{k=1}^{K} \widetilde{B}_{\ell,\kappa,n_\ell,k}. \tag{3.6}$$

This gives a jackknife estimate of bias.[5]

[5] In many cases, one might use a delete-one or a delete-k jackknife with samples of size $n - k$ to correct for bias. For our purposes, however, the jackknife samples must be a fixed, constant, multiplicative factor of n in order for the result in Theorem 4.3 of Kneip et al., 2013 to hold.

Theorem 4.3 of Kneip et al. (2013) establishes, under appropriate regularity conditions and provided $p+q \leq 4$ when VRS-DEA estimators are used,

$$\sqrt{n_\ell}\left(\widehat{\mu}_{\ell,n_\ell} - \widehat{B}_{\ell,\kappa,n_\ell} - \mu_{\ell,\theta} + R_{\ell,n_\ell,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2_{\ell,\theta}) \tag{3.7}$$

for the two groups $\ell \in \{1,2\}$, where $R_{\ell,n_\ell,\kappa} = o(n_\ell^{-\kappa})$ and $\sigma_{\ell,\theta} = \text{VAR}(\theta(X,Y) \mid (X,Y) \in G_\ell)$. If CRS-DEA estimators are used and Ψ^∂ is globally CRS, then the result holds for $p+q \leq 5$. On the other hand, if FDH estimators are used, the result is valid only for $p+q \leq 3$. See Kneip et al. (2013) for additional details.

Alternatively, if $p+q > 4$ and VRS-DEA estimators are used (or if $p+q > 5$ with CRS-DEA estimators, or $p+q > 3$ with FDH estimators), then Theorem 4.4 of Kneip et al. (2013) is applicable. For $\ell \in \{1,2\}$, let $n_{\ell,\kappa} = [n_\ell^{2\kappa}]$; then $n_{\ell,\kappa} < n_\ell$ for $\kappa < 1/2$. Let $\mathcal{X}^*_{\ell,n_{\ell,\kappa}}$ be a random subset of $n_{\ell,\kappa}$ input-output pairs from $\mathcal{X}_{\ell,n_\ell}$. Then let

$$\widehat{\mu}_{\ell,n_{\ell,\kappa}} = n_{\ell,\kappa}^{-1} \sum_{(X_{\ell,i},Y_{\ell,i}) \in \mathcal{X}^*_{\ell,n_{\ell,\kappa}}} \widehat{\theta}\left(X_{\ell,i}, Y_{\ell,i}) \mid \mathcal{X}_{\ell,n_\ell}\right), \tag{3.8}$$

noting that while the summation is over only the input-output pairs in $\mathcal{X}^*_{\ell,n_{\ell,\kappa}}$, the efficiency estimates under the summation sign are computed using all of the input-output pairs in $\mathcal{X}_{\ell,n_\ell}$. Then by Kneip et al. (2013, Theorem 4.4), for each group $\ell = 1,2$,

$$n_\ell^{\kappa}\left(\widehat{\mu}_{\ell,n_{\ell,\kappa}} - \widehat{B}_{\ell,\kappa,n_\ell} - \mu_{\ell,\theta} + R_{\ell,n_\ell,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2_{\ell,\theta}) \tag{3.9}$$

under suitable regularity conditions.

For all values of $p+q$, Theorem 4.1 of Kneip et al. (2013) indicates that the variances $\sigma^2_{\ell,\theta}$ are estimated consistently by the sample variances $\widehat{\sigma}^2_{\ell,\theta,n_\ell}$ within each group $\ell \in \{1,2\}$; i.e.,

$$\widehat{\sigma}^2_{\ell,\theta,n_\ell} = n_\ell^{-1} \sum_{i=1}^{n_\ell} \left[\widehat{\theta}(X_{\ell,i}, Y_{\ell,i} \mid \mathcal{X}_\ell) - \widehat{\mu}_{\ell,n_\ell}\right]^2 \xrightarrow{p} \sigma^2_\ell. \tag{3.10}$$

The independence of the two samples plays a crucial role, and avoids complications due to covariances.

It is well-known that two sequences of independent variables, each having a normal limiting distribution, possess a joint limiting bivariate normal distribution with independent marginals given by the individual normal limits. Consequently, the difference of the two random, independent sequences has a limiting normal distribution given by the difference of the two normal limits. Therefore, using VRS-DEA estimators with $p+q \leq 4$ (or CRS-DEA estimators with $p+q \leq 5$, or FDH estimators with $p+q \leq 3$),

$$\widehat{\tau}_{1,n_1,n_2} = \frac{(\widehat{\mu}_{1,n_1} - \widehat{\mu}_{2,n_2}) - \left(\widehat{B}_{1,\kappa,n_1} - \widehat{B}_{2,\kappa,n_2}\right) - (\mu_{1,\theta} - \mu_{2,\theta})}{\sqrt{\frac{\widehat{\sigma}^2_{1,\theta,n_1}}{n_1} + \frac{\widehat{\sigma}^2_{2,\theta,n_2}}{n_2}}} \xrightarrow{\mathcal{L}} N(0,1), \tag{3.11}$$

provided $n_1/n_2 \to c > 0$ as $n_1, n_2 \to \infty$, where c is a constant. Then if $\widehat{p} = 1 - \Phi\left(\widehat{\tau}_{1,n_1,n_2}\right)$ is sufficiently small, perhaps less than .1, .05, or .01 and where $\Phi(\cdot)$ denotes the standard normal distribution function, one may reject the null hypothesis $H_0\colon \mu_{1,\theta} = \mu_{2,\theta}$ in favor of the alternative hypothesis $H_1\colon \mu_{1,\theta} > \mu_{2,\theta}$. One could also use (3.11) to construct confidence intervals for $(\mu_{1,\theta} - \mu_{2,\theta})$.

In situations where $p+q > 4$ with VRS-DEA estimators (or $p+q > 5$ with CRS-DEA estimators, or $p+q > 3$ with FDH estimators), a similar test statistic can be obtained using (3.9) in place of equations (3.7). Using similar reasoning, it is easy to see that

$$\widehat{\tau}_{2,n_{1,\kappa},n_{2,\kappa}} = \frac{\left(\widehat{\mu}_{1,n_{1,\kappa}} - \widehat{\mu}_{2,n_{2,\kappa}}\right) - \left(\widehat{B}_{1,\kappa,n_1} - \widehat{B}_{2,\kappa,n_2}\right) - (\mu_{1,\theta} - \mu_{2,\theta})}{\sqrt{\frac{\widehat{\sigma}^2_{1,\theta,n_1}}{n_{1,\kappa}} + \frac{\widehat{\sigma}^2_{2,\theta,n_2}}{n_{2,\kappa}}}} \xrightarrow{\mathcal{L}} N(0,1), \tag{3.12}$$

again provided provided $n_1/n_2 \to c > 0$ as $n_1, n_2 \to \infty$. Note that the same estimates for the variances and biases are used in (3.12) as in (3.11). The only difference between (3.11) and (3.12) is in the number of observations used to compute the sample means.

3.2 Testing returns to scale

Unlike the situation described in Section 3.1 where the researcher faces two independent groups of observations and wants to test whether mean efficiency is the same in the two groups, one may face a single iid sample $\mathcal{X}_n = \{(X_i, Y_i\}_{i=1}^n$ of n input-output pairs and wish to test the null hypothesis of constant returns to scale versus the alternative hypothesis of variable returns to scale. Under the alternative hypothesis, Ψ is strictly convex, while under the null, Ψ is only weakly convex. Under the null, both the VRS-DEA and CRS-DEA estimators of $\theta(X,Y)$ are consistent, but under the alternative, only the VRS-DEA estimator is consistent.

Consider the sample means

$$\widehat{\mu}^{\text{full}}_{\text{VRS},n} = n^{-1} \sum_{i=1}^{n} \widehat{\theta}_{\text{VRS}}\left(X_i, Y_i \mid \mathcal{X}_n\right) \tag{3.13}$$

and

$$\widehat{\mu}^{\text{full}}_{\text{CRS},n} = n^{-1}\sum_{i=1}^{n}\widehat{\theta}_{\text{CRS}}\left(X_i, Y_i \mid \mathcal{X}_n\right) \tag{3.14}$$

computed using all of the n observations in $\mathcal{X}_n$. By construction, $\widehat{\theta}_{\text{CRS}}\left(X_i, Y_i \mid \mathcal{X}_n\right) \leq \widehat{\theta}_{\text{VRS}}\left(X_i, Y_i \mid \mathcal{X}_n\right) \leq 1$ and hence $\widehat{\mu}^{\text{full}}_{\text{VRS},n} - \widehat{\mu}^{\text{full}}_{\text{CRS},n} \geq 0$. Under the null, one would expect $\widehat{\mu}^{\text{full}}_{\text{VRS},n} - \widehat{\mu}^{\text{full}}_{\text{CRS},n}$ to be "small," while under the alternative $\widehat{\mu}^{\text{full}}_{\text{VRS},n} - \widehat{\mu}^{\text{full}}_{\text{CRS},n}$ is expected to be "large."

Clearly, the variance of the difference $\widehat{\mu}^{\text{full}}_{\text{VRS},n} - \widehat{\mu}^{\text{full}}_{\text{CRS},n}$ is $\text{VAR}\left(\widehat{\mu}^{\text{full}}_{\text{VRS},n}\right) + \text{VAR}\left(\widehat{\mu}^{\text{full}}_{\text{CRS},n}\right) - 2\text{COV}\left(\widehat{\mu}^{\text{full}}_{\text{VRS},n},\ \widehat{\mu}^{\text{full}}_{\text{CRS},n}\right)$. By Theorem 4.1 of Kneip et al. (2013), the first two terms sum to $2n^{-1}\sigma^2_\theta$, and under the null

$$\begin{aligned}
\text{COV}\left(\widehat{\mu}^{\text{full}}_{\text{VRS},n},\ \widehat{\mu}^{\text{full}}_{\text{CRS},n}\right) &= \text{COV}\left[\left(\widehat{\mu}^{\text{full}}_{\text{VRS},n} - E\left(\widehat{\mu}^{\text{full}}_{\text{VRS},n}\right)\right)\left(\widehat{\mu}^{\text{full}}_{\text{CRS},n} - E\left(\widehat{\mu}^{\text{full}}_{\text{CRS},n}\right)\right)\right] \\
&= \text{COV}\left[\left(\overline{\theta}_n - \mu_\theta + o_p\left(n^{-1/2}\right)\right),\ \left(\overline{\theta}_n - \mu_\theta + o_p\left(n^{-1/2}\right)\right)\right] \\
&= \text{VAR}\left(\overline{\theta}_n\right) + o\left(n^{-1}\right) \\
&= n^{-1}\sigma^2_\theta + o\left(n^{-1}\right)
\end{aligned} \tag{3.15}$$

under the null, where $\overline{\theta}_n = n^{-1}\sum_{i=1}^{n}\theta(X_i, Y_i)$, which is unobserved. Consequently, a test statistic using the difference in the sample means given by (3.13)–(3.14) will have a degenerate distribution under the null since the asymptotic variance of $\left(\widehat{\mu}^{\text{full}}_{\text{VRS},n} - \widehat{\mu}^{\text{full}}_{\text{CRS},n}\right)$ is zero. In other words, the density of $n^a\left(\widehat{\mu}^{\text{full}}_{\text{VRS},n} - \widehat{\mu}^{\text{full}}_{\text{CRS},n}\right)$ collapses to a Dirac delta function at zero for any power $a \leq 1/2$ of n. This is true regardless of the dimensionality $(p+q)$.

In order to obtain non-degenerate test statistics, randomly split the sample into two samples $\mathcal{X}_{1,n_1}$, $\mathcal{X}_{2,n_2}$ such that $\mathcal{X}_{1,n_1} \cup \mathcal{X}_{2,n_2} = \mathcal{X}_n$ and $\mathcal{X}_{1,n_1} \cap \mathcal{X}_{2,n_2} = \emptyset$, where $n_1 = [n/2]$ and $n_2 = n - n_1$. Next, let

$$\widehat{\mu}_{\text{VRS},n_1} = n_1^{-1}\sum_{(X_i,Y_i)\in\mathcal{X}_{1,n_1}}\widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_{1,n_1}) \tag{3.16}$$

and

$$\widehat{\mu}_{\text{CRS},n_2} = n_2^{-1}\sum_{(X_i,Y_i)\in\mathcal{X}_{2,n_2}}\widehat{\theta}_{\text{CRS}}(X_i, Y_i \mid \mathcal{X}_{2,n_2}). \tag{3.17}$$

In addition, let

$$\widehat{\sigma}^2_{\text{VRS},n_1} = n_1^{-1}\sum_{(X_i,Y_i)\in\mathcal{X}_{1,n_1}}\left[\widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_{1,n_1}) - \widehat{\mu}_{\text{VRS},n_1}\right]^2 \tag{3.18}$$

and

$$\widehat{\sigma}^2_{\text{CRS},n_2} = n_2^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}_{2,n_2}} \left[\widehat{\theta}_{\text{CRS}}(X_i, Y_i \mid \mathcal{X}_{2,n_2}) - \widehat{\mu}_{\text{CRS},n_2}\right]^2. \tag{3.19}$$

Theorem 4.1 of Kneip et al. (2013) establishes that both $\widehat{\mu}_{\text{VRS},n_1}$ and $\widehat{\mu}_{\text{CRS},n_2}$ are consistent estimators of $\mu_\theta = E(\theta(X,Y))$ under the null hypothesis of CRS, and that both (3.18) and (3.19) consistently estimate the variances of the VRS and CRS efficiency estimators.

Park et al. (2010) prove that the CRS efficiency estimator converges at rate $n^{2/(p+q)}$ under CRS, whereas Kneip et al. (1998) prove that the VRS efficiency estimator converges at rate $n^{2/(p+q+1)}$ under variable (but not constant) returns to scale. The following theorem establishes the convergence rate of the VRS efficiency estimator when Ψ^∂ is globally CRS; this is needed to construct bias corrections similar to those used above in Section 3.1 and in Kneip et al. (2013).

The theorem that follows gives some new and unexpected results. Among other things, the theorem establishes that when Ψ^∂ is globally CRS, the VRS-DEA estimator attains the faster convergence rate of the CRS-DEA estimator.

Theorem 3.1. *Under Assumptions 2.1–A.1, A.4, and A.6, the following conditions hold:*

(i) For any fixed (x,y) in the interior of $\mathcal{D}$

$$\widehat{\theta}_{VRS}(x, y \mid \mathcal{X}_n) - \theta(x,y) = O_P(n^{-\frac{2}{p+q}}). \tag{3.20}$$

(ii) If $p+q=2$, then $E\left[\widehat{\theta}_{VRS}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i,Y_i)\right] = O(n^{-1}\log n)$. If $p+q>2$, then there exists a constant $0 < D_1 < \infty$ such that for all $i,j \in \{1,\ldots,n\}$, $i \neq j$,

$$E\left[\widehat{\theta}_{VRS}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i,Y_i)\right] = D_1 n^{-\frac{2}{p+q}} + O\left(n^{-\frac{3}{p+q}}(\log n)^{\frac{p+q+3}{p+q}}\right), \tag{3.21}$$

$$VAR\left(\widehat{\theta}_{VRS}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i,Y_i)\right) = O\left(n^{-\frac{3}{p+q}}(\log n)^{\frac{3}{p+q}}\right), \tag{3.22}$$

and

$$\left|COV\Big(\widehat{\theta}_{VRS}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i,Y_i), \widehat{\theta}_{VRS}(X_j, Y_j \mid \mathcal{X}_n) - \theta(X_j,Y_j)\Big)\right|$$
$$= O\left(n^{-\frac{p+q+1}{p+q}}(\log n)^{\frac{p+q+1}{p+q}}\right) = o\left(n^{-1}\right). \tag{3.23}$$

The value of the constant D_1 depends on f and on the structure of the set $\mathcal{D} \subset \Psi$.

A proof is given in the Appendix, in Section A.2:.

By virtue of Theorem 3.1, we can build a test statistic as follows. First, in order to construct the bias corrections, set $k = 1$ and split each of the two subsamples $\mathcal{X}_{\ell,n_\ell}$, $\ell \in \{1, 2\}$ randomly into two mutually exclusive and collectively exhaustive parts $\mathcal{X}^{(1)}_{\ell,m_{\ell,1,k}}$ and $\mathcal{X}^{(2)}_{\ell,m_{\ell,2,k}}$ as described above in Section 3.1. For each part $j \in \{1, 2\}$ of $\mathcal{X}_{1,n_1,k}$, compute

$$\widehat{\mu}^{(j)}_{\mathrm{VRS},m_{1,j},k} = m^{-1}_{1,j} \sum_{(X_i,Y_i)\in\mathcal{X}^{(j)}_{1,m_{1,j}}} \widehat{\theta}_{\mathrm{VRS}} \left(X_i, Y_i \mid \mathcal{X}^{(j)}_{1,m_{1,j},k} \right). \tag{3.24}$$

Similarly, for each part $j \in \{1, 2\}$ of $\mathcal{X}_{2,n_2,k}$, compute

$$\widehat{\mu}^{(j)}_{\mathrm{CRS},m_{2,j},k} = m^{-1}_{2,j} \sum_{(X_i,Y_i)\in\mathcal{X}^{(j)}_{2,m_{2,j}}} \widehat{\theta}_{\mathrm{CRS}} \left(X_i, Y_i \mid \mathcal{X}^{(j)}_{2,m_{2,j},k} \right). \tag{3.25}$$

Then let

$$\widetilde{\mu}^{*}_{\mathrm{VRS},n_1,k} = 0.5 \left(\widehat{\mu}^{(1)}_{\mathrm{VRS},m_{1,1},k} + \widehat{\mu}^{(2)}_{\mathrm{VRS},m_{1,2},k} \right) \tag{3.26}$$

and

$$\widetilde{\mu}^{*}_{\mathrm{CRS},n_2,k} = 0.5 \left(\widehat{\mu}^{(1)}_{\mathrm{CRS},m_{2,1},k} + \widehat{\mu}^{(2)}_{\mathrm{CRS},m_{2,2},k} \right). \tag{3.27}$$

Analogous to (3.5), compute (for the kth split)

$$\widetilde{B}_{\mathrm{VRS},\kappa,n_1,k} = (2^\kappa - 1)^{-1} \left(\widetilde{\mu}^{*}_{\mathrm{VRS},n_1,k} - \widehat{\mu}_{\mathrm{VRS},n_1} \right) \tag{3.28}$$

and

$$\widetilde{B}_{\mathrm{CRS},\kappa,n_2,k} = (2^\kappa - 1)^{-1} \left(\widetilde{\mu}^{*}_{\mathrm{CRS},n_2,k} - \widehat{\mu}_{\mathrm{CRS},n_2} \right), \tag{3.29}$$

where $\kappa = 2/(p+q)$. For $\ell \in \{1,\ 2\}$, shuffle the observations in the subsamples $\mathcal{X}_{\ell,n_\ell}$ and split again; then repeat the above steps for $k = 2,\ \ldots,\ K$. Finally, the necessary bias corrections are given by

$$\widehat{B}_{\mathrm{VRS},\kappa,n_1} = K^{-1} \sum_{k=1}^{K} \widetilde{B}_{\mathrm{VRS},\kappa,n_1,k} \tag{3.30}$$

and

$$\widehat{B}_{\mathrm{CRS},\kappa,n_1} = K^{-1} \sum_{k=1}^{K} \widetilde{B}_{\mathrm{CRS},\kappa,n_1,k}. \tag{3.31}$$

Under the null hypothesis of constant returns to scale, and following the reasoning used in Section 3.1, Theorem 4.2 of Kneip et al. (2013) together with Theorem 3.1 given above ensure that

$$\widehat{\tau}_{3,n} = \frac{(\widehat{\mu}_{\text{VRS},n_1} - \widehat{\mu}_{\text{CRS},n_2}) - \left(\widehat{B}_{\text{VRS},\kappa,n_1} - \widehat{B}_{\text{CRS},\kappa,n_2}\right)}{\sqrt{\frac{\widehat{\sigma}^2_{\text{VRS},n_1}}{n_1} + \frac{\widehat{\sigma}^2_{\text{CRS},n_2}}{n_2}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{3.32}$$

provided $(p+q) \leq 5$.

Alternatively, if $(p+q) > 5$, the sample means must be computed using subsets of the available observations. For $\ell \in \{1,2\}$ and $\mathcal{X}^*_{\ell,n_{\ell,\kappa}}$ defined as in Section 3.1, let

$$\widehat{\mu}_{\text{VRS},n_{1,\kappa}} = n_{1,\kappa}^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}^*_{1,n_{1,\kappa}}} \widehat{\theta}(X_i, Y_i \mid \mathcal{X}_{1,n_1}) \tag{3.33}$$

and

$$\widehat{\mu}_{\text{CRS},n_{2,\kappa}} = n_{2,\kappa}^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}^*_{2,n_{2,\kappa}}} \widehat{\theta}(X_i, Y_i \mid \mathcal{X}_{2,n_2}). \tag{3.34}$$

As in (3.8), the summations in (3.33)–(3.34) are over subsets of the observations used to compute the efficiency estimates under the summation signs. Again under the null hypothesis of constant returns to scale, by Theorem 4.4 of Kneip et al. (2013) and Theorem 3.1 that appears above ensure

$$\widehat{\tau}_{4,n} = \frac{\left(\widehat{\mu}_{\text{VRS},n_{1,\kappa}} - \widehat{\mu}_{\text{CRS},n_{2,\kappa}}\right) - \left(\widehat{B}_{\text{VRS},\kappa,n_1} - \widehat{B}_{\text{CRS},\kappa,n_2}\right)}{\sqrt{\frac{\widehat{\sigma}^2_{\text{VRS},n_1}}{n_{1,\kappa}} + \frac{\widehat{\sigma}^2_{\text{CRS},n_2}}{n_{2,\kappa}}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{3.35}$$

for $(p+q) > 5$.

Depending on the value of $(p+q)$, either $\widehat{\tau}_{3,n}$ or $\widehat{\tau}_{4,n}$ can be used to test the null hypothesis of constant returns to scale, with critical values obtained from the standard normal distribution. In particular, for $j \in \{3,\ 4\}$, the null hypothesis of constant returns to scale is rejected if $\widehat{p} = 1 - \Phi(\widehat{\tau}_{j,n})$ is less than, say, .1, .05, or .01.

3.3 Testing convexity of the attainable set

Situations where one might want to test whether the production set Ψ is convex versus non-convex resemble the situation in Section 3.2 in that the researcher is faced with a single iid sample $\mathcal{X}_n = \{(X_i, Y_i)\}_{i=1}^n$. Under the null hypothesis of convexity, both the FDH and VRS-DEA estimators are consistent, but under the alternative, only the FDH estimator is consistent. It might be tempting to compute the sample mean

$$\widehat{\mu}_{\text{FDH},n}^{\text{full}} = n^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}_n} \widehat{\theta}_{\text{FDH}}\left(X_i, Y_i \mid \mathcal{X}_n\right) \tag{3.36}$$

using the full set of observations in $\mathcal{X}_n$ and use this with (3.13) to construct a test statistic based on the difference $\widehat{\mu}_{\text{VRS},n}^{\text{full}} - \widehat{\mu}_{\text{FDH},n}^{\text{full}}$. By construction, $\widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_n) \leq \widehat{\theta}_{\text{FDH}}(X_i, Y_i \mid \mathcal{X}_n) \leq 1$ and therefore $\widehat{\mu}_{\text{FDH},n}^{\text{full}} - \widehat{\mu}_{\text{VRS},n}^{\text{full}} \geq 0$. Under the null, $\widehat{\mu}_{\text{VRS},n}^{\text{full}} - \widehat{\mu}_{\text{FDH},n}^{\text{full}}$ is expected to be "small," while under the alternative the difference is expected to be "large."

Such an approach is doomed to failure for reasons similar to those given at the beginning of Section 3.2. Using Theorem 4.1 of Kneip et al. (2013) and reasoning similar to the argument at the beginning of Section 3.2, it is easy to show that $n^a\left(\widehat{\theta}_{\text{FDH}}(X_i, Y_i \mid \mathcal{X}_n) - \widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_n)\right)$ converges under the null to a degenerate distribution for any power $a \leq 1/2$ of n; i.e., the asymptotic variance of the statistic is zero, and the density of the statistic converges to a Dirac delta function at zero under the null.

As in Section 3.2, the sample $\mathcal{X}_n$ can be repeatedly split into two parts $\mathcal{X}_{1,n_1}$ and $\mathcal{X}_{2,n_2}$ such that $\mathcal{X}_{1,n_1} \cap \mathcal{X}_{2,n_2} = \emptyset$ and $\mathcal{X}_{1,n_1} \cup \mathcal{X}_{2,n_2} = \mathcal{X}_n$. Here, however, the two efficiency estimators have different convergence rates under the null. The FDH estimator converges at rate $n^{1/(p+q)}$ (Park et al., 2000), while the VRS-DEA estimator converges at rate $n^{2/(p+q+1)}$ under strict convexity (Kneip et al., 1998), or at rate $n^{2/(p+q)}$ under weak convexity by Theorem 3.1. This difference can be exploited by setting $n_1^{2/(p+q+1)} = n_2^{1/(p+q)}$ and $n_1 + n_2 = n$ for a given sample size n, and then solving for n_1 and n_2. There is no closed-form solution, but it is easy to find a numerical solution by writing $n - n_1 - n_1^{2(p+q)/(p+q+1)} = 0$; the root of this equation is bounded between 0 and $n/2$, and can be found by simple bisection. Letting n_1 equal the integer part of the solution and setting $n_2 = n - n_1$ gives the desired subsample sizes with $n_2 > n_1$. Using the larger subsample $\mathcal{X}_{2,n_2}$ to compute the FDH estimates and the smaller subsample $\mathcal{X}_{1,n_1}$ to compute the VRS-DEA estimates allocates observations from the original sample $\mathcal{X}_n$ efficiently in the sense that more observations are used to mitigate the slower convergence rate of the FDH estimator.

Once the original sample has been split, compute $\widehat{\mu}_{\text{VRS},n_1}$ using (3.16) and

$$\widehat{\mu}_{\text{FDH},n_2} = n_2^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}_{2,n_2}} \widehat{\theta}_{\text{FDH}}\left(X_i, Y_i \mid \mathcal{X}_{2,n_2}\right). \tag{3.37}$$

In addition, let

$$\widehat{\sigma}^2_{\text{FDH},n_2} = n_2^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}_{2,n_2}} \left[\widehat{\theta}_{\text{FDH}}(X_i, Y_i \mid \mathcal{X}_{2,n_2}) - \widehat{\mu}_{\text{FDH},n_2}\right]^2. \tag{3.38}$$

Theorem 4.1 of Kneip et al. (2013) establishes that both $\widehat{\mu}_{\text{VRS},n_1}$ and $\widehat{\mu}_{\text{FDH},n_2}$ are consistent estimators of $\mu_\theta = E(\theta(X,Y))$ under the null hypothesis of convexity, and that both $\widehat{\sigma}^2_{\text{VRS},n_1}$ and $\widehat{\sigma}^2_{\text{FDH},n_2}$ given in (3.18) and (3.38) consistently estimate the variances of the VRS and FDH efficiency estimators.

In order to construct the bias corrections, for $k = 1, \ldots, K$, split each of the two subsamples $\mathcal{X}_{\ell,n_\ell}$, $\ell \in \{1,2\}$ randomly into two mutually exclusive and collectively exhaustive parts $\mathcal{X}^{(1)}_{\ell,m_{\ell,1},k}$ and $\mathcal{X}^{(2)}_{\ell,m_{\ell,2},k}$ as described above in Section 3.1. Compute $\widehat{B}_{\text{VRS},\kappa_1,n_1}$ as described in Section 3.2 using (3.24), (3.26), and (3.28) with $\kappa_1 = 2/(p+q+1)$ replacing κ. For each part $j \in \{1,2\}$ of $\mathcal{X}_{2,n_2,k}$, compute

$$\widehat{\mu}^{(j)}_{\text{FDH},m_{2,j},k} = m_{2,j}^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}^{(j)}_{2,m_{2,j},k}} \widehat{\theta}_{\text{FDH}}\left(X_i, Y_i \mid \mathcal{X}^{(j)}_{2,m_{2,j},k}\right). \tag{3.39}$$

and let

$$\widetilde{\mu}^*_{\text{FDH},n_2,k} = 0.5\left(\widehat{\mu}^{(1)}_{\text{FDH},m_{2,1,k}} + \widehat{\mu}^{(2)}_{\text{FDH},m_{2,2,k}}\right). \tag{3.40}$$

Then compute

$$\widetilde{B}_{\text{FDH},\kappa_2,n_2,k} = \left(2^{\kappa_2} - 1\right)^{-1}\left(\widetilde{\mu}^*_{\text{FDH},n_2,k} - \widehat{\mu}_{\text{FDH},n_2}\right), \tag{3.41}$$

using (3.39) and (3.40), and where $\kappa_2 = 1/(p+q)$. Finally, compute the FDH bias correction

$$\widehat{B}_{\text{FDH},\kappa_2,n_2} = K^{-1}\sum_{k=1}^{K} \widetilde{B}_{\text{FDH},\kappa_2,n_2,k}. \tag{3.42}$$

Under the null hypothesis of convexity of Ψ, and following the reasoning used in Sections 3.1–3.2, Theorem 4.2 of Kneip et al. (2013) ensures that

$$\widehat{\tau}_{5,n} = \frac{\left(\widehat{\mu}_{\text{FDH},n_2} - \widehat{\mu}_{\text{VRS},n_1}\right) - \left(\widehat{B}_{\text{FDH},\kappa_2,n_2} - \widehat{B}_{\text{VRS},\kappa_1,n_1}\right)}{\sqrt{\frac{\widehat{\sigma}^2_{\text{FDH},n_2}}{n_2} + \frac{\widehat{\sigma}^2_{\text{VRS},n_1}}{n_1}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{3.43}$$

provided $(p+q) \leq 3$ since the FDH convergence rate dominates that of the VRS-DEA estimator.

Alternatively, if $(p+q) > 3$, the sample means must be computed using subsets of $\mathcal{X}_{1,n_1}$ and $\mathcal{X}_{2,n_2}$. For $\ell \in \{1,2\}$, let $\kappa = \kappa_2 = 1/(p+q)$ and let $\mathcal{X}_{\ell,n_{\ell,\kappa}}$ be defined as in Section 3.1. Compute $\widehat{\mu}_{\text{VRS},n_{1,\kappa}}$ using (3.33), and compute

$$\widehat{\mu}_{\text{FDH},n_{2,\kappa}} = n_{2,\kappa}^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}^*_{2,n_{2,\kappa}}} \widehat{\theta}(X_i, Y_i \mid \mathcal{X}_{2,n_2}). \tag{3.44}$$

Here again, as in (3.33) and (3.34), the summation in (3.44) is over a subset of the observations used to compute the efficiency estimates under the summation sign. Then under the null hypothesis of convexity for Ψ,

$$\widehat{\tau}_{6,n} = \frac{\left(\widehat{\mu}_{\text{FDH},n_{2,\kappa}} - \widehat{\mu}_{\text{VRS},n_{1,\kappa}}\right) - \left(\widehat{B}_{\text{FDH},\kappa_2,n_2} - \widehat{B}_{\text{VRS},\kappa_1,n_1}\right)}{\sqrt{\frac{\widehat{\sigma}^2_{\text{FDH},n_2}}{n_{2,\kappa}} + \frac{\widehat{\sigma}^2_{\text{VRS},n_1}}{n_{1,\kappa}}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{3.45}$$

for $(p+q) > 3$ by Theorem 4.4 of Kneip et al. (2013).

Depending on whether $(p+q) \leq 3$ or $(p+q) > 3$, either $\widehat{\tau}_{5,n}$ or $\widehat{\tau}_{6,n}$ can be used to test the null hypothesis of constant returns to scale, with critical values obtained from the standard normal distribution. In particular, for $j \in \{5,6\}$, the null hypothesis of convexity of Ψ is rejected if $\widehat{p} = 1 - \Phi(\widehat{\tau}_{j,n})$ is less than a suitably small value, e.g., .1, .05, or .01.

4 Monte Carlo Evidence

4.1 Experimental Framework

We perform three sets of Monte Carlo experiments to examine the performance of the tests described above in Section 3. In the first set of experiments, we consider the size and power properties of the test of equality of mean efficiency across two groups. In the next two sets of experiments, we consider size and power properties of (i) the test of convexity of the production set Ψ and (ii) returns to scale of the technology Ψ^∂.

In the experiments examining the test of mean efficiency across two groups of observations, we consider sample sizes $n_1 = n_2 \in \{50,\ 100,\ 200,\ 1,000,\ 10,000,\ 20,000\}$ and data-generating processes (DGPs) with $p = q = 1$, $p = q = 2$, and $p = q = 3$. In the experiments with the returns to scale and convexity tests, we consider individual sample sizes

$n = \{50,\ 100,\ 200,\ 1,000,\ 10,000,\ 20,000\}$ and DGPs with $q = 1$ and $p \in \{1,\ 2,\ 3,\ 4,\ 5\}$.[6] In each experiment, we perform 1,000 Monte Carlo trials. On each Monte Carlo trial, we simulate data from a known, "true" model, compute the relevant test statistic, and then compute the corresponding p-value using the standard normal quantile function. In the tables that follow, we report the proportion (among 1,000 Monte Carlo trials) of cases where we reject the null hypothesis of equivalent means, constant returns to scale, or convexity of Ψ in tests of nominal sizes .10, .05, and .01.

In the first set of experiments, where we test the equality of mean efficiency across two samples of sizes $n_1 = n_2$, we simulate data by first generating $(p+q)$-tuples $\boldsymbol{u} = \left[\boldsymbol{u}_p', \boldsymbol{u}_q'\right]'$ uniformly distributed on a unit sphere centered at the origin in $\mathbb{R}^{p+q}$, where $\boldsymbol{u}_p$ and $\boldsymbol{u}_q$ are vectors of length p and q, respectively. We then set $\boldsymbol{x} = (1 - |\boldsymbol{u}_p|)\theta^{-1}$ and $\boldsymbol{y} = |\boldsymbol{u}_q|$, where θ is a draw from the distribution with density

$$f(t \mid \lambda_k) = \begin{cases} \lambda_k t^{-2} e^{-\lambda_k (t^{-1}-1)} & \forall\ t \in (0,1], \\ 0 & \text{otherwise}, \end{cases} \tag{4.1}$$

with $k = 1$ or 2 depending on whether observations are generated for sample 1 or 2.[7] We set $\lambda_1 = 2$ and consider $\lambda_2 \in \{2.0,\ 1.9,\ 1.8,\ \ldots,\ 1.0,\ 0.75,\ 0.5\}$.

In the second set of experiments examining the returns-to-scale test, we model the technology by

$$Y = g\left(\prod_{j=1}^{p}\left(\widetilde{X}^j\right)^{1/p}\right), \tag{4.2}$$

where $\widetilde{X}^j$ is the efficient level of the jth input and the function $g()\colon \mathbb{R}^1_+ \mapsto \mathbb{R}^1_+$ is either homogeneous of degree 1 under the null hypothesis of CRS, or is not homogeneous under the alternative hypothesis of variable returns to scale. To describe the function $g(\cdot)$, consider the transformation $(x, y) \mapsto (s, t)$ such that $s = \sqrt{2} - \frac{x+y}{\sqrt{2}}$ and $t = \frac{x-y}{\sqrt{2}}$. In (s,t)-space, the coordinate system relative to that in (x,y)-space has been rotated through an clockwise angle of $3\pi/4$ radians and then shifted by a distance of $\sqrt{2}$ along a 45-degree ray from the origin in

[6] Of course, situations involving more than one output can be easily handled using our methods; here, we use only one output to simplify the process of simulating data. In all of the theoretical results about properties of DEA and FDH estimators, including Korostelev et al. (1995a, 1995b), Park et al. (2000), Park et al. (2010), Kneip et al. (1998), Kneip et al., (2008, 2011), Gijbels et al. (1999), and Wilson (2011), it is the dimensionality $(p+q)$ rather than the ratio p/q that is important for determining properties of the estimators; consequently, we expect no loss of generality from simulating only one output.

[7] Note that $1/\theta$ has exponential density $f(t \mid \lambda_k) = \lambda_k e^{\lambda_k (t-1)}\ \forall\ t \in [1, \infty)$.

(x, y)-space. In (s, t)-space, the function $g(\cdot)$ corresponds to

$$t = c\left(a^2 + \delta^2 s^2\right)^{1/2} - d \tag{4.3}$$

where $a = 0.5$, $c = 0.75$, $d = 0.375$, and $\delta \in \{0.0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.6,\ 0.8,\ 1.0,\ 1.2,\ 1.4\}$. The transformation from (x, y)-space to (s, t)-space is easily inverted; given a point (s, t), $x = \frac{\sqrt{2}-s+t}{\sqrt{2}}$ and $y = \frac{\sqrt{2}-s-t}{\sqrt{2}}$.

The function $g(\cdot)$ is illustrated in Figure 1 for the various values of δ. For $\delta = 0$, (4.3) is a flat, horizontal line in (s, t)-space so that $t = 0\ \forall s$ as depicted in the first panel of Figure 1; in (x, y)-space, this corresponds to a 45-degree line from the origin as illustrated in the second panel of Figure 3.2. Setting $\delta > 0$ results in a convex (from below) curve in (s, t)-space, and a concave (from below) curve in (x, y)-space, with curvature increasing with δ as shown in both panels of Figure 1. In the left-hand panel of Figure 1, the triangle with corners at $(-\sqrt{2}, 0)$, $(\sqrt{2}, 0)$, and $(0, \sqrt{2}, 0)$ formed by the dashed lines and the horizontal solid line corresponds to the triangle with corners at (0,0), $(0, 2)$, and $(2, 2)$ in the right-hand panel. For δ strictly greater than 0 but less than about 1.41, $|\frac{\partial t}{\partial s}|$ calculated from (4.3) is less than one, and hence $g(\cdot)$ is monotonically increasing for in (x, y)-space within the triangle described above and depicted in the right-hand panel of Figure 1.

Data for the returns-to-scale experiments are generated by first computing for a value of δ in the set given above, the corresponding value of s, denoted $s_{\max}$, where the curve in (4.3) intersects the line $t = \sqrt{2} - s$, and then generating uniform random numbers s_i on the interval $\left(\max(-s_{\max}, -0.9\sqrt{2}),\ \min(s_{\max}, 0.9\sqrt{2})\right)$. Plugging these into (4.3) for s gives corresponding values t_i; pairs (s_i, t_i) are then transformed to pairs $(\widetilde{X}_i, Y_i)$ using the inverse transformation described above. If $p = 1$, then $X_i = \theta^{-1}\widetilde{X}_i$ where θ is a draw from the density in (4.1) parameterized by setting $\lambda = 2$. If $p > 1$, then generate a pair (s_i, t_i) as before and transform to a pair (V_i, Y_i) using the same inverse transformation describe above (here, the scalar $\widetilde{X}_i$ has been relabeled V_i). Generate a $(p \times 1)$ vector $\boldsymbol{u}$ of uniform deviates on $(0,\ 1)$. Then set $V_i = \prod_{j=1}^{p}\left(\widetilde{X}_i^j\right)^{1/p}$; in terms of (4.2), we have $Y_i = g(V_i)$. Now write $p \log V_i = \sum_{j=1}^{p} \log \widetilde{X}_i^j = \boldsymbol{i}_p'\boldsymbol{W}_i$, where $\boldsymbol{W}_i$ is a p-vector with jth element $\log \widetilde{X}_i^j$. Finally, set $X_i = \theta^{-1} \exp \boldsymbol{W}_i = \theta^{-1} \exp\left(\frac{\boldsymbol{u}}{\boldsymbol{i}_p'\boldsymbol{u}} p \log V_i\right)$, where θ is a draw from the density in (4.1), again with $\lambda = 2$. The vector $\boldsymbol{u}$ of uniform deviates serves to divide the scalar quantity $p \log V_i$ into p additive components, which are transformed to efficient input levels, and then projected away from the frontier Ψ^∂ by multiplying by θ^{-1}.

A similar simulation strategy is used for the third set of experiments that examine performance the convexity test. The technology is again described by (4.2), but the function $g()\colon \mathbb{R}^1_+ \mapsto \mathbb{R}^1_+$ is redefined. Here, the transformation $(x, y) \mapsto (s, t)$ is such that $s = t - \sqrt{2} + x\sqrt{2}$ and $t = \frac{y-x}{\sqrt{2}}$; then $x = \frac{s-t+\sqrt{2}}{\sqrt{2}}$ and $y = \frac{t+s}{\sqrt{2}} + 1$. Hence in (s, t)-space, the coordinate system relative to that in (x, y)-space has been rotated through a counterclockwise angle of $\pi/4$ radians and then shifted by a distance of $\sqrt{2}$ along a 45-degree ray from the origin in (x, y) space. Data are simulated as described above for the returns to scale test with the same values of δ, resulting in the function g depicted in Figure 2. In addition, a strictly convex (from above) version of g is simulated using $\delta = 1.4$ and the transformation used to generate data for the experiments with the returns to scale test.

In each of the three sets of experiments, $K = 100$ subsample splits were used to compute the bias corrections.

4.2 Results of simulation experiments

Monte Carlo estimates of rejection rates for the two-sample test using the VRS-DEA estimator with nominal test sizes of .1, .05, and .01 are shown in Table 1 for 2-6 dimensions. The test statistic $\widehat{\tau}_{1,n_1,n_2}$ in (3.11) is used in the first three sets of results depicted in Table 1, where $(p + q) = 2$, 3, and 4. The test statistic $\widehat{\tau}_{2,n_1,n_2}$ given in (3.12) is used in the last two sets of results in Table 1, where $(p + q) = 5$ and 6. For each sample size, there are 13 rows of results, with the first row giving rejection rates when the null is true, and rows 2–13 giving rejection rates for increasing departures from the null.

A broad overview of the results in Table 1 indicates that for a given nominal test-size, the realized rejection rate approaches the nominal size as sample sizes increase when the null is true. In addition, for a given departure from the null, the power of the test increases with sample size (although not monotonically in every case). In addition, looking left to right in the table, there is a noticeable improvement in terms of achieved size of the tests while moving from 4 to 5 dimensions, i.e., when the statistic $\widehat{\tau}_{2,n_1,n_2}$ begins to be used. This improvement comes at the expense of decreased power, however, reflecting the usual tradeoff between size and power of a test.

For 200 or fewer observations, the results in Table 1 indicate that the realized test sizes are too large; but with 1,000 observations, the realized test sizes are close to their corresponding nominal levels. Although one might prefer a conservative test to a less-conservative test in small samples, to the extent that realized sizes do not match nominal sizes, one can compensate by choosing smaller test-sizes, or making appropriate caveats when just barely rejecting the null in real-world applications. The good news is that the difference between realized and nominal test sizes is smaller for 5-6 dimensions, which are common in applications, than for 2–4 dimensions.

The next set of results given in Table 2 illustrates the performance of the returns to scale test, again for 2–6 dimensions (recall that here, $q = 1$, while $p = 1$ through 5). The test statistic $\widehat{\tau}_{3,n}$ given in (3.32) is used for 2–5 dimensions, and the test statistic $\widehat{\tau}_{4,n}$ given in (3.35) is used when there are 6 dimensions. Looking at the table, overall conclusions similar to those drawn for the test of equivalent means can be drawn: the returns to scale test improves in terms of size and power as sample size increases, and while there is a price to pay for increasing the number of dimensions, there is a noticeable improvement in the size (but at the expense of reduced power) of the test when going from 5 to 6 dimensions, i.e., when the statistic $\widehat{\tau}_{4,n}$ can be used.

It is interesting to note that in the experiments for the equivalent means test, *two* samples of sizes 50, 100, ... were generated, whereas in the experiments for the returns to scale test, only *one* sample of size 50, 100, ... was generated. Comparing the results for $n = 100$ in Table 2 with the results for $n_1 = n_2 = 50$ in Table 1 suggests, that for 100 total observations in either case, the size-performance of the returns to scale test is slightly better than that for the means test for 2 or 3 dimensions, and considerably better for 4 dimensions. Similar observations hold for $n = 200$ in Table 2 versus $n_1 = n_2 = 100$ in Table 1. Apparently, it is "easier" to test for constant versus variable returns to scale than to test whether mean efficiencies are equal.

Table **??** gives results for the simulations for the convexity test. For each sample size, 11 rows give rejection rates at the three nominal test sizes considered and the five different dimensionalities. The first row in each case corresponds to the case where $g()$ is strictly convex (from above), while the second row in each case corresponds to the case where $g()$ is linear, i.e., weakly convex. Rows 2–11 correspond to increasing departures from the null hypothesis of convexity.

Overall conclusions similar to those drawn in the two previous sets of results can be drawn from the results in Table **??**, too; i.e., realized size improves with increasing sample size, and there is a price to pay for increasing dimensionality, except in going from 3 to 4 dimensions, where the test statistic $\widehat{\tau}_{6,n}$ given in (3.45) begins to be used instead of $\widehat{\tau}_{5,n}$ given in (3.43), there is a noticeable improvement in performance in terms of size (but not power). At sample sizes of $n = 1,000$ or less, when $g()$ is strictly convex, the realized test sizes are much closer to the corresponding nominal sizes than when $g()$ is only weakly convex. For example, with $n = 50$, at the five-percent level, the estimated rejection rates for $p + q = 2$ and 3 are 8.6 and 10.1 percent where $g()$ is strictly convex, but 29.8 and 31.1 percent where $g()$ is only weakly convex. The situation is somewhat better with large dimensionality, however; for $p + q = 4$, 5, and 6, the estimated rejection rates are 15.3 to 16.2 percent at the five-percent level when $g()$ is only weakly convex. The estimated rejection rates improve (i.e., move closer to the nominal rates) as sample size increases, but even with $n = 20,000$, the rates are larger than the nominal values when $g()$ is only weakly convex, and slightly so even if $g()$ is strictly convex.

Nonetheless, the results in Table **??** suggest that if one is comfortable working at the 5 percent level in ordinary situations, he might want to work at the 1 percent level when testing convexity. Or, if estimated p-values are used, the researcher should be wary of rejecting the null hypothesis of convexity if the estimated p value is, for example, 0.06–0.01. But if the estimated p value is, e.g., of order 10^{-4}, it seems reasonable to reject convexity. As is often the case in hypothesis testing, some caution is warranted.

As noted above at the end of Section 4.1, the experiments whose results are displayed in Tables 1–**??** were conducted while averaging the bias-corrections discussed in Section 3 over $K = 100$ random splits of the samples (or subsamples, in the case of the returns to scale and convexity tests). Doing this, however, adds to the computational burden. To examine the potential gains from the averaging, we also conducted experiments along the lines described above, but with no averaging of the bias corrections (i.e., in terms of the notation in Section 3, with $K = 1$). In order to conserve space, we do not include the results of these experiments here, but they are available from the authors on request.

The experiments with $K = 1$ indicate that there are substantial gains to averaging the bias-corrections, at least for small to moderate sample sizes. For the test of equivalent means, with $n_1 = n_2 = 50$, and $\lambda_2 = 2.00$ (so that the null is true), the experiments with $K = 1$ yield realized test-sizes at the five-percent level of 0.146, 0.227, 0.352, 0.103, and 0.110 for 2–6 dimensions. For comparison, the corresponding realized sizes in Table 1 are 0.117, 0.190, 0.283, 0.090, and 0.102. The differences in corresponding realized sizes become smaller with $n_1 = n_2 = 100$ and $n_1 = n_2 = 200$, and are insignificantly different when $n_1 = n_2 = 1000$ except in the case $p = q = 2$, where the realized size is 0.134, compared to 0.096 in Table 1. Similar differences can be seen when the results for the returns to scale and convexity tests with $K = 1$ are compared to results with $K = 100$. We suggest averaging the bias corrections when using fewer than 1,000 observations, but there is little or no apparent gain from doing so when with sample sizes greater than 1,000.

5 Summary and Conclusions

We have presented tests of equivalent means across groups of producers, constant versus variable returns to scale of the frontier Ψ^{∂}, and convexity versus non-convexity of the production set Ψ based on the new central limit theorem results of Kneip et al. (2013). Our tests rely on asymptotic normality of the test statistics, and thus avoid the complication and computational burden of bootstrapping. The Monte Carlo results we presented in Section 4 indicate that performance of the tests, in terms of realized sizes, improves as sample size increases.

The Monte Carlo results also indicate that our tests tend to over-reject, particularly in samples of less than a few hundred observations. As noted above in Section 4.2, the experimental results provide some practical guidance for applied researchers; i.e., one should be cautious in drawing conclusions when one of our tests just barely rejects the null. On the other hand, one can be more confident when estimated p-values are 0.01 or less. To give an example of how the tests might be used, Apon et al. (2013) examine research output by eight different academic departments across U.S. universities, and whether those that have on-campus access to high performance computing (HPC) facilities are more efficient than those that do not have access to HPC. Apon et al. find arguably clear evidence of significantly greater efficiency for departments with on-campus access to HPC in six cases, with p-values ranging from about 10^{-7} to 10^{-137}. In the two cases where the null hypothesis of equivalent means could not be rejected, the p-values were 0.9999 or greater. At least in the study by Apon et al., there seems to be little ambiguity about whether to reject null hypotheses of equivalent means; i.e., while the Monte Carlo results presented above in Section 4.2 indicate that our tests tend to over-reject, the p-values obtained by Apon et al. leave little doubt. Not every research project will yield such clear-cut results, but some will.

Appendix A: Technical Details

A.1: Additional Assumptions

The next two assumptions are required for each central limit theorem result (for means of FDH, VRS-DEA, and CRS-DEA estimates) established by Kneip et al. (2013).

Assumption A.1. *(i) The random variables (X, Y) possess a joint density f with support $\mathcal{D} \subset \Psi$; and (ii) f is continuously differentiable on $\mathcal{D}$.*

Assumption A.2. *(i) $\mathcal{D}^* := \{\theta(x, y)x, y) \mid (x, y) \in \mathcal{D}\} \subset \mathcal{D}$; (ii) $\mathcal{D}^*$ is compact; and (iii) $f(\theta(x, y)x, y) > 0$ for all $(x, y) \in \mathcal{D}$.*

In the case of FDH estimators, the central limit theorem results in Kneip et al. (2013) require the next assumption.

Assumption A.3. *(i) $\theta(x, y)$ is twice continuously differentiable on $\mathcal{D}$; and (ii) all the first-order partial derivatives of $\theta(x, y)$ with respect to x and y are nonzero at any point $(x, y) \in \mathcal{D}$.*

Recalling that the free disposability assumed in Assumption 2.2 implies that the frontier is weakly monotone, Assumption A.3 strengthens this by requiring the frontier to be strictly monotone with no constant segments.

Stronger assumptions are required by the DEA estimators. Both the VRS-DEA and CRS-DEA estimators require the next assumption.

Assumption A.4. *$\theta(x, y)$ is three times continuously differentiable on $\mathcal{D}$.*

In addition, the VRS-DEA estimator requires the following assumption.

Assumption A.5. *$\mathcal{D}$ is* almost strictly convex*; i.e., for any $(x, y), (\widetilde{x}, \widetilde{y}) \in \mathcal{D}$ with $(\frac{x}{\|x\|}, y) \neq (\frac{\widetilde{x}}{\|\widetilde{x}\|}, \widetilde{y})$, the set $\{(x^*, y^*) \mid (x^*, y^*) = (x, y) + \alpha((\widetilde{x}, \widetilde{y}) - (x, y))$ for some $0 < \alpha < 1\}$ is a subset of the* interior *of $\mathcal{D}$.*

For the case of the CRS-DEA estimator, Assumption A.5 must be replaced by the following condition.

Assumption A.6. *(i) For any $(x, y) \in \Psi$ and any $a \in [0, \infty)$, $(ax, ay) \in \Psi$; (ii) the support $\mathcal{D} \subset \Psi$ of f is such that for any (x, y), $(\widetilde{x}, \widetilde{y}) \in \mathcal{D}$ with $(\frac{x}{\|x\|}, \frac{y}{\|y\|}) \neq (\frac{\widetilde{x}}{\|\widetilde{x}\|}, \frac{\widetilde{y}}{\|\widetilde{y}\|})$, the set*

$\{(x^*, y^*) \mid (x^*, y^*) = (x, y) + \alpha((\widetilde{x}, \widetilde{y}) - (x, y))$ *for some* $0 < \alpha < 1\}$ *is a subset of the* interior *of* $\mathcal{D}$*; (iii)* $\mathcal{D}$ *is a connected set; and (iv)* $(x, y) \notin \mathcal{D}$ *for any* $(x, y) \in \mathbb{R}^p_+ \times \mathbb{R}^q$ *with* $y^1 = 0$*, where* y^1 *denotes the first element of the vector* y*.*

To summarize, all of the central limit theorem results obtained by Kneip et al. (2013) and used in Section 3 depend on Assumptions 2.1–A.2. In addition to these assumptions, the results involving FDH estimators require Assumption A.3. The VRS-DEA and CRS-DEA estimators require the stronger Assumption A.4 in place of Assumption A.3. The VRS-DEA estimator also requires Assumption A.5, while the VRS-CRS estimator instead requires Assumption A.6. It is important to note that the conditions on the structure of Ψ (and $\mathcal{D}$) given in Assumptions A.5 and A.6 are incompatible. It is not possible that both assumptions hold simultaneously.

A.2: Proof of Theorem 3.1

The construction follows the arguments used in the proofs of Theorems 3.1 and 3.2 in Kneip et al. (2013). We therefore need some additional notation. Consider the transformations $x^* = x/y^1$, $y^* = y/y^1 = (1, y^2/y^1, \ldots, y^q/y^1)'$ and $\widetilde{y} = (y^2/y^1, \ldots, y^q/y^1)' \in \mathbb{R}^{q-1}$, for all $y = (y^1, \ \ldots, \ y^q)'$ with $y^1 > 0$ With respect to the $(q-1)$-dimensional output variable $\widetilde{y}$, a production set $\widetilde{\Psi} := \{(x^*, \widetilde{y}) \mid (x^*, (1, \widetilde{y})') \in \Psi$ can be defined. By the CRS-assumption, corresponding efficiencies are given by $\theta^*(x^*, \widetilde{y}) := \theta(x^*, (1, \widetilde{y})') = \theta(x, y)$. Furthermore, the density f of (X_i, Y_i) induces a density f^* of $(X_i^*, \widetilde{Y}_i)$. Smoothness of θ and f translates into a corresponding smoothness of θ^* and f^*.

Consider a point (x, y) in the interior of $\mathcal{D}$, and let $\mathcal{V}(x^*)$ denote the $(p-1)$-dimensional linear space of all vectors $z \in \mathbb{R}^p$ such that $z^T x^* = 0$, and let $\Psi^*(x^*)$ denote the set of all $(z, \widetilde{y}) \in \mathcal{V}(x^*) \times \mathbb{R}^{q-1}$ with $(\gamma \frac{x^*}{||x^*||} + z, \widetilde{y}) \in \widetilde{\mathcal{D}}$ for some $\gamma > 0$. This introduces another coordinate system whose properties are extensively discussed in Kneip et al. (2008) and Kneip et al. (2013). In particular, the efficient boundary of Ψ^* can now be described by the function $g_{x^*}(z, \widetilde{y}) := \inf \left\{ \gamma \mid \left(\gamma \frac{x^*}{||x^*||} + z, \widetilde{y} \right) \in \Psi^* \right\}$.

Let $\widetilde{\mathcal{X}}_n := \{(X_i^*, \widetilde{Y}_i), \ i = 1, \ldots, n\}$. Since (x, y) is in the interior of $\mathcal{D}$, the probability that $(x^*, \widetilde{y})$ is in the convex hull of $\mathcal{X}_n^*$ tends to 1 as $n \to \infty$. But then it has been shown in Kneip et al. (2013) that the CRS-estimator of $\theta(x, y)$ exactly coincides with a VRS-DEA estimator based on the reduced sample of observations $\widetilde{\mathcal{X}}_n$:

$$\begin{aligned}\widehat{\theta}_{\text{CRS}}(x, y \mid \mathcal{X}_n) &= \widehat{\theta}_{\text{VRS}}(x^*, \widetilde{y} \mid \widetilde{\mathcal{X}}_n) \\ &= \min_{\boldsymbol{\omega}} \Big\{ \sum_{i=1}^{n} \omega_i \frac{g_{x^*}(\theta^*(X_i^*, \widetilde{Y}_i) Z_i, \widetilde{Y}_i)}{\|x^*\| \theta^*(X_i^*, \widetilde{Y}_i)} \mid \boldsymbol{Z}\boldsymbol{\omega} = 0,\ \widetilde{\boldsymbol{Y}}\boldsymbol{\omega} = \widetilde{y},\ \boldsymbol{i}_n'\boldsymbol{\omega} = 1,\ \boldsymbol{\omega} \in \mathbb{R}_+^n \Big\},\end{aligned} \tag{A.1}$$

where $\boldsymbol{i}_n$ is defined as in Section 2, ω_i represents the ith element of $\boldsymbol{\omega}$, $\theta_i^* = \theta(X_i^*, \widetilde{Y}_i)$, $Z_i = X_i^* - \dfrac{\boldsymbol{x}^{*T} X_i^*}{||\boldsymbol{x}^*||^2}\boldsymbol{x}^*$ is a $(p \times 1)$ vector, while $\boldsymbol{Y} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_n)$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ are $((q-1) \times n)$ and $(p \times n)$ matrices, respectively.

Kneip et al. (2013) also show that when including the point $(0, 0)$, then $\widehat{\theta}_{\text{VRS}}(x, y \mid \mathcal{X}_n \cup \{(0, 0)\})$ is obtained by minimizing (A.1) with respect to the additional constraint $\sum_{i=1}^n \omega_i^* \frac{y^1}{Y_i^1} \leq 1$. When turning to the usual VRS-DEA estimator $\widehat{\theta}_{\text{VRS}}(x, y \mid \mathcal{X}_n)$, then the same type of arguments yield

$$\begin{aligned}&\widehat{\theta}_{\text{VRS}}(x, y \mid \mathcal{X}_n) \\ &= \min_{\boldsymbol{\omega}} \Big\{ \sum_{i=1}^{n} \omega_i \frac{g_{x^*}(\theta^*(X_i^*, \widetilde{Y}_i) Z_i, \widetilde{Y}_i)}{\|x^*\| \theta^*(X_i^*, \widetilde{Y}_i)} \mid \boldsymbol{Z}\boldsymbol{\omega} = 0,\ \widetilde{\boldsymbol{Y}}\boldsymbol{\omega} = \widetilde{y},\ \boldsymbol{i}_n'\boldsymbol{\omega} = 1,\ \sum_{i=1}^{n} \omega_i \frac{y^1}{Y_i^1} = 1,\ \boldsymbol{\omega} \in \mathbb{R}_+^n \Big\}.\end{aligned} \tag{A.2}$$

Now define

$$\begin{aligned}\widehat{\theta}^{+}_{\text{CRS}}(x, y \mid \mathcal{X}_n) = \min_{\boldsymbol{\omega}} \Big\{ &\sum_{i: Y_i^1 > y^1} \omega_i \frac{g_{x^*}(\theta^*(X_i^*, \widetilde{Y}_i) Z_i, \widetilde{Y}_i)}{\|x^*\| \theta^*(X_i^*, \widetilde{Y}_i)} \mid \\ &\sum_{i: Y_i^1 > y^1} \omega_i = 1, \sum_{i: Y_i^1 > y^1} \omega_i Z_i^* = 0, \sum_{i: Y_i^1 > y^1} \omega_i \widetilde{Y}_i = \widetilde{y},\ \boldsymbol{\omega} \in \mathbb{R}_+^n \Big\},\end{aligned} \tag{A.3}$$

and similarly define $\widehat{\theta}^{-}_{\text{CRS}}(x, y \mid \mathcal{X}_n)$ by only using observations with $Y_i^1 < y^1$.

Let $\boldsymbol{\omega}^+$ and $\boldsymbol{\omega}^-$ denote the vectors $\boldsymbol{\omega} \in \mathbb{R}_+^n$ providing the minimal values of $\widehat{\theta}^{+}_{\text{CRS}}(x, y \mid \mathcal{X}_n)$ and $\widehat{\theta}^{-}_{\text{CRS}}(x, y \mid \mathcal{X}_n)$, respectively. Without restriction, $\omega_i^+ = 0$ whenever $Y_i^1 \leq y^1$, and $\omega_i^- = 0$ whenever $Y_i^1 \geq y^1$. Obviously, $s_+ := \sum_{i=1}^n \omega_i^+ \frac{y^1}{Y_i^1} < 1$ while $s_- := \sum_{i=1}^n \omega_i^- \frac{y^1}{Y_i^1} > 1$. Hence there exists an $0 < \alpha < 1$ such that $\alpha s_+ + (1 - \alpha)s_- = 1$, and thus $\sum_{i=1}^n \omega_i^* \frac{y^1}{Y_i^1} = 1$ for $\boldsymbol{\omega}^* := \alpha\boldsymbol{\omega}^+ + (1 - \alpha)\boldsymbol{\omega}^-$. Moreover, the vector $\boldsymbol{\omega}^*$ satisfies all constraints in (A.2). We can conclude that

$$\widehat{\theta}_{\text{CRS}}(x, y \mid \mathcal{X}_n) \leq \widehat{\theta}_{\text{VRS}}(x, y \mid \mathcal{X}_n) \leq \alpha\widehat{\theta}^{+}_{\text{CRS}}(x, y \mid \mathcal{X}_n) + (1 - \alpha)\widehat{\theta}^{-}_{\text{CRS}}(x, y \mid \mathcal{X}_n). \tag{A.4}$$

But since the point (x,y) is in the interior of $\mathcal{D}$, $\widehat{\theta}^{+}_{\text{CRS}}(x,y \mid \mathcal{X}_n)$ and $\widehat{\theta}^{-}_{\text{CRS}}(x,y \mid \mathcal{X}_n)$ are CRS-estimators based on subsamples of observations, where the size of each subsample increases proportional to n. This implies $\widehat{\theta}^{+}_{\text{CRS}}(x,y \mid \mathcal{X}_n) - \theta(x,y) = O_P(n^{-\frac{2}{p+q}})$ as well as $\widehat{\theta}^{-}_{\text{CRS}}(x,y \mid \mathcal{X}_n) - \theta(x,y) = O_P(n^{-\frac{2}{p+q}})$, and assertion (i) is an immediate consequence of (A.4).

Consider (ii). With $\nu_n := b(\frac{\log n}{n})^{\frac{1}{p+q}}$ for some $b > 0$ let $C(x,y;\nu_n^2,\nu_n)$ denote the set of all (x',y') with $1-\theta^*(x'^*,\widetilde{y}') \geq \nu_n^2$, $|z'^j| \leq \nu_n$, $j = 1,\ldots,p-1$, and $|\widetilde{y}'^j - \widetilde{y}^j| \leq \nu_n$, $j = 1,\ldots,q-1$. If $\boldsymbol{\omega}^{opt}$ denotes the vector $\boldsymbol{\omega} \in \mathbb{R}^n_+$ providing the minimal value of $\widehat{\theta}_{\text{VRS}}(x,y \mid \mathcal{X}_n)$ in (A.2), then a straightforward generalization of the localization arguments given in Kneip et al. (2008) and Kneip et al. (2013) shows that with probability tending to 1,

$$\omega_i^{opt} = 0 \text{ for all } i = 1,\ldots,n \text{ with } (X_i,Y_i) \notin C(x,y;\nu_n^2,\nu_n). \tag{A.5}$$

Let $\bar{f}_1$ denote the marginal density of Y_i^1. Assumptions A.2 and A.6 imply that $\bar{f}_1$ has a compact support $[y^1_{min}, y^1_{max}] \subset \mathbb{R}_+$ with $y^1_{min} > 0$. Moreover, $\bar{f}_1$ is continuous, and $\bar{f}_1(y_1) > 0$ for any $y^1 \in [y^1_{min}, y^1_{max}]$. For $y^1_{min} < y^1 < y^1_{max}$ let $\pi^+(y^1) = P(Y_i^1 > y^1)$ as well as $\pi^-(y^1) = P(Y_i^1 < y^1)$. Obviously, the number of observations (X_i,Y_i) with $Y_i^1 > y^1$ varies around $n\pi^+(y^1)$, while the number of observations (X_i,Y_i) with $Y_i^1 < y^1$ varies around $n\pi^-(y^1)$.

Using (A.5), a straightforward generalization of the arguments in the proof of Theorem 3.1 of Kneip et al. (2013) now may be used to show that for any $y^1_{min} < y^1 < y^1_{max}$,

$$\begin{aligned}\left| E\left(\widehat{\theta}_{\text{VRS}}(X_i,Y_i \mid \mathcal{X}_n) - \theta(X_i,Y_i) \middle| Y_i^1 = y^1\right) - n^{-\frac{2}{p+q}} D(y^1)\right| \\ \leq A\left(\frac{1}{n\min\{\pi^+(y^1),\pi^-(y^1)\}}\right)^{\frac{3}{p+q}} (\log n)^{\frac{p+q+3}{p+q}},\end{aligned} \tag{A.6}$$

where $0 < D(y^1) < \infty$ is a measurable function of y^1, while $0 < A < \infty$ is a constant which does not depend on y^1. The exact analytical structure of $D(y^1)$ is difficult to evaluate, but Theorem 3.2 of Kneip et al. (2013) together with (A.4) and our maintained distributional assumptions imply that there are constants $0 < A_1 < \infty$ and $0 < A_2 < \infty$ such that

$$D(y^1) \leq A_1\left(\frac{1}{\min\{\pi^+(y^1),\pi^-(y^1)\}}\right)^{\frac{2}{p+q}} \leq A_2\left(\frac{1}{\min\{y^1 - y^1_{min}, y^1_{max} - y^1\}}\right)^{\frac{2}{p+q}}. \tag{A.7}$$

If $p+q > 2$, then the integral $\int_{y^1_{min}}^{y^1_{max}} \left(\frac{1}{\min\{y^1-y^1_{min},y^1_{max}-y^1\}}\right)^{\frac{2}{p+q}} \bar{f}_1(y^1)dy^1$ is necessarily finite. Since $E\left(\widehat{\theta}_{\text{VRS}}(X_i,Y_i \mid \mathcal{X}_n) - \theta(X_i,Y_i)\right) = \int_{y^1_{min}}^{y^1_{max}} E\left(\widehat{\theta}_{\text{VRS}}(X_i,Y_i \mid \mathcal{X}_n) - \theta(X_i,Y_i) \middle| Y_i^1 = y^1\right) \bar{f}_1(y^1)dy^1$, relation (3.21) then follows from (A.6) and (A.7).

Note that for $p+q=2$,

$$\int_{y^1_{min}+n^{-1}}^{y^1_{max}-n^{-1}} \left(\frac{1}{\min\{y^1 - y^1_{min}, y^1_{max} - y^1\}} \right) \bar{f}_1(y^1) dy^1 = O(\log n). \tag{A.8}$$

Since furthermore $0 \leq \widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i) \leq 1$ for all i, and $P(Y_i^1 \in [y^1_{min}, y^1_{min} + n^{-1}]) = O(n^{-1})$, $P(Y_i^1 \in [y^1_{max} - n^{-1}, y^1_{max}]) = O(n^{-1})$, (A.6) and (A.7) yield $E\left(\widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) = O(n^{-1}\log n)$ for $p+q=2$.

Using the localization result (A.5), assertions (3.22) and (3.23) follow from straightforward generalizations of the arguments used in the proofs of Theorems 3.1 and 3.2 of Kneip et al. (2013) in order to derive variances and covariances of VRS- and CRS-estimators. ■

References

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

Apon, A. W., L. B. Ngo, M. E. Payne, and P. W. Wilson (2013), Assessing the effect of high performance computing capabilities on academic research output. Unpublished working paper, Department of Economics and School of Computing, Clemson University, Clemson, SC 29634 USA.

Banker, R. D., A. Charnes, and W. W. Cooper (1984), Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science* 30, 1078–1092.

Charnes, A., W. W. Cooper, and E. Rhodes (1978), Measuring the efficiency of decision making units, *European Journal of Operational Research* 2, 429–444.

— (1981), Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through, *Management Science* 27, 668–697.

Daouia, A., L. Simar, and P. W. Wilson (2013), Measuring firm performance using nonparametric quantile-type distances. Discussion paper, Institut de Statistique Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

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

Deprins, D., L. Simar, and H. Tulkens (1984), Measuring labor inefficiency in post offices, in M. M. P. Pestieau and H. Tulkens, eds., *The Performance of Public Enterprises: Concepts and Measurements*, Amsterdam: North-Holland, pp. 243–267.

Färe, R. (1988), *Fundamentals of Production Theory*, Berlin: Springer-Verlag.

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

Gattoufi, S., M. Oral, and A. Reisman (2004), Data envelopment analysis literature: A bibliography update (1951–2001), *Socio-Economic Planning Sciences* 38, 159–229.

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.

Jeong, S. O. (2004), Asymptotic distribution of DEA efficiency scores, *Journal of the Korean Statistical Society* 33, 449–458.

Jeong, S. O. and B. U. Park (2006), Large sample approximation of the distribution for convex-hull estimators of boundaries, *Scandinavian Journal of Statistics* 33, 139–151.

Kantorovich, L. V. (1939), *Mathematicheskiye Metody Organizatsiyi i Planirovaniya Proizvodstva*, Leningrad: Leningrad State University Press.

Kneip, A., B. Park, and L. Simar (1998), A note on the convergence of nonparametric DEA efficiency measures, *Econometric Theory* 14, 783–793.

Kneip, A., L. Simar, and P. W. Wilson (2008), Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models, *Econometric Theory* 24, 1663–1697.

— (2011), A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators, *Computational Economics* Forthcoming.

— (2013), Central limit theorems for DEA scores: When bias can kill the variance. Discussion paper #2013/12, Institut de Statistique Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Koopmans, T. C. (1951), An analysis of production as an efficient combination of activities, in T. C. Koopmans, ed., *Activity Analysis of Production and Allocation*, New York: John-Wiley and Sons, Inc., pp. 33–97. Cowles Commission for Research in Economics, Monograph 13.

Korostelev, A., L. Simar, and A. B. Tsybakov (1995a), Efficient estimation of monotone boundaries, *The Annals of Statistics* 23, 476–489.

— (1995b), On estimation of monotone and convex boundaries, *Publications de l'Institut de Statistique de l'Université de Paris XXXIX* 1, 3–18.

Park, B. U., S.-O. Jeong, and L. Simar (2010), Asymptotic distribution of conical-hull estimators of directional edges, *Annals of Statistics* 38, 1320–1340.

Park, B. U., L. Simar, and C. Weiner (2000), FDH efficiency scores from a stochastic point of view, *Econometric Theory* 16, 855–877.

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

Simar, L. and A. Vanhems (2012), Probabilistic characterization of directional distances and their robust versions, *Journal of Econometrics* 166, 342–354.

Simar, L., A. Vanhems, and P. W. Wilson (2012), Statistical inference for dea estimators of directional distances, *European Journal of Operational Research* 220, 853–864.

Simar, L. and P. W. Wilson (2000), Statistical inference in nonparametric frontier models: The state of the art, *Journal of Productivity Analysis* 13, 49–78.

— (2013), Estimation and inference in nonparametric frontier models: Recent developments and perspectives, *Foundations and Trends in Econometrics* 5, 183–337.

Wilson, P. W. (2011), Asymptotic properties of some non-parametric hyperbolic efficiency estimators, in I. Van Keilegom and P. W. Wilson, eds., *Exploring Research Frontiers in Contemporary Statistics and Econometrics*, Berlin: Springer-Verlag, pp. 115–150.

Table 1: Rejection Rates for Test of Equivalent Means

$n_1 = n_2$	λ_2	$p=1,\ q=1$			$p=2,\ q=1$			$p=2,\ q=2$			$p=3,\ q=2$			$p=3,\ q=3$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
50	2.00	0.181	0.117	0.036	0.266	0.190	0.075	0.374	0.283	0.169	0.130	0.090	0.035	0.155	0.102	0.038
	1.90	0.191	0.118	0.038	0.283	0.198	0.089	0.394	0.295	0.171	0.139	0.088	0.036	0.152	0.102	0.038
	1.80	0.229	0.148	0.055	0.308	0.219	0.101	0.410	0.334	0.197	0.138	0.086	0.037	0.147	0.098	0.042
	1.70	0.274	0.192	0.083	0.349	0.270	0.142	0.460	0.365	0.230	0.141	0.087	0.034	0.158	0.103	0.042
	1.60	0.354	0.253	0.129	0.435	0.325	0.186	0.511	0.430	0.282	0.145	0.100	0.041	0.167	0.113	0.054
	1.50	0.442	0.343	0.187	0.521	0.422	0.256	0.569	0.491	0.348	0.163	0.110	0.044	0.176	0.123	0.060
	1.40	0.553	0.445	0.255	0.624	0.529	0.332	0.629	0.564	0.434	0.174	0.115	0.046	0.199	0.131	0.066
	1.30	0.669	0.570	0.362	0.727	0.645	0.464	0.711	0.644	0.513	0.188	0.134	0.060	0.218	0.138	0.071
	1.20	0.781	0.678	0.483	0.816	0.744	0.581	0.790	0.732	0.599	0.198	0.145	0.064	0.231	0.171	0.079
	1.10	0.881	0.806	0.615	0.887	0.848	0.708	0.867	0.821	0.704	0.227	0.155	0.072	0.259	0.191	0.093
	1.00	0.936	0.899	0.754	0.932	0.906	0.815	0.926	0.896	0.804	0.258	0.181	0.085	0.291	0.212	0.110
	0.75	0.995	0.991	0.959	0.998	0.994	0.966	0.991	0.983	0.962	0.365	0.262	0.140	0.388	0.294	0.166
	0.50	1.000	1.000	0.997	1.000	1.000	1.000	0.999	0.999	0.998	0.521	0.404	0.228	0.545	0.431	0.255
100	2.00	0.149	0.083	0.018	0.223	0.143	0.061	0.270	0.199	0.095	0.101	0.054	0.019	0.128	0.091	0.044
	1.90	0.171	0.106	0.030	0.238	0.164	0.068	0.309	0.214	0.110	0.096	0.050	0.017	0.133	0.090	0.046
	1.80	0.226	0.147	0.051	0.296	0.214	0.109	0.363	0.288	0.154	0.112	0.060	0.018	0.134	0.095	0.050
	1.70	0.336	0.221	0.097	0.395	0.312	0.162	0.464	0.371	0.231	0.116	0.061	0.020	0.145	0.097	0.053
	1.60	0.469	0.373	0.181	0.528	0.429	0.252	0.581	0.492	0.331	0.136	0.075	0.023	0.155	0.103	0.052
	1.50	0.606	0.511	0.318	0.666	0.582	0.382	0.705	0.632	0.468	0.153	0.087	0.026	0.168	0.117	0.056
	1.40	0.765	0.658	0.470	0.802	0.723	0.543	0.836	0.767	0.616	0.169	0.112	0.030	0.179	0.127	0.064
	1.30	0.895	0.825	0.621	0.895	0.843	0.708	0.915	0.879	0.765	0.198	0.123	0.042	0.197	0.140	0.071
	1.20	0.961	0.929	0.798	0.957	0.926	0.831	0.972	0.943	0.880	0.219	0.147	0.052	0.215	0.162	0.082
	1.10	0.990	0.978	0.924	0.984	0.979	0.928	0.986	0.984	0.950	0.253	0.182	0.071	0.239	0.179	0.091
	1.00	0.999	0.995	0.980	1.000	0.994	0.981	0.996	0.994	0.984	0.302	0.209	0.092	0.259	0.205	0.111
	0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.421	0.324	0.159	0.349	0.262	0.159
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.557	0.434	0.234	0.492	0.373	0.220

Table 1: Rejection Rates for Test of Equivalent Means (continued)

$n_1 = n_2$	λ_2	$p=1,\ q=1$			$p=2,\ q=1$			$p=2,\ q=2$			$p=3,\ q=2$			$p=3,\ q=3$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
200	2.00	0.144	0.086	0.015	0.185	0.113	0.038	0.236	0.172	0.061	0.108	0.062	0.024	0.123	0.066	0.023
	1.90	0.193	0.118	0.044	0.215	0.150	0.065	0.301	0.204	0.091	0.109	0.062	0.024	0.121	0.069	0.024
	1.80	0.324	0.223	0.103	0.376	0.271	0.128	0.419	0.329	0.174	0.112	0.066	0.025	0.126	0.075	0.022
	1.70	0.522	0.400	0.210	0.552	0.463	0.254	0.597	0.500	0.330	0.120	0.069	0.029	0.138	0.085	0.022
	1.60	0.738	0.623	0.397	0.743	0.651	0.465	0.757	0.679	0.524	0.133	0.075	0.032	0.147	0.086	0.025
	1.50	0.894	0.833	0.637	0.881	0.825	0.664	0.887	0.832	0.713	0.154	0.089	0.036	0.161	0.101	0.031
	1.40	0.967	0.944	0.844	0.966	0.939	0.846	0.967	0.938	0.862	0.178	0.108	0.043	0.173	0.119	0.040
	1.30	0.996	0.988	0.954	0.989	0.984	0.953	0.994	0.989	0.962	0.197	0.132	0.055	0.190	0.131	0.055
	1.20	0.999	0.999	0.993	0.999	0.998	0.986	1.000	0.999	0.994	0.238	0.156	0.063	0.220	0.147	0.065
	1.10	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000	1.000	0.279	0.191	0.083	0.255	0.172	0.083
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.317	0.225	0.105	0.303	0.205	0.099
	0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.474	0.351	0.182	0.446	0.332	0.157
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.664	0.521	0.318	0.607	0.488	0.271
1000	2.00	0.124	0.071	0.026	0.146	0.084	0.020	0.151	0.096	0.031	0.105	0.055	0.012	0.087	0.044	0.017
	1.90	0.327	0.227	0.100	0.359	0.255	0.113	0.379	0.273	0.132	0.108	0.057	0.015	0.087	0.049	0.016
	1.80	0.755	0.639	0.408	0.773	0.675	0.471	0.798	0.706	0.503	0.116	0.064	0.022	0.097	0.053	0.015
	1.70	0.971	0.943	0.854	0.974	0.947	0.853	0.980	0.966	0.888	0.140	0.077	0.024	0.121	0.064	0.017
	1.60	1.000	1.000	0.987	0.999	0.999	0.988	0.998	0.997	0.993	0.160	0.097	0.033	0.148	0.075	0.025
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.185	0.125	0.040	0.172	0.105	0.031
	1.40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.235	0.149	0.055	0.192	0.138	0.039
	1.30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.279	0.186	0.081	0.229	0.162	0.062
	1.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.334	0.236	0.111	0.281	0.192	0.081
	1.10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.408	0.296	0.137	0.338	0.236	0.111
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.474	0.357	0.173	0.400	0.290	0.143
	0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.680	0.550	0.329	0.582	0.454	0.237
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.886	0.790	0.553	0.794	0.659	0.406

Table 1: Rejection Rates for Test of Equivalent Means (continued)

$n_1 = n_2$	λ_2	$p=1,\ q=1$			$p=2,\ q=1$			$p=2,\ q=2$			$p=3,\ q=2$			$p=3,\ q=3$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
10000	2.00	0.101	0.045	0.006	0.090	0.047	0.007	0.128	0.054	0.017	0.115	0.054	0.015	0.112	0.060	0.008
	1.90	0.973	0.950	0.856	0.982	0.959	0.861	0.977	0.958	0.846	0.117	0.057	0.018	0.117	0.068	0.011
	1.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.130	0.077	0.020	0.118	0.076	0.017
	1.70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.154	0.095	0.029	0.137	0.080	0.022
	1.60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.205	0.126	0.044	0.152	0.095	0.037
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.272	0.171	0.062	0.186	0.126	0.053
	1.40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.350	0.242	0.094	0.235	0.151	0.070
	1.30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.424	0.318	0.138	0.320	0.196	0.091
	1.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.504	0.391	0.198	0.367	0.256	0.121
	1.10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.597	0.478	0.273	0.440	0.336	0.151
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.693	0.571	0.356	0.522	0.404	0.194
	0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.893	0.822	0.597	0.747	0.636	0.378
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.988	0.958	0.870	0.921	0.847	0.626
20000	2.00	0.120	0.067	0.011	0.113	0.061	0.017	0.091	0.049	0.014	0.101	0.058	0.009	0.101	0.055	0.010
	1.90	1.000	0.998	0.995	1.000	1.000	0.992	1.000	1.000	0.995	0.103	0.059	0.013	0.112	0.057	0.015
	1.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.128	0.065	0.018	0.120	0.070	0.018
	1.70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.164	0.094	0.030	0.150	0.085	0.020
	1.60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.219	0.133	0.044	0.178	0.110	0.034
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.277	0.199	0.064	0.216	0.146	0.054
	1.40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.367	0.255	0.103	0.267	0.180	0.076
	1.30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.464	0.338	0.164	0.344	0.229	0.096
	1.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.578	0.436	0.232	0.409	0.311	0.143
	1.10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.684	0.569	0.311	0.489	0.380	0.191
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.780	0.679	0.417	0.588	0.460	0.245
	0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.949	0.901	0.729	0.817	0.712	0.449
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.992	0.990	0.937	0.962	0.902	0.732

Table 2: Rejection Rates for Returns to Scale Test

n	δ	$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
50	0.0	0.180	0.109	0.035	0.267	0.182	0.077	0.327	0.258	0.139	0.329	0.252	0.146	0.224	0.146	0.061
	0.1	0.223	0.132	0.048	0.287	0.195	0.083	0.336	0.247	0.142	0.333	0.263	0.146	0.224	0.150	0.063
	0.2	0.373	0.255	0.091	0.374	0.271	0.124	0.403	0.312	0.178	0.378	0.301	0.175	0.250	0.166	0.074
	0.3	0.620	0.456	0.187	0.560	0.436	0.222	0.555	0.447	0.276	0.517	0.438	0.283	0.334	0.246	0.120
	0.4	0.717	0.503	0.184	0.703	0.516	0.236	0.718	0.566	0.307	0.690	0.565	0.322	0.423	0.333	0.157
	0.6	0.814	0.611	0.262	0.806	0.625	0.285	0.840	0.663	0.372	0.844	0.681	0.381	0.474	0.385	0.182
	0.8	0.869	0.689	0.316	0.854	0.676	0.326	0.890	0.719	0.411	0.889	0.741	0.431	0.490	0.406	0.185
	1.0	0.896	0.730	0.364	0.892	0.715	0.367	0.911	0.756	0.434	0.916	0.774	0.456	0.501	0.421	0.190
	1.2	0.910	0.753	0.406	0.900	0.733	0.376	0.927	0.781	0.463	0.934	0.810	0.486	0.515	0.432	0.197
	1.4	0.920	0.772	0.445	0.919	0.759	0.416	0.941	0.807	0.483	0.946	0.824	0.519	0.527	0.443	0.207
100	0.0	0.173	0.112	0.032	0.233	0.161	0.057	0.284	0.201	0.093	0.287	0.205	0.094	0.172	0.105	0.023
	0.1	0.235	0.139	0.048	0.272	0.179	0.071	0.294	0.207	0.100	0.295	0.206	0.090	0.173	0.104	0.023
	0.2	0.429	0.302	0.132	0.416	0.294	0.139	0.380	0.277	0.151	0.350	0.262	0.134	0.207	0.125	0.033
	0.3	0.749	0.617	0.346	0.689	0.576	0.329	0.646	0.523	0.300	0.579	0.474	0.287	0.314	0.213	0.073
	0.4	0.856	0.677	0.323	0.866	0.699	0.366	0.844	0.668	0.359	0.823	0.664	0.380	0.397	0.288	0.121
	0.6	0.912	0.761	0.441	0.910	0.791	0.473	0.921	0.786	0.460	0.931	0.781	0.454	0.440	0.331	0.158
	0.8	0.924	0.798	0.487	0.928	0.820	0.521	0.938	0.800	0.488	0.947	0.818	0.497	0.446	0.355	0.169
	1.0	0.931	0.809	0.533	0.934	0.827	0.550	0.944	0.826	0.510	0.957	0.827	0.532	0.465	0.362	0.170
	1.2	0.935	0.816	0.550	0.934	0.842	0.572	0.943	0.834	0.530	0.963	0.839	0.549	0.479	0.382	0.177
	1.4	0.941	0.823	0.564	0.937	0.847	0.598	0.944	0.835	0.539	0.964	0.844	0.555	0.483	0.393	0.189

Table 2: Rejection Rates for Returns to Scale Test (continued)

n	δ	$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
200	0.0	0.142	0.076	0.021	0.216	0.141	0.038	0.237	0.160	0.060	0.273	0.191	0.086	0.148	0.085	0.024
	0.1	0.208	0.130	0.037	0.254	0.178	0.054	0.271	0.183	0.068	0.280	0.196	0.090	0.153	0.081	0.023
	0.2	0.526	0.388	0.168	0.462	0.334	0.173	0.419	0.307	0.145	0.384	0.282	0.143	0.188	0.111	0.031
	0.3	0.934	0.860	0.613	0.853	0.738	0.494	0.759	0.658	0.450	0.740	0.642	0.417	0.339	0.216	0.081
	0.4	0.933	0.823	0.537	0.929	0.812	0.533	0.942	0.834	0.541	0.940	0.826	0.547	0.405	0.308	0.160
	0.6	0.943	0.842	0.593	0.942	0.847	0.613	0.954	0.862	0.624	0.964	0.862	0.632	0.426	0.331	0.184
	0.8	0.946	0.850	0.631	0.942	0.848	0.623	0.953	0.861	0.638	0.968	0.870	0.650	0.424	0.342	0.189
	1.0	0.948	0.854	0.631	0.942	0.845	0.625	0.948	0.860	0.638	0.963	0.871	0.648	0.425	0.346	0.193
	1.2	0.948	0.855	0.633	0.939	0.849	0.627	0.946	0.861	0.637	0.963	0.864	0.640	0.435	0.351	0.202
	1.4	0.949	0.857	0.643	0.939	0.845	0.639	0.946	0.863	0.643	0.960	0.862	0.644	0.441	0.353	0.205
1000	0.0	0.101	0.048	0.015	0.166	0.098	0.031	0.188	0.110	0.032	0.218	0.146	0.050	0.129	0.070	0.026
	0.1	0.266	0.172	0.051	0.304	0.192	0.056	0.263	0.166	0.050	0.266	0.182	0.072	0.137	0.074	0.027
	0.2	0.907	0.830	0.646	0.820	0.711	0.457	0.701	0.576	0.327	0.603	0.492	0.275	0.212	0.127	0.038
	0.3	1.000	1.000	1.000	0.998	0.998	0.990	0.999	0.999	0.975	0.995	0.989	0.948	0.540	0.403	0.178
	0.4	0.954	0.873	0.714	0.958	0.871	0.737	0.977	0.924	0.795	0.969	0.922	0.799	0.408	0.303	0.173
	0.6	0.955	0.873	0.716	0.961	0.875	0.728	0.972	0.916	0.787	0.966	0.916	0.793	0.391	0.293	0.178
	0.8	0.958	0.874	0.721	0.959	0.879	0.724	0.969	0.912	0.791	0.962	0.908	0.787	0.383	0.285	0.177
	1.0	0.959	0.876	0.725	0.959	0.882	0.724	0.966	0.909	0.781	0.959	0.905	0.779	0.373	0.280	0.180
	1.2	0.958	0.874	0.728	0.959	0.882	0.726	0.966	0.907	0.779	0.957	0.904	0.779	0.365	0.282	0.177
	1.4	0.960	0.875	0.731	0.959	0.882	0.728	0.963	0.904	0.777	0.957	0.900	0.776	0.369	0.282	0.178

Table 2: Rejection Rates for Returns to Scale Test (continued)

n	δ	$p = 1$			$p = 2$			$p = 3$			$p = 4$			$p = 5$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
10000	0.0	0.106	0.051	0.009	0.102	0.057	0.013	0.149	0.092	0.019	0.203	0.117	0.036	0.133	0.074	0.018
	0.1	0.793	0.676	0.398	0.604	0.448	0.198	0.520	0.381	0.155	0.461	0.330	0.152	0.165	0.092	0.027
	0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	0.999	0.993	0.426	0.294	0.124
	0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.965	0.926	0.781
	0.4	0.969	0.915	0.830	0.964	0.921	0.837	0.972	0.938	0.859	0.979	0.942	0.865	0.350	0.254	0.150
	0.6	0.970	0.917	0.833	0.964	0.921	0.844	0.971	0.934	0.852	0.977	0.940	0.859	0.329	0.245	0.144
	0.8	0.970	0.916	0.835	0.963	0.919	0.845	0.970	0.934	0.846	0.974	0.939	0.853	0.327	0.246	0.144
	1.0	0.970	0.919	0.837	0.965	0.920	0.845	0.968	0.930	0.852	0.974	0.936	0.855	0.327	0.242	0.138
	1.2	0.970	0.920	0.840	0.966	0.920	0.847	0.968	0.931	0.849	0.972	0.935	0.850	0.321	0.238	0.135
	1.4	0.970	0.921	0.839	0.964	0.918	0.843	0.967	0.929	0.849	0.969	0.935	0.849	0.324	0.238	0.134
20000	0.0	0.093	0.042	0.008	0.112	0.060	0.010	0.149	0.074	0.015	0.171	0.101	0.027	0.119	0.061	0.015
	0.1	0.956	0.907	0.731	0.835	0.722	0.433	0.686	0.558	0.302	0.607	0.465	0.207	0.162	0.087	0.021
	0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.536	0.410	0.176
	0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.990	0.947
	0.4	0.971	0.936	0.859	0.968	0.929	0.844	0.977	0.933	0.856	0.976	0.947	0.869	0.323	0.235	0.143
	0.6	0.972	0.935	0.858	0.971	0.937	0.845	0.973	0.934	0.851	0.977	0.945	0.871	0.306	0.235	0.140
	0.8	0.972	0.936	0.859	0.970	0.933	0.844	0.970	0.928	0.850	0.973	0.943	0.865	0.290	0.227	0.138
	1.0	0.972	0.935	0.860	0.969	0.933	0.846	0.970	0.926	0.855	0.973	0.941	0.862	0.284	0.216	0.136
	1.2	0.973	0.936	0.862	0.969	0.933	0.845	0.968	0.924	0.854	0.971	0.938	0.860	0.282	0.216	0.131
	1.4	0.973	0.936	0.862	0.969	0.937	0.847	0.968	0.924	0.852	0.971	0.938	0.860	0.283	0.213	0.129

Table 3: Rejection Rates for Convexity Test

n	δ	— $p = 1$ —			— $p = 2$ —			— $p = 3$ —			— $p = 4$ —			— $p = 5$ —		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
50	1.4*	0.134	0.086	0.023	0.138	0.101	0.045	0.086	0.049	0.017	0.099	0.058	0.027	0.123	0.081	0.032
	0.0	0.381	0.298	0.154	0.383	0.311	0.188	0.222	0.154	0.066	0.194	0.153	0.080	0.220	0.162	0.098
	0.1	0.393	0.309	0.161	0.401	0.321	0.198	0.228	0.154	0.064	0.194	0.156	0.084	0.220	0.164	0.097
	0.2	0.452	0.361	0.209	0.433	0.355	0.233	0.242	0.168	0.072	0.215	0.164	0.087	0.234	0.168	0.105
	0.3	0.556	0.473	0.313	0.493	0.410	0.279	0.265	0.181	0.086	0.240	0.180	0.102	0.259	0.190	0.114
	0.4	0.714	0.642	0.503	0.583	0.502	0.357	0.310	0.232	0.111	0.279	0.207	0.118	0.285	0.212	0.129
	0.6	0.866	0.840	0.731	0.708	0.654	0.535	0.412	0.319	0.190	0.365	0.292	0.173	0.362	0.279	0.175
	0.8	0.950	0.920	0.859	0.829	0.784	0.688	0.525	0.429	0.293	0.464	0.397	0.244	0.471	0.369	0.242
	1.0	0.974	0.962	0.926	0.881	0.853	0.781	0.627	0.542	0.397	0.568	0.483	0.338	0.582	0.487	0.328
	1.2	0.984	0.977	0.954	0.910	0.895	0.830	0.698	0.622	0.485	0.647	0.567	0.410	0.663	0.575	0.425
	1.4	0.987	0.981	0.964	0.929	0.910	0.862	0.754	0.681	0.542	0.701	0.624	0.474	0.697	0.629	0.493
100	1.4*	0.122	0.068	0.019	0.119	0.076	0.029	0.089	0.063	0.012	0.091	0.064	0.016	0.106	0.066	0.021
	0.0	0.324	0.222	0.093	0.400	0.325	0.163	0.201	0.132	0.053	0.159	0.111	0.044	0.188	0.134	0.051
	0.1	0.345	0.241	0.098	0.408	0.346	0.177	0.210	0.139	0.058	0.163	0.109	0.046	0.188	0.135	0.054
	0.2	0.424	0.327	0.159	0.453	0.378	0.223	0.220	0.152	0.063	0.178	0.118	0.050	0.207	0.141	0.063
	0.3	0.625	0.531	0.342	0.561	0.468	0.325	0.257	0.179	0.073	0.192	0.138	0.061	0.236	0.163	0.077
	0.4	0.841	0.788	0.660	0.712	0.642	0.482	0.330	0.253	0.121	0.238	0.163	0.079	0.291	0.203	0.108
	0.6	0.964	0.950	0.893	0.868	0.820	0.728	0.497	0.386	0.249	0.345	0.265	0.130	0.413	0.310	0.186
	0.8	0.992	0.985	0.972	0.952	0.934	0.876	0.657	0.563	0.394	0.511	0.404	0.233	0.565	0.453	0.294
	1.0	0.999	0.998	0.988	0.987	0.974	0.943	0.773	0.702	0.540	0.639	0.551	0.371	0.688	0.599	0.423
	1.2	1.000	1.000	0.998	0.989	0.988	0.970	0.849	0.780	0.642	0.719	0.637	0.468	0.774	0.688	0.538
	1.4	1.000	1.000	0.999	0.993	0.990	0.977	0.890	0.819	0.700	0.776	0.701	0.533	0.821	0.749	0.601

*The production set Ψ is strictly convex.

Table 3: Rejection Rates for Convexity Test (continued)

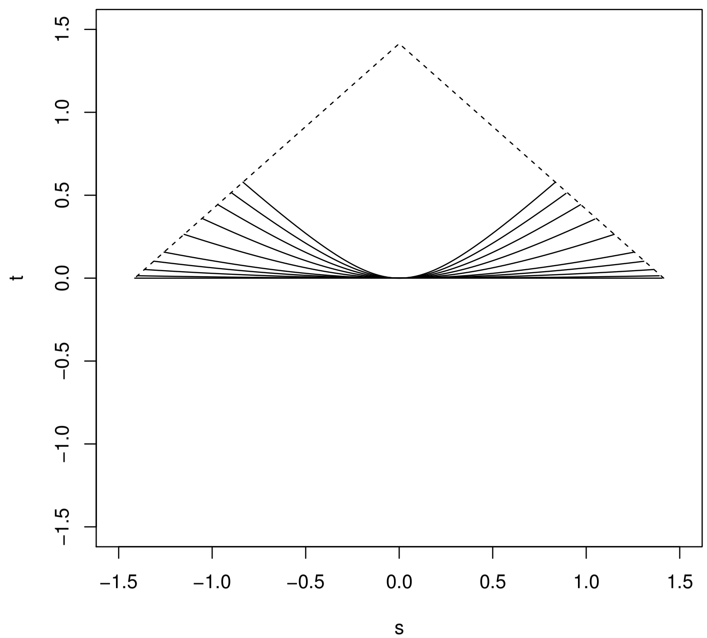
n	δ	$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
200	1.4*	0.101	0.056	0.014	0.125	0.080	0.033	0.104	0.066	0.019	0.099	0.069	0.016	0.095	0.062	0.020
	0.0	0.243	0.151	0.062	0.413	0.301	0.165	0.208	0.127	0.056	0.189	0.117	0.050	0.153	0.105	0.040
	0.1	0.266	0.169	0.075	0.431	0.321	0.176	0.219	0.135	0.058	0.196	0.123	0.052	0.156	0.107	0.041
	0.2	0.435	0.326	0.159	0.513	0.409	0.234	0.242	0.151	0.064	0.206	0.140	0.058	0.159	0.111	0.047
	0.3	0.757	0.661	0.467	0.659	0.573	0.405	0.296	0.213	0.085	0.237	0.166	0.069	0.188	0.128	0.057
	0.4	0.961	0.930	0.845	0.840	0.789	0.667	0.404	0.308	0.144	0.308	0.226	0.110	0.232	0.168	0.082
	0.6	0.998	0.996	0.985	0.965	0.952	0.918	0.656	0.554	0.340	0.494	0.389	0.223	0.418	0.304	0.162
	0.8	1.000	1.000	1.000	0.996	0.993	0.981	0.822	0.753	0.576	0.689	0.571	0.397	0.611	0.501	0.309
	1.0	1.000	1.000	1.000	0.999	0.999	0.997	0.917	0.862	0.738	0.803	0.717	0.543	0.765	0.653	0.453
	1.2	1.000	1.000	1.000	0.999	0.999	0.999	0.949	0.917	0.829	0.883	0.810	0.651	0.838	0.766	0.565
	1.4	1.000	1.000	1.000	1.000	1.000	1.000	0.974	0.945	0.878	0.916	0.872	0.718	0.881	0.830	0.677
1000	1.4*	0.099	0.053	0.013	0.121	0.068	0.016	0.100	0.061	0.020	0.115	0.062	0.019	0.096	0.059	0.015
	0.0	0.177	0.113	0.034	0.356	0.253	0.103	0.159	0.116	0.042	0.168	0.108	0.041	0.163	0.100	0.030
	0.1	0.247	0.161	0.062	0.405	0.283	0.123	0.165	0.118	0.044	0.174	0.115	0.042	0.167	0.105	0.030
	0.2	0.657	0.549	0.336	0.606	0.495	0.271	0.201	0.138	0.058	0.197	0.131	0.050	0.187	0.122	0.036
	0.3	0.978	0.959	0.906	0.920	0.878	0.733	0.326	0.218	0.105	0.261	0.179	0.073	0.227	0.151	0.053
	0.4	1.000	1.000	1.000	0.998	0.996	0.987	0.583	0.446	0.226	0.450	0.305	0.155	0.329	0.240	0.096
	0.6	1.000	1.000	1.000	1.000	1.000	1.000	0.868	0.795	0.596	0.759	0.654	0.411	0.624	0.487	0.293
	0.8	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.952	0.836	0.932	0.872	0.691	0.844	0.739	0.508
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.991	0.951	0.978	0.959	0.856	0.942	0.884	0.696
	1.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.984	0.995	0.981	0.933	0.975	0.943	0.804
	1.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	1.000	0.991	0.963	0.992	0.969	0.892

Table 3: Rejection Rates for Convexity Test (continued)

n	δ	$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
10000	1.4*	0.108	0.055	0.009	0.138	0.070	0.014	0.097	0.054	0.013	0.108	0.067	0.024	0.127	0.079	0.029
	0.0	0.139	0.076	0.014	0.310	0.205	0.058	0.140	0.068	0.024	0.169	0.105	0.032	0.169	0.111	0.048
	0.1	0.458	0.320	0.139	0.427	0.314	0.116	0.149	0.076	0.028	0.176	0.113	0.032	0.181	0.115	0.049
	0.2	0.998	0.996	0.985	0.920	0.882	0.750	0.248	0.133	0.040	0.211	0.133	0.046	0.204	0.127	0.057
	0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.531	0.386	0.149	0.361	0.248	0.109	0.301	0.219	0.087
	0.4	1.000	1.000	1.000	1.000	1.000	1.000	0.859	0.766	0.501	0.637	0.498	0.268	0.524	0.380	0.199
	0.6	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.987	0.933	0.939	0.887	0.681	0.838	0.725	0.484
	0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.995	0.982	0.921	0.968	0.932	0.749
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.986	0.997	0.984	0.925
	1.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.999	0.997	0.970
	1.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	0.999	0.988
20000	1.4*	0.102	0.053	0.011	0.153	0.080	0.022	0.130	0.063	0.020	0.104	0.059	0.023	0.107	0.070	0.027
	0.0	0.123	0.067	0.018	0.284	0.176	0.058	0.165	0.097	0.022	0.143	0.083	0.028	0.157	0.094	0.045
	0.1	0.581	0.444	0.217	0.479	0.342	0.150	0.178	0.099	0.023	0.146	0.087	0.031	0.162	0.096	0.046
	0.2	1.000	1.000	1.000	0.984	0.972	0.906	0.263	0.182	0.063	0.199	0.126	0.048	0.190	0.116	0.051
	0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.533	0.409	0.204	0.381	0.259	0.107	0.306	0.208	0.083
	0.4	1.000	1.000	1.000	1.000	1.000	1.000	0.903	0.816	0.558	0.707	0.552	0.284	0.538	0.407	0.198
	0.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.985	0.955	0.914	0.740	0.884	0.779	0.528
	0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.987	0.945	0.989	0.963	0.811
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.988	0.998	0.994	0.948
	1.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	1.000	0.999	0.990
	1.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.997

Figure 1: Function $g(\cdot)$ for Tests of Returns to Scale

(s, t)-space



(x, y)-space

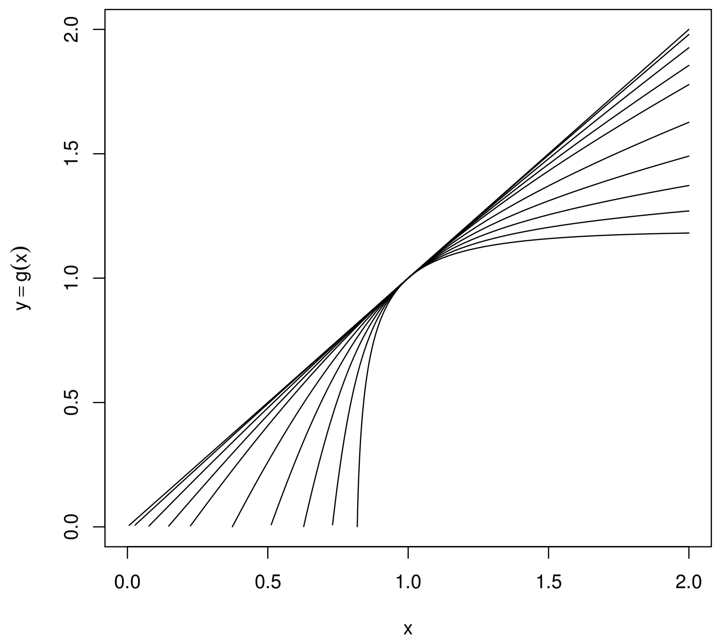


Figure 2: Function $g(\cdot)$ for Tests of Convexity