Testing Hypotheses in Nonparametric Models of Production

Alois Kneip
Department of Economics and Hausdorff Center for Mathematics
Universität Bonn
Bonn, Germany
email: akneip@uni-bonn.de

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

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

February 2015

Abstract

Data envelopment analysis (DEA) and free disposal hull (FDH) estimators are widely used to estimate efficiency of production. Practitioners use DEA estimators far more frequently than FDH estimators, implictly assuming that production sets are convex. Moreover, use of the constant returns to scale (CRS) version of the DEA estimator requires an assumption of CRS. Although bootstrap methods have been developed for making inference about the efficiencies of individual units, until now no methods exist 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. We use the central limit theorem results of Kneip et al. (2015) to develop additional theoretical results permitting consistent tests of model structure and provide Monte Carlo evidence on the performance of the tests in terms of size and power. 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. Using a sample of U.S. commercial banks, we test and reject convexity of the production set, calling into question results from numerous banking studies that have imposed convexity assumptions.

Keywords: asymptotic, inference, hypothesis test, efficiency, DEA, FDH.

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 Koopmans (1951) and Debreu (1951); 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; A search on google.com on 18 August 2014 using the keywords "dea" and "efficiency" yielded approximately 14.2 million results. 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, 2015). In addition, Kneip et al. (2015) 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. In particular, applied researchers must choose between FDH estimators (that do not impose convexity on the production set) and DEA estimators (that do impose convexity); among DEA estimators, one must choose between those that impose constant versus non-constant returns to scale. These choices suggest various hypothesis tests; yet in most applied work using FDH or DEA estimators, the choice between estimators is typically made arbitrarily. This paper extends the results of Kneip et al. (2015) 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. (2015) 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 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. (2015) 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, as noted above. 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 asymptotic results required for implementing the tests and making appropriate inference. Section 4 presents bootstrap versions of the tests developed in Section 3. The performance of the tests in finite samples is examined in a series of Monte Carlo experiments described in Section 5. In Section 6, we illustrate the application of our tests by revisiting the empirical study of Aly et al. (1990), who used DEA estimators, thus imposing convexity on the production set, to examine technical efficiency of U.S. commercial banks. We test and soundly reject the convexity imposed by Aly et al., calling into question the myriad other studies that have used DEA estimators to examine banks' inefficiency. 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 $(x, \widetilde{y}) \in \Psi$.*

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

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

was proposed by Deprins et al. (1984) to estimate Ψ. Replacing Ψ with $\widehat{\Psi}_{\mathrm{FDH}}(\mathcal{X}_n)$ on the right-hand side (RHS) of (2.3) yields the FDH estimator $\widehat{\theta}_{\mathrm{FDH}}(x, y \mid \mathcal{X}_n)$ of $\theta(x, y)$. Note that Afriat (1972, Theorem 1.1) defines a left- (but not right-) continuous function similar to the FDH estimator $\widehat{\Psi}_{\mathrm{FDH}}(\mathcal{X}_n)$ for the case $p \geq 1,\ q = 1$. Note, however, that $\widehat{\Psi}_{\mathrm{FDH}}(\mathcal{X}_n)$ 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.

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 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. (2014). 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. (2015) 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

3.1.1 Firms face same or different frontiers

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 Group 1 has greater mean efficiency. More formally, one might test the null hypothesis $H_0\colon \mu_{1,\theta} = \mu_{2,\theta}$ versus the alternative hypothesis $H_1\colon \mu_{1,\theta} > \mu_{2,\theta}$. *Mutatis mutandis*, alternative tests with a two-sided alternative or with other measures of efficiency would follow a similar procedure.

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.

In this version of the test, no restrictions are made on whether firms in the two groups face operate in the same production set, i.e., whether they face the same frontier. Below, in Section 3.1.2, we show how the ideas in this section can be adapted to test $H_0'\colon \mu_{1,\theta} = \mu_{2,\theta} \cap \Psi^{\partial 1} = \Psi^{\partial 2}$, where $\Psi^{\partial j}$ is the frontier faced by firms in the j group versus $H_1'\colon \mu_{1,\theta} \neq \mu_{2,\theta} \cup \Psi^{\partial 1} \neq \Psi^{\partial 2}$.

In many applications, one might want to test whether the frontier faced by two groups is the same. For example, Charnes et al. (1981) implicitly assumed that the two groups of schools in their study faced a common frontier, but maintained assumptions should be tested with data. Another example is provided by the literature on measurement of *changes* in productivity using Malmquist indices, which are often decomposed into terms representing changes in efficiency (for which our test is useful), changes in the technology, and other factors. At present, we are unable to offer a test of changes or differences in technologies across time or groups of firms, but work on this topic is underway. As will be seen below, simply testing whether mean efficiency is equivalent for two groups of firms involves considerable difficulties.

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. (2015) establishes (under appropriate regularity conditions; see Section A.1: 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.

The Lindeberg-Feller and other central limit theorems fail when FDH, VRS-DEA, or CRS-DEA estimators are averaged as in (3.1)–(3.2) due to the fact that while averaging drives the variance to zero, it does not diminish the bias. From Kneip et al. (2015) it can be seen that, unless $(p+q)$ is very small, scaling sample means such as (3.1)–(3.2) by a power of the sample size to stabilize the bias results in a degenerate statistic with the variance converging to zero. On the other hand, scaling if the power of the sample size is chosen to stabilize the variance, the bias explodes. Consequently, Kneip et al. (2015) use a bias estimator 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

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

giving a jackknife estimate of bias. Averaging in (3.6) reduces the variance by a factor of K^{-1}, as is usual with jackknife estimators. Note that 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. (2015) to hold.

Theorem 4.3 of Kneip et al. (2015) 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. (2015) 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 (3.7) does not hold, but Theorem 4.4 of Kneip et al. (2015) 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

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

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. (2015, 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. (2015) 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.,

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

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. Alternatively, 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 $n_1/n_2 \to c > 0$ 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.

Note that the results in (3.11) and (3.12) hold for *any* values of $\mu_{1,\theta}$ and $\mu_{2,\theta}$. Hence, if one tests $H_0\colon \mu_{1,\theta} = \mu_{2,\theta}$ versus an alternative hypothesis such as $H_1\colon \mu_{1,\theta} > \mu_{2,\theta}$ or perhaps $H_1\colon \mu_{1,\theta} \neq \mu_{2,\theta}$, the (asymptotic) distribution of the test statistic will be known under the null and up to $\mu_{1,\theta}$, $\mu_{2,\theta}$ under the alternative hypothesis. Consequently, given two independent samples, one can either (i) compute under the null (so that $(\mu_{1\theta} - \mu_{2,\theta}) = 0$) $\widehat{\tau}_{1,n_1,n_2}$ or $\widehat{\tau}_{2,n_{1,\kappa},n_{2,\kappa}}$ as appropriate, and compare the resulting value against a critical value from the $N(0,1)$ distribution, or (ii) use (3.11) and (3.12) to estimate a confidence interval for $(\mu_{1\theta} - \mu_{2,\theta})$. If the estimated interval does not include 0, one would reject the null; otherwise, one would fail to reject the null. Furthermore, the outcome will be the same under either approach; i.e., for a given test size, both approaches will either reject or fail to reject the null. It clearly follows that for a given departure from the null, the tests will reject the null with probability tending to one as $n_1, n_2 \to \infty$, and hence the tests are consistent. In addition, we expect that for given values of n_1 and n_2, the tests will reject more frequently with increasing departure from the null; this is confirmed by results from Monte Carlo experiments presented below.

3.1.2 Firms face the same frontier

As noted above, the tests presented in Section 3.1.1 make no restriction regarding whether firms in the two groups face a common frontier. It is straightforward to modify the previous tests in order to test equivalent means *and* a common frontier, i.e., $H_0' \colon \mu_{1,\theta} = \mu_{2,\theta} \cap \Psi^{\partial 1} = \Psi^{\partial 2}$ versus $H_1' \colon \mu_{1,\theta} \neq \mu_{2,\theta} \cup \Psi^{\partial 1} \neq \Psi^{\partial 2}$.

Let $n = n_1 + n_2$ and $\mathcal{X}_n = \mathcal{X}_{1,n_1} \cup \mathcal{X}_{2,n_2}$. Define $\widehat{\mu}'_{1,n_1}$ and $\widehat{\mu}'_{2,n_2}$ by replacing $\mathcal{X}_{1,n_1}$ and $\mathcal{X}_{2,n_2}$ which define the reference sets over which $\widehat{\theta}$ is computed with $\mathcal{X}_n$ in (3.1)–(3.2), but keep the summation over observations in $\mathcal{X}_{1,n_1}$ in (3.1) and over observations in $\mathcal{X}_{2,n_2}$ in (3.2). In other words, the means in (3.1)–(3.2) are computed as before, but the individual efficiencies are estimated using the combined sample.

The bias correction terms are computed analogously. Let $\mathcal{X}^{(1)}_{n/2}$ be the set of the first $[n/2]$ observations in $\mathcal{X}_n$, and let $\mathcal{X}^{(2)}_{n/2}$ be the set of remaining observations in $\mathcal{X}_n$. Then compute the bias corrections $\widehat{B}_{1,\kappa,n_1}$ and $\widehat{B}_{2,\kappa,n_2}$ using (3.3)–(3.6), but replacing in (3.3) the set of observations used to compute $\widehat{\theta}$ with $\mathcal{X}^{(j)}_{n/2}$, $j \in \{1,\ 2\}$. The sample variances in (3.10) should be computed as before, but the efficiency estimates under the summation sign should be computed using the full sample $\mathcal{X}_n$ instead of the individual samples $\mathcal{X}_{\ell,n_\ell}$.

Finally, for $(p+q) \leq 4$, 5, or 3 (for VRS-DEA, CRS-DEA, or FDH estimators), a test statistic $\widehat{\tau}'_{1,n_1,n_2}$ with limiting standard normal distribution can be computed by replacing on the right-hand side of (3.11) with their counterparts where efficiency have been computed using $\mathcal{X}_n$ or $\mathcal{X}^{(j)}_{n/2}$, $j \in \{1,\ 2\}$, as described above. For $(p+q) > 4$, 5, or 3 (corresponding to the estimators listed above) a test statistic $\widehat{\tau}'_{2,n_{1,\kappa},n_{2,\kappa}}$ with limiting standard normal distribution can be computed by replacing terms on the right-hand side of (3.12) with the appropriate counterparts. The sample means can be computed as just described, but with summation over subsets $n_{1,\kappa}$ and $n_{2,\kappa}$ observations instead of n_1, n_2 observations. For purposes of implementing the test, one can either use critical values from the $N(0,1)$ distribution, or use the bootstrap to estimate a confidence interval as described later in Section 4.

It is important to note that the sample means $\widehat{\tau}'_{1,n_1,n_2}$ and $\widehat{\tau}'_{2,n_{1,\kappa},n_{2,\kappa}}$ remain asymptotically uncorrelated, since summation is over disjoint sets of independent observations. However, using Theorems 3.1–3.3 in Kneip et al. (2015), it can be shown that the means have covariance of order $o(n^{-1})$, which leads to additional bias in finite samples that is not present in the statistics considered in Section 3.1.2. This finite sample covariance results from using the same observations to estimate individual efficiency scores, even though the resulting estimates are summed over disjoint sets to construct $\widehat{\tau}'_{1,n_1,n_2}$ and $\widehat{\tau}'_{2,n_{1,\kappa},n_{2,\kappa}}$.

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

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

and

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

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 $\widehat{\mu}_{\text{VRS},n}^{\text{full}} - \widehat{\mu}_{\text{CRS},n}^{\text{full}} \geq 0$. Under the null, one would expect $\widehat{\mu}_{\text{VRS},n}^{\text{full}} - \widehat{\mu}_{\text{CRS},n}^{\text{full}}$ to be "small," while under the alternative $\widehat{\mu}_{\text{VRS},n}^{\text{full}} - \widehat{\mu}_{\text{CRS},n}^{\text{full}}$ is expected to be "large."

Clearly, the variance of the difference $\widehat{\mu}_{\text{VRS},n}^{\text{full}} - \widehat{\mu}_{\text{CRS},n}^{\text{full}}$ is $\text{VAR}\left(\widehat{\mu}_{\text{VRS},n}^{\text{full}}\right) + \text{VAR}\left(\widehat{\mu}_{\text{CRS},n}^{\text{full}}\right) - 2\text{COV}\left(\widehat{\mu}_{\text{VRS},n}^{\text{full}},\ \widehat{\mu}_{\text{CRS},n}^{\text{full}}\right)$. By Theorem 4.1 of Kneip et al. (2015), the first two terms sum to $2n^{-1}\sigma_\theta^2$, and under the null

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

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. (2015) 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. (2015).

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

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

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

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

Then let

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

and

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

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

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

and

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

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

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

and

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

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. (2015) together with Theorem 3.1 given above ensure that

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

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. (2015) and Theorem 3.1 that appears above ensure

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

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.

A formal proof of consistency of the tests is beyond the scope of this paper. However, $\widehat{\theta}_{\mathrm{VRS}}(x,y)$, and hence $\widehat{\mu}_{\mathrm{VRS},n_1}$ and $\widehat{\mu}_{\mathrm{VRS},n_{1,\kappa}}$, remain consistent under the alternative hypothesis of variable returns to scale. Under the alternative, $\widehat{\theta}_{\mathrm{CRS}}(x,y)$ would estimate distance to the conical hull of a variable returns-to-scale production set; in view of the results of Park et al. (2010), it is easy to imagine the estimator would remain a consistent of distance to the boundary of the conical hull of Ψ, but with a slower convergence rate since the support of (X,Y) would lie away from the conical hull in most regions. In any case, it is clear that the two estimators of mean efficiency will diverge under the alternative, and the test statistics will have nonzero means, enabling the statistics to identify departures from the null. The simulation results presented in Section 5, while short of a proof, indicate that the tests are consistent.

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

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

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 $\widehat{\mu}^{\mathrm{full}}_{\mathrm{VRS},n} - \widehat{\mu}^{\mathrm{full}}_{\mathrm{FDH},n}$. By construction, $\widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_n) \leq \widehat{\theta}_{\mathrm{FDH}}(X_i, Y_i \mid$

$\mathcal{X}_n) \leq 1$ and therefore $\widehat{\mu}^{\text{full}}_{\text{FDH},n} - \widehat{\mu}^{\text{full}}_{\text{VRS},n} \geq 0$. Under the null, $\widehat{\mu}^{\text{full}}_{\text{VRS},n} - \widehat{\mu}^{\text{full}}_{\text{FDH},n}$ 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. (2015) 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 $n_1^{2/(p+q+1)} = \beta n_2^{1/(p+q)}$ and $n_1 + n_2 = n$ for a given sample size n, where β is a constant, and then solving for n_1 and n_2. Below, in Section 5, we examine the performance of the test for several choices of β. There is no closed-form solution, but it is easy to find a numerical solution by writing $n - n_1 - \beta^{-1} n_1^{2(p+q)/(p+q+1)} = 0$; 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. (2015) 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. (2015) 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,

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

for $(p+q) > 3$ by Theorem 4.4 of Kneip et al. (2015).

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.

Here again, a formal proof of consistency of the tests would be difficult and is beyond the scope of this paper for reasons similar to those given at the end of Section 3.2. Nonetheless, as in the previous case, the two means and their estimators will diverge under the alternative hypothesis of non-convexity, and hence the statistics will have non-zero mean, allowing the statistics to distinguish the alternative from the null hypothesis.

As a final remark, note that the idea of using the entire sample to estimate the efficiency scores of individual firms could be used here, as well as in the test of returns to scale. This might seem natural, since presumably there is only one frontier (otherwise, it would make little sense to test for CRS versus VRS or convexity versus non-convexity). However, as mentioned above in Section 3.1.1, using all of the data to compute the individual efficiency estimates results in finite-sample correlation between the two sample means used to define the test statistics. As will be seen below, in the case of testing for equivalent means across two groups of firms, this tends to worsen the size properties of the test relative to the case where the two sub-samples are kept entirely separate.

4 Bootstrap Tests

Under the null hypothesis of equivalent means, the statistics in (3.11) and (3.12) are asymptotically pivotal, as are the modified versions described in Section 3.1.2. Similarly, the statistics in (3.32) and (3.35) are asymptotically pivotal under the null hypothesis of CRS, and the statistics in (3.43) and (3.45) are asymptotically pivotal under the null hypothesis of convexity of the production set Ψ. Moreover, each of these statistics are asymptotically normal under the appropriate null hypotheses. Bootstrap methods sometimes provide better performance than tests based on asymptotic normality, especially when asymptotically pivotal statistics are available.

A bootstrap test of $H_0\colon \mu_{1,\theta} = \mu_{2,\theta}$ can be implemented by estimating a bootstrap confidence interval, and rejecting H_0 if the resulting interval does not cover zero. Implementation requires re-ordering the computations leading up to the bias-correction in (3.6). To illustrate, suppose the VRS-DEA estimator is used, and $(p+q) \leq 4$. Let $(X_{i\ell}, Y_{i,\ell})$, $i = 1, \ldots, n_\ell$ denote the ith observation in group $\ell \in \{1, 2\}$. For a given $k \in \{1, \ldots, K\}$, define

$$\widehat{\theta}^*_{i\ell;k} = \begin{cases} \widehat{\theta}\left(X_{i\ell}, Y_{i\ell} \mid \mathcal{X}^{(1)}_{\ell,m_{\ell,1},k}\right) & \text{if } (X_{i\ell}, Y_{i\ell}) \in \mathcal{X}^{(1)}_{\ell,m_{\ell,1},k}; \\ \widehat{\theta}\left(X_{i\ell}, Y_{i\ell} \mid \mathcal{X}^{(2)}_{\ell,m_{\ell,1},k}\right) & \text{if } (X_{i\ell}, Y_{i\ell}) \in \mathcal{X}^{(2)}_{\ell,m_{\ell,1},k}. \end{cases}$$

Let

$$\overline{\theta}^*_{i\ell;k} = (2^\kappa - 1)^{-1} \left(\widehat{\theta}^*_{i\ell;k} - \widehat{\theta}(X_{i\ell}, Y_{i\ell} \mid \mathcal{X}_{\ell,n_\ell})\right) \tag{4.1}$$

and then set

$$\overline{\theta}_{i\ell} = K^{-1} \sum_{k=1}^{K} \overline{\theta}^*_{i\ell;k}, \tag{4.2}$$

providing a bias correction for $\widehat{\theta}(X_{i\ell}, Y_{i\ell} \mid \mathcal{X}_{\ell,n_\ell})$, the original VRS-DEA estimator for group ℓ.

Note that for a given $k \in \{1, \ldots, K\}$,

$$\begin{aligned} n_\ell^{-1} \sum_{i=1}^{n_\ell} \overline{\theta}_{i\ell} &= (2^\kappa - 1)^{-1}\, n_\ell^{-1} \Big[\sum_{\mathcal{I}_1} \widehat{\theta}\left(X_{i\ell}, Y_{i\ell} \mid \mathcal{X}^{(1)}_{\ell,m_{\ell,1},k}\right) + \sum_{\mathcal{I}_2} \widehat{\theta}\left(X_{i\ell}, Y_{i\ell} \mid \mathcal{X}^{(2)}_{\ell,m_{\ell,2},k}\right) \\ &\qquad - \sum_{i=1}^{n_\ell} \widehat{\theta}\left(X_{i\ell}, Y_{i\ell} \mid \mathcal{X}_{\ell,n_\ell}\right) \Big] \\ &= (2^\kappa - 1)^{-1} \left(\frac{m_{\ell,1}}{n_\ell} \widehat{\mu}^{(1)}_{\ell,m_{\ell,1},k} + \frac{m_{\ell,2}}{n_\ell} \widehat{\mu}^{(2)}_{\ell,m_{\ell,2},k} \right) \end{aligned} \tag{4.3}$$

where $\mathcal{I}_1$ and $\mathcal{I}_2$ denote summation over $(X_{i\ell}, Y_{i\ell}) \in \mathcal{X}^{(1)}_{\ell,m_{\ell,1},k}$ and $(X_{i\ell}, Y_{i\ell}) \in \mathcal{X}^{(2)}_{\ell,m_{\ell,2},k}$, respectively. If n_ℓ is even, then $\frac{m_{\ell,1}}{n_\ell} = \frac{m_{\ell,2}}{n_\ell} = 0.5$ and $n_\ell^{-1} \sum_{i=1}^{n_\ell} \overline{\theta}^*_{i\ell;k} = \widetilde{B}_{\ell,\kappa,n_\ell,k}$ given in (3.5). If n_ℓ is odd, then $n_\ell^{-1} \sum_{i=1}^{n_\ell} \overline{\theta}^*_{i\ell;k} \approx \widetilde{B}_{\ell,\kappa,n_\ell,k}$, and the difference is negligible as $n_\ell \to \infty$. Hence the bias correction in (3.6) can now be written as

$$\widehat{B}_{\ell,\kappa,n_\ell} = n_\ell \sum_{i=1}^{n_\ell} \overline{\theta}_{i\ell}. \tag{4.4}$$

.

Now let $\mathcal{Z}_{\ell,n_\ell} = \left\{\left(\widehat{\theta}_{i\ell}, \overline{\theta}_{i\ell}\right)\right\}_{i=1}^{n_\ell}$ for $\ell \in \{1,\ 2\}$ where $\widehat{\theta}_{i\ell}$ denotes the original VRS-DEA estimate $\widehat{\theta}(X_{i\ell}, Y_{i\ell} \mid \mathcal{X}_{\ell,n_\ell})$ for observation i in group ℓ. In order to simplify notation, let $\zeta = (\mu_{1,\theta} - \mu_{2,\theta})$, $\widehat{\zeta}_{n_1,n_2} = (\widehat{\mu}_{1,n_1} - \widehat{\mu}_{2,n_2}) - \left(\widehat{B}_{1,\kappa,n_1} - \widehat{B}_{2,\kappa,n_2}\right)$, and $\widehat{s}_{n_1,n_2} = \left(n_1^{-1}\widehat{\sigma}^2_{1,\theta,n_1} + n_2^{-1}\widehat{\sigma}^2_{2,\theta,n_2}\right)^{1/2}$ so that (3.11) can be written as $\widehat{\tau}_{1,n_1,n_2} = (\widehat{\zeta}_{n_1,n_2} - \zeta)/\widehat{s}_{n_1,n_2}$. To implement the bootstrap test, compute the test statistic $\widehat{\tau}_{1,n_1,n_2}$ in (3.11) and then perform the following steps:

[1] For $b = 1$, draw n_1 pairs from $\mathcal{Z}_{1,n_1}$ independently, uniformly, and with replacement to construct a bootstrap sample $\mathcal{Z}^*_{1,n_1,b} = \left\{\left(\widehat{\theta}^*_{i1}, \overline{\theta}^*_{i1}\right)\right\}_{i=1}^{n_1}$. Draw similarly n_2 times from $\mathcal{Z}_{2,n_2}$ to construct the bootstrap sample $\mathcal{Z}^*_{2,n_2,b} = \left\{\left(\widehat{\theta}^*_{i2}, \overline{\theta}^*_{i2}\right)\right\}_{i=1}^{n_2}$.

[2] Compute the bootstrap statistic $\widehat{\tau}^*_{1,n_1,n_2,b} = \frac{\widehat{\zeta}^*_{n_1,n_2,b} - \widehat{\zeta}_{n_1,n_2}}{\widehat{s}^*_{n_1,n_2,b}}$ where

$$\zeta^*_{n_1,n_2,b} = \left(n_1^{-1}\sum_{i=1}^{n_1}\widehat{\theta}^*_{i1} - n_2^{-1}\sum_{i=1}^{n_2}\widehat{\theta}^*_{i2}\right) - \left(n_1^{-1}\sum_{i=1}^{n_1}\overline{\theta}^*_{i1} - n_2^{-1}\sum_{i=1}^{n_2}\overline{\theta}^*_{i2}\right), \tag{4.5}$$

$$\widehat{s}^*_{n_1,n_2,b} = \left(n_1^{-1}\widehat{\sigma}^{2*}_{1,\theta,n_1} + n_2^{-1}\widehat{\sigma}^{2*}_{2,\theta,n_2}\right)^{1/2}, \tag{4.6}$$

and

$$\widehat{\sigma}^{2*}_{\ell,\theta,n_\ell} = n_\ell^{-1}\sum_{i=1}^{n_\ell}\left(\widehat{\theta}^*_{i\ell} - n_\ell^{-1}\sum_{i=1}^{n_\ell}\widehat{\theta}^*_{i\ell}\right)^2 + \left(\overline{\theta}^*_{i\ell} - n_\ell^{-1}\sum_{i=1}^{n_\ell}\overline{\theta}^*_{i\ell}\right)^2 \tag{4.7}$$

for $\ell \in \{1,\ 2\}$.

[3] Repeat steps [1]–[2] for $b = 1,\ \ldots,\ B$ to produce a set $\mathcal{B} = \left\{\widehat{\tau}^*_{1,n_1,n_2,b}\right\}_{b=1}^{B}$ of bootstrap values.

[4] For a test of size α, construct a $(1-\alpha)\times 100$=percent confidence interval estimate

$$\left[\widehat{\zeta}_{n_1,n_2} - z^*_{1-\alpha/2}\widehat{s}_{n_1,n_2},\ \ \widehat{\zeta}_{n_1,n_2} - z^*_{\alpha/2}\widehat{s}_{n_1,n_2}\right]$$

by finding the $(\frac{\alpha}{2})$, $(1-\frac{\alpha}{2})$ percentiles $z^*_{\alpha/2}$, $z^*_{1-\alpha/2}$ of the set $\mathcal{B}$ of bootstrap values.

[5] If the interval computed in step 4 does not include 0, reject H_0; otherwise, fail to reject H_0.

The bootstrap described here is a naive bootstrap, i.e., resampling is accomplished by drawing from the empirical distribution of the observed input-output pairs. Here, the statistic of interest is a pivotal statistic under the null, and has asymptotically normally distribution. This is very different from inference about the efficiency of a fixed point, where the naive bootstrap cannot provide consistent inference (see Simar and Wilson, 2015 and the references cited therein for additional discussion).

Extension of the naive bootstrap outlined above to cases where the VRS-DEA estimator is used with $(p+q) > 4$ is straightforward. For this case, compute the bias corrections $\widehat{B}_{\ell,\kappa,n_\ell}$ for groups $\ell \in \{1,\ 2\}$ as before. Then let $\mathcal{Z}_{\ell,n_{\ell,\kappa}} = \left\{\left(\widehat{\theta}_{i\ell}, \overline{\theta}_{i\ell}\right)\right\}_{i=1}^{n_{\ell,\kappa}}$ denote a subset of size $n_{\ell,\kappa}$ of $\mathcal{Z}_{\ell,n_\ell}$ for for groups $\ell \in \{1,\ 2\}$ such that $\widehat{\mu}_{1,n_{1,\kappa}}$ is computed by summing over all $\widehat{\theta}_{i1}$ in $\mathcal{Z}_{1,n_{1,\kappa}}$ and $\widehat{\mu}_{2,n_{2,\kappa}}$ is computed by summing over all $\widehat{\theta}_{i2}$ in $\mathcal{Z}_{2,n_{2,\kappa}}$. To simplify notation again, let $\widehat{\psi}_{n_{1,\kappa},n_{2,\kappa}} = \left(\widehat{\mu}_{1,n_{1,\kappa}} - \widehat{\mu}_{2,n_{2,\kappa}}\right) - \left(\widehat{B}_{1,\kappa,n_1} - \widehat{B}_{2,\kappa,n_2}\right)$, and $\widehat{s}_{n_{1,\kappa},n_{2,\kappa}} = \left(n_{1,\kappa}^{-1}\widehat{\sigma}^2_{1,\theta,n_1} + n_{2,\kappa}^{-1}\widehat{\sigma}^2_{2,\theta,n_2}\right)^{1/2}$. Then (3.12) can be written as $\widehat{\tau}_{2,n_{1,\kappa},n_{2,\kappa}} = (\widehat{\psi}_{n_{1,\kappa},n_{2,\kappa}} - \zeta)/\widehat{s}_{n_{1,\kappa},n_{2,\kappa}}$.

To implement the bootstrap test, compute the test statistic $\widehat{\tau}_{2,n_{1,\kappa},n_{2,\kappa}}$ in (3.12) and then perform the following steps:

[1] For $b = 1$, draw $n_{1,\kappa}$ pairs from $\mathcal{Z}_{1,n_{1,\kappa}}$ independently, uniformly, and with replacement to construct a bootstrap sample $\mathcal{Z}^*_{1,n_{1,\kappa},b} = \left\{\left(\widehat{\theta}^*_{i1}, \overline{\theta}^*_{i1}\right)\right\}_{i=1}^{n_{1,\kappa}}$. Draw similarly $n_{2,\kappa}$ times from $\mathcal{Z}_{2,n_{2,\kappa}}$ to construct the bootstrap sample $\mathcal{Z}^*_{2,n_{2,\kappa},b} = \left\{\left(\widehat{\theta}^*_{i2}, \overline{\theta}^*_{i2}\right)\right\}_{i=1}^{n_{2,\kappa}}$.

[2] Compute the bootstrap statistic $\widehat{\tau}^*_{2,n_{1,\kappa},n_{2,\kappa},b} = \frac{\widehat{\zeta}^*_{n_{1,\kappa},n_{2,\kappa},b} - \widehat{\zeta}_{n_{1,\kappa},n_{2,\kappa}}}{\widehat{s}^*_{n_{1,\kappa},n_{2,\kappa},b}}$ where

$$\zeta^*_{n_{1,\kappa},n_{2,\kappa},b} = \left(n_{1,\kappa}^{-1}\sum_{i=1}^{n_{1,\kappa}}\widehat{\theta}^*_{i1} - n_{2,\kappa}^{-1}\sum_{i=1}^{n_{2,\kappa}}\widehat{\theta}^*_{i2}\right) - \left(n_{1,\kappa}^{-1}\sum_{i=1}^{n_{1,\kappa}}\overline{\theta}^*_{i1} - n_{2,\kappa}^{-1}\sum_{i=1}^{n_{2,\kappa}}\overline{\theta}^*_{i2}\right), \tag{4.8}$$

$$\widehat{s}^*_{n_{1,\kappa},n_{2,\kappa},b} = \left(n_{1,\kappa}^{-1}\widehat{\sigma}^{2*}_{1,\theta,n_{1,\kappa}} + n_{2,\kappa}^{-1}\widehat{\sigma}^{2*}_{2,\theta,n_{2,\kappa}}\right)^{1/2}, \tag{4.9}$$

and

$$\widehat{\sigma}^{2*}_{\ell,\theta,n_{\ell,\kappa}} = n_{\ell,\kappa}^{-1}\sum_{i=1}^{n_{\ell,\kappa}}\left(\widehat{\theta}^*_{i\ell} - n_{\ell,\kappa}^{-1}\sum_{i=1}^{n_{\ell,\kappa}}\widehat{\theta}^*_{i\ell}\right)^2 + \left(\overline{\theta}^*_{i\ell} - n_{\ell,\kappa}^{-1}\sum_{i=1}^{n_{\ell,\kappa}}\overline{\theta}^*_{i\ell}\right)^2 \tag{4.10}$$

for $\ell \in \{1,\ 2\}$.

[3] Repeat steps [1]–[2] for $b = 2,\ \ldots,\ B$ to produce a set $\mathcal{B} = \left\{\widehat{\tau}^*_{2,n_{1,\kappa},n_{2,\kappa},b}\right\}_{b=1}^{B}$ of bootstrap values.

Steps [4]–[5] are analogous to those for the previous bootstrap test, with the obvious changes in notation.

Bootstrap tests for constant versus variable returns to scale and for convexity versus non-convexity of Ψ can be constructed similarly using the appropriate efficiency estimators and corresponding rates of convergence.

5 Monte Carlo Evidence

5.1 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 $p \in \{1,\ 2,\ 3,\ 4,\ 5\}$. 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.

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

$$f(t \mid \lambda_k) = \begin{cases} \lambda_k t^{-2} e^{-\lambda_k (t^{-1}-1)} & \forall\, t \in (0,1], \\ 0 & \text{otherwise}, \end{cases} \tag{5.1}$$

with $k = 1$ or 2 depending on whether observations are generated for sample 1 or 2. Note that $1/\theta$ has exponential density $f(t \mid \lambda_k) = \lambda_k e^{\lambda_k(t-1)}\ \forall\ t \in [1, \infty)$. Moreover, given a $(p+q)$-vector $\boldsymbol{v}$ of draws from the uniform distribution on $[0,1]$, $\boldsymbol{u} = \boldsymbol{v}(\boldsymbol{v}'\boldsymbol{v})^{-1/2}$ is a vector of coordinates from a uniform distribution on the unit sphere in $\mathbb{R}^{p+q}$. Setting $\boldsymbol{y} = |\boldsymbol{u}_q|$ amounts to reflecting any point that lies below one or more of the $\boldsymbol{u}_p$ axes around those axes. Similarly, $-|\boldsymbol{u}_p|$ reflects around the $\boldsymbol{u}_q$ axes, but in negative directions; adding 1 shifts the resulting points to the positive orthant in $\mathbb{R}^{p+q}$. This amounts to generating uniform points on a unit sphere centered at $\left[\boldsymbol{1}_p', \boldsymbol{0}_q'\right]'$, reflecting the points so that all lie on the part of the sphere in the unit hypercube with in the positive orthant with a corner at the origin, and then projecting points away from this "frontier" in the input directions. 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 set $q = 1$ and model the technology by

$$Y = g\left(\prod_{j=1}^{p}\left(\widetilde{X}^j\right)^{1/p}\right), \tag{5.2}$$

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

$$t = c\left(a^2 + \delta^2 s^2\right)^{1/2} - d \tag{5.3}$$

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 $\delta = 0$, (5.3) 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 1. 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 $(0, \sqrt{2})$ 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 (5.3) 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 (5.3) intersects the line $t = \sqrt{2} - s$, and then generating uniform random numbers s_i on the interval $\big(\max(-s_{\max}, -0.9\sqrt{2}),\ \min(s_{\max}, 0.9\sqrt{2})\big)$. Plugging these into (5.3) 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 (5.1) 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 (5.2), 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 (5.1), 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 (5.2), 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.

5.2 Results of simulation experiments

Monte Carlo estimates of rejection rates for the two-sample test (without assuming a common frontier) using the VRS-DEA estimator and relying on the asymptotic normality results 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 $\widehat{\tau}_{2,n_{1,\kappa},n_{2,\kappa}}$ 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.

Recall from the discussion in Section 3.1.1 that sample means based on subsets of the observations in groups 1 and 2 must be used when the number of dimensions (i.e., $p+q$) exceeds 4 when the VRS-DEA estimator is used; hence among the five cases $(p + q) \in \{2, \ \ldots, \ 6\}$

represented in Table 1, different statistics are used for the first 3 cases and the last 2 cases. Per the discussion at the end of Section 4 in Kneip et al. (2015), the statistic used for the first 3 cases in Table 1 is scaled by $n^{1/2}$ and hence neglects a term of order $O(n^{-1/10})$, whereas the statistic used for the last 3 cases is scaled by n^{κ} and hence the neglected term is of order $O(n^{-1/5})$. As with coverages of confidence intervals for mean efficiency examined by Kneip et al. (2015), performance in terms of realized size of our tests using the VRS-DEA estimator improves when moving from $(p + q) = 4$ to $(p + q) = 5$ because of the different scaling.

For 1,000 or fewer observations in each sample, the results in Table 1 for $(p + q) = 2$, 3, for 4 indicate that the realized test sizes are too large. The same problem occurs for $n_1 = n + 2 = 50$, 100, or 200 when there are five or six dimensions and the statistic $\widehat{\tau}_{2,n_{1,\kappa},n_{2,\kappa}}$ based on sub-samples is used. Even with $n_1 = n_2 = 10,000$, the rejection rate for $p = q = 2$ remains at 13.3 percent when the nominal test size is .10. On the other hand, 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.

Analogous to Table 1, Table 2 shows a similar set of results for rejection rates for tests of equivalent means across two samples using the bootstrap method described in Section 4. Examination of the rejection rates in the first row of the cells corresponding to different sample sizes reveals that the bootstrap yields realized test sizes smaller than those obtained in Table 1 based on asymptotic normality. For $n_1 = n_2 = 50$ or 100, the bootstrap gives test sizes that are typically closer to nominal test sizes than does asymptotic normality. However, as sample sizes increase beyond 100, the test sizes in Table 2 continue to become smaller. For $n_1 = n + 2 = 1,000$ and larger, the bootstrap gives conservative tests, with sizes smaller than nominal values. However, the power of the tests is good in the sense that for given departures from the null, the bootstrap yields rejection rates that in many cases are larger than those obtained with the tests relying on asymptotic normality in Table 1.

Similar experiments were conducted to examine the performance of tests of equivalent means *and* common frontiers as discussed above in Section 3.1.2. Performance of tests based on both asymptotic normality and bootstrap methods were examined. Detailed results are available from the authors in a separate appendix; we do not give tables of rejection rates here in order to conserve space. Overall, the results can be summarized by noting that the rejection rates obtained with either the asymptotic normal or the bootstrap version of the tests are greater than the corresponding rejection rates in Tables 1 and 2. This is to be expected since the simulated model is the same in both cases, but the hypothesis that is tested is more restrictive than the one that is tested in Table 1. In addition, the comparison between the asymptotic normal and bootstrap versions of the test is qualitatively similar to the previous comparison; i.e., the bootstrap gives smaller rejection rates when the null is true than does asymptotic normality, with the bootstrap version of the test eventually becoming conservative as sample sizes increase. As in the previous comparison, the bootstrap in many cases results in tests with more than power than the tests based on asymptotic normality.

The next set of results given in Table 3 illustrate the performance of the returns to scale test relying on asymptotic normality, 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 3 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 3 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 4 shows rejection rates obtained with the bootstrap test of constant versus variable returns to scale using the ideas presented above in Section 4. As with the tests of equivalent means, here the bootstrap yields realized test sizes that are smaller than the tests based on asymptotic normality whose results are given in Table 3. As the sample size increases, the bootstrap tests eventually become conservative, with realized sizes smaller than the nominal sizes. However, the power remains strong; as departure from the null increases (i.e., as δ

increases from 0.0), the rejection rates often reach or exceed those obtained in Table 3.

Table 5 gives results for the simulations for the convexity test relying on asymptotic normality as presented in Section 3.3. 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 *concave* (from above), while the second row in each case corresponds to the case where $g()$ is linear, i.e., weakly convex, reflecting the fact that the null is a composite hypothesis. Rows 2–11 correspond to increasing departures from the null hypothesis of convexity. The results shown in Table 5 were obtained by setting $\beta = 1$ to determine the sample split as described in Section 3.3.

Overall conclusions similar to those drawn in the two previous sets of results can be drawn from the results in Table 5, 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). In addition, rejection rates are smaller when $g()$ is strictly concave than when it is weakly convex. The test tends to over-reject, although not in every case. As with the test of equivalent means, the situation is somewhat better with larger dimensionality, however, where subsamples of observations are used; for $n = 1,000$ and $p + q = 4$, 5, and 6, the estimated rejection rates are between 5.5 and 8.7 percent at the five-percent level when $g()$ is 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 expected in a number of cases when $g()$ is only weakly convex.

Rejection rates for the bootstrap test of convexity are presented in Table 6, where the layout is similar to that in Table 5. The sample split is again determined using $\beta = 1$ as discussed in Section 3.3. The results shown in Table 6 indicate that with $(p + q) = 2$ or 3, and $n = 50$, 100, or 200, the bootstrap yields realized test sizes that are much closer to nominal test sizes than does the test based on asymptotic normality. As sample size increases beyond 200, the bootstrap test eventually becomes conservative, with realized sizes less than nominal sizes. Nonetheless, the bootstrap test has power almost as good as the test relying on asymptotic normality. For example, with nominal test size .10, $n = 20,000$, and $(p + q) = 6$, the rejection rate for the bootstrap test when $\delta = 0.0$ is only 3.2 percent, versus 10.2 percent with the asymptotic normal version of the test in Table 5. However, when $\delta = 0.4$, the bootstrap test rejects at 59.5 percent, versus only 45.2 percent with the asymptotic normal test.

As discussed in Section 3.3, a given sample $\mathcal{X}_n$ of input-output pairs must be split into two parts to implement the test of convexity of Ψ. The rejection rates reported in Tables 5 and 6 were obtained using $\beta = 1$ to determine the sample splits, but of course other values could be used. We repeated the simulations described above for the test of convexity, allowing the value of the constant β to vary in order to check the sensitivity of the realized test sizes with respect to the choice of value for β. Figure 3 shows the results of these experiments for tests relying on asymptotic normality, with $\beta \in \{8.0,\ 4.0,\ 2,\ .01.5,\ 1.0,\ 0.8,\ 0.7,\ 0.6,\ 0.5,\ 0.4,\ 0.3,\ 0.2\}$, under the null hypothesis with $\delta = 0.0$. Recalling the expression given in Section 3.3, decreasing values of β correspond to increasing values of n_2 and greater proportions of the observations being used by the FDH estimator, as opposed to the VRS-DEA estimator. The six panels in Figure 3 show the estimated rejection rates for each of the six sample sizes n considered in Table 5. Within each panel, rejection rates for nominal test size .05 and number of dimension $(p + q) \in \{1,\ 2,\ \ldots,\ 6\}$ are plotted as functions of β, and a horizontal line intersecting the vertical axis at .05 is plotted for reference.

The results shown in Figure 3 indicate that rejection rates typically decrease with β over the range of values considered. The results also suggest that the choice of value for β matters less in terms of achieved test size as n increases. Moreover, for $(p + q) = 4$, 5, or 6, setting $\beta = 1$ seems to be a reasonable choice. For $(p + q) = 2$ or 3, the performance of the convexity test, in terms of size, might be improved by decreasing the value of β below 1 (the value used to produce the results in Table 5). Similar plots for the bootstrap version of the convexity test show a similar pattern; these are omitted here to conserve space, but are available in the separate appendix mentioned earlier.

There is, unfortunately, no theory to guide the choice of value for β in applications, but one might adapt the data-driven method for choosing subsample sizes used by Simar and Wilson (2011) by looking at p-values corresponding to a grid of values of β, and then looking for a range of values of β over which the variation in the corresponding p-values is minimized. Note that a similar issue exists for the test of returns to scale; the results in Table 3 were produced by splitting the sample evenly, given that the VRS-DEA and CRS-DEA estimators have the same convergence rate under the null, but one could give more observations to either of the two estimators and fewer observations to the other estimator.

As noted above at the end of Section 5.1, the experiments whose results are displayed in Tables 1–6 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 appear in the previously mentioned, separate appendix that is 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 using asymptotic normality, 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 0.150, 0.216, 0.336, 0.177, and 0.190 for 2–6 dimensions. For comparison, the corresponding realized sizes in Table 1 are 0.120, 0.179, 0.275, 0.160, and 0.156. 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 $n_1 = n_2 = 1000$. Similar differences can be seen when the results for the returns to scale and convexity tests with $K = 1$ are compared to results with $K = 100$. When bootstrap tests are used, the difference between realized test sizes with and without averaging are larger than with the tests that rely on asymptotic normality. The results again suggest that averaging should be employed for sample sizes up to about 1,000. Without averaging, the bootstrap test of equivalent means yields realized sizes much closer to nominal sizes for $n_1 = n_2 = 10,000$ or 20,000 than when the bias corrections are averaged, though the results at these sample sizes are no better than those obtained with the asymptotic normal version of the test. Similar observations hold for the tests of constant versus variable returns to scale and convexity versus non-convexity of Ψ. 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.

6 An Empirical Application: U.S. Commercial Banks

Aly et al. (1990) examine a random sample of 322 independent U.S. commercial banks operating at year-end 1986. Aly et al. estimate various types of efficiency, including technical efficiency, in each case using VRS-DEA estimators and thereby imposing convexity on the production set while allowing for variable returns to scale.

In order to test the convexity assumption imposed by Aly et al., data from the December 31, 2013 Federal Deposit Insurance Corporation Reports of Condition and Reports of Income (Call Reports) were used to reconstruct the input-output variables used by Aly et al.. These include $p = 3$ inputs (labor, physical capital, and loanable funds) and $q = 5$ outputs (real estate loans, commercial and industrial loans, consumer loans, all other loans, and demand deposits). Treatment of demand deposits as an output reflects the view that transaction accounts are primarily a service provided by banks. Loanable funds include time and savings deposits, notes and debentures, and other borrowed funds. Loans and deposits are measured by the stock in current dollars as of December 31, 2013. Physical capital is measured by the book value in current dollars of premises and equipment, while labor is measured in units of full-time equivalents. After omitting observations with missing values for one or more variables, $n = 5,551$ observations remain.

Applying the FDH and VRS-DEA estimators in (2.6) and (2.7) to each observation in the sample and then averaging yields sample means of 0.9424 and 0.4677 for the FDH and VRS-DEA estimates, respectively. However, recall from Section 3.3 that in order to test convexity of the production set, the sample must first be split into two unequal parts, reflecting the fact that the FDH estimator converges more slowly than the VRS-DEA estimator. Following the method described in Section 3.3 and setting $\beta = 1$ results in subsamples of sizes $n_1 = 128$ and $n_2 = 5,551$. Since $p + q = 8$, the test statistic $\widehat{\tau}_{6,n}$ defined in (3.45) must be used, and the data yield a value $\widehat{\tau}_{6,n} \approx 3.302$. The corresponding p-value (assuming asymptotic normality) is approximately 0.00048, while the p-value based on a bootstrap with 2,000 replications is 0.04. These results, combined with the facts (i) that the mean of the FDH estimates is more than two times the mean of the VRS-DEA estimates and (ii) that the FDH estimator remains consistent under non-convexity while the VRS-DEA estimator does not, suggest that the null hypothesis of convexity should be rejected.

We applied the same test on the original data used by Aly et al. (1990) and obtained a value of 3.920 for the test statistic $\widehat{\tau}_{6,n}$ (we are grateful to Richard Grabowski for providing the original data). Using asymptotic normality, this gives a p-value of approximately 0.000044, while bootstrapping (with 2,000 replications) yields a p-value of 0.249. Hence with the original data used by Aly et al., the tests give ambiguous results. Although our simulations indicate that the bootstrap leads to conservative tests, our simulation results also indicate that tests based on asymptotic normality tend to over-reject. The mean VRS-DEA and FDH estimates for all of the 322 observations in the data used by Aly et al. are 0.8021 and 0.9957, respectively, and so the difference in mean efficiencies is less than in the previous case. Note, however, that Aly et al. sampled only *independent* banks, excluding members of multi-bank holding companies, yielding a more homogeneous, less variable sample than what might have been obtained otherwise.

Dozens, perhaps hundreds, of published studies (see the bibliography of Gattoufi et al., 2004 have used DEA estimators (usually, VRS-DEA, but sometimes CRS-DEA) to examine efficiency among commercial banks. While these have used various specifications of inputs and outputs, most of these specifications are at least similar to the variable specification of Aly et al. (1990), where banks are viewed as financial intermediaries that combine labor, capital, and borrowed money (i.e., deposits) to make loans of various types. Rejection of the convexity assumption here with the 2013 data raises questions about whether other banking studies that have imposed convexity on the production set provide reasonable estimates of inefficiency. The rejection of convexity here is also consistent with recent results obtained by Wheelock and Wilson (2011, 2012), Hughes and Mester (2014), and others who have found evidence of increasing returns to scale throughout the size-ranges of banks and credit unions by estimating cost functions using nonparametric regression techniques.

If one adopts the view that banks' technologies are heterogeneous, then estimating a single frontier for all banks amounts to estimation of a meta-frontier as described by O'Donnell et al. (2008). On the other hand, banks may have different business plans, and may choose to operate in different parts of the production set, while facing a common frontier.

7 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. Asymptotic normal distributions of our test statistics follow from the new central limit theorem results of Kneip et al. (2015). For each hypothesis tested, we consider both tests relying on asymptotic normality for critical vales as well as tests employing bootstrap methods. The Monte Carlo results we presented in Section 5 indicate that performance of the tests relying on asymptotic normality, in terms of realized sizes, improves as sample size increases.

The Monte Carlo results also indicate that our tests that rely on asymptotic normality to find critical values tend to over-reject, particularly in samples of less than a few hundred observations. The bootstrap test of equivalent means provides improved size properties, while the bootstrap tests of constant returns to scale and of convexity tend to under-reject in most cases, providing conservative tests.

As noted above in Section 5.2, the experimental results provide some practical guidance for applied researchers; i.e., one should be cautious in drawing conclusions when one of our tests relying on asymptotic normality for critical values just barely rejects the null. On the other hand, one can be more confident when estimated p-values are 0.01 or less, as with the empirical application in Section 6. Alternatively, the bootstrap tests can be used with little increase in computational burden.

To give another 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 5.2 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.

Acknowledgments: 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.

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. (2015).

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. (2015) 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. (2015) and used in Section 3 depend on Assumptions 2.1–2.2 and A.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 CRS-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. (2015). 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. (2015). 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. (2015) that the CRS-estimator of $\theta(x, y)$ exactly coincides with a VRS-DEA estimator based on the reduced sample of observations $\widetilde{\mathcal{X}}_n$:

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

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. (2015) 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

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

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

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

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

Now consider assertion (ii) of the theorem. 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 $1-\theta^*(x'^*,\widetilde{y}')\ge\nu_n^2$, $|z'^j|\le\nu_n$, $j=1,\ldots,p-1$, and $|\widetilde{y}'^j-\widetilde{y}^j|\le\nu_n$, $j=1,\ldots,q-1$. If $\boldsymbol{\omega}^{opt}$ denotes the vector $\boldsymbol{\omega}\in\mathbb{R}^n_+$ providing the minimal value of $\widehat{\theta}_{\mathrm{VRS}}(x,y\mid\mathcal{X}_n)$ in (A.2), then a straightforward generalization of the localization arguments given in Kneip et al. (2008) and Kneip et al. (2015) 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. (2015) now may be used to show that for any $y^1_{min}<y^1<y^1_{max}$,

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

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. (2015) 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

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

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 $E\left(\widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) = \int_{y^1_{min}}^{y^1_{max}} E\left(\widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\big|Y_i^1 = y^1\right) \bar{f}_1(y^1)dy^1$, 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 $0 \leq \widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i) \leq 1$ 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 $E\left(\widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) = O(n^{-1}\log n)$ 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. (2015) 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.

Aly, H. Y., R. G. Grabowski, C. Pasurka, and N. Rangan (1990), Technical, scale, and allocative efficiencies in U.S. banking: an empirical investigation, *Review of Economics and Statistics* 72, 211–218.

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 (2014), Measuring firm performance using nonparametric quantile-type distances, *Econometric Reviews* Forthcoming.

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.

Hughes, J. P. and L. J. Mester (2014), Who said large banks don't experience scale economies? evidence from a risk-rturn-driven cost function, *Journal of Financial Intermediation* Forthcoming.

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.

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* 38, 483–515.

— (2015), When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores, *Econometric Theory* Forthcoming.

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.

O'Donnell, C. J., D. S. P. Rao, and G. E. Battese (2008), Metafrontier frameworks for the study of firm-level efficiencies and technology ratios, *Empirical Economics* 34, 231–255.

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.

— (2011), Inference by the m out of n bootstrap in nonparametric frontier models, *Journal of Productivity Analysis* 36, 33–53.

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

— (2015), Statistical approaches for non-parametric frontier models: A guided tour, *International Statistical Review* Forthcoming.

Wheelock, D. C. and P. W. Wilson (2011), Are credit unions too small?, *Review of Economics and Statistics* 93, 1343–1359.

— (2012), Do large banks have lower costs? new estimates of returns to scale for U.S. banks, *Journal of Money, Credit, and Banking* 44, 171–199.

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 Two-Sample Test of Equivalent Means using Asymptotic Normality

$n_1 = n_2$	λ_2	$\widehat{\tau}_{1,n_1,n_2}$									$\widehat{\tau}_{2,n_{1,\kappa},n_{2,\kappa}}$					
		$p=1,\ q=1$			$p=2,\ q=1$			$p=2,\ q=2$			$p=3,\ q=2$			$p=3,\ q=3$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
50	2.00	0.192	0.120	0.046	0.254	0.179	0.076	0.357	0.275	0.154	0.238	0.160	0.059	0.229	0.156	0.065
	1.90	0.206	0.128	0.045	0.280	0.192	0.076	0.391	0.296	0.170	0.231	0.153	0.054	0.214	0.148	0.056
	1.80	0.255	0.174	0.064	0.297	0.208	0.109	0.412	0.331	0.197	0.238	0.165	0.067	0.239	0.163	0.073
	1.70	0.287	0.188	0.078	0.342	0.253	0.137	0.462	0.381	0.245	0.241	0.172	0.078	0.243	0.160	0.073
	1.60	0.355	0.245	0.116	0.422	0.320	0.167	0.524	0.447	0.305	0.293	0.211	0.097	0.306	0.229	0.098
	1.50	0.459	0.371	0.189	0.494	0.409	0.232	0.593	0.521	0.378	0.346	0.255	0.134	0.326	0.237	0.117
	1.40	0.565	0.464	0.263	0.591	0.509	0.340	0.672	0.588	0.433	0.405	0.302	0.159	0.362	0.268	0.143
	1.30	0.671	0.585	0.370	0.697	0.616	0.443	0.752	0.698	0.565	0.462	0.361	0.196	0.434	0.346	0.180
	1.20	0.754	0.676	0.479	0.813	0.737	0.565	0.809	0.767	0.646	0.541	0.442	0.251	0.493	0.393	0.222
	1.10	0.867	0.802	0.631	0.865	0.812	0.667	0.885	0.854	0.761	0.609	0.509	0.316	0.573	0.457	0.289
	1.00	0.930	0.887	0.753	0.930	0.900	0.794	0.919	0.894	0.816	0.710	0.607	0.435	0.632	0.521	0.321
	0.75	0.997	0.993	0.961	0.997	0.991	0.962	0.986	0.979	0.955	0.868	0.800	0.617	0.795	0.725	0.524
	0.50	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	0.971	0.948	0.853	0.943	0.892	0.740
100	2.00	0.149	0.091	0.025	0.218	0.146	0.059	0.310	0.216	0.114	0.172	0.104	0.039	0.154	0.085	0.028
	1.90	0.174	0.104	0.032	0.244	0.164	0.062	0.332	0.247	0.123	0.167	0.107	0.044	0.147	0.086	0.026
	1.80	0.251	0.174	0.063	0.278	0.194	0.101	0.359	0.266	0.140	0.208	0.127	0.053	0.188	0.120	0.033
	1.70	0.345	0.242	0.101	0.407	0.315	0.181	0.473	0.389	0.246	0.223	0.147	0.061	0.209	0.147	0.055
	1.60	0.493	0.378	0.188	0.543	0.447	0.259	0.582	0.485	0.335	0.276	0.197	0.085	0.246	0.170	0.059
	1.50	0.622	0.514	0.313	0.688	0.602	0.387	0.712	0.636	0.474	0.353	0.265	0.125	0.300	0.213	0.102
	1.40	0.774	0.678	0.478	0.804	0.731	0.544	0.801	0.746	0.610	0.418	0.331	0.160	0.340	0.235	0.112
	1.30	0.880	0.817	0.640	0.897	0.846	0.727	0.894	0.860	0.750	0.522	0.416	0.229	0.428	0.329	0.173
	1.20	0.955	0.911	0.807	0.962	0.935	0.849	0.954	0.932	0.853	0.661	0.560	0.347	0.494	0.381	0.216
	1.10	0.989	0.978	0.923	0.986	0.968	0.923	0.981	0.972	0.933	0.755	0.661	0.436	0.580	0.469	0.283
	1.00	0.998	0.996	0.970	0.999	0.995	0.980	0.997	0.994	0.978	0.850	0.785	0.580	0.662	0.562	0.355
	0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.967	0.934	0.824	0.887	0.812	0.619
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.996	0.963	0.974	0.945	0.837

Table 1: Rejection Rates for Two-Sample Test of Equivalent Means using Asymptotic Normality (continued)

		$\widehat{\tau}_{1,n_1,n_2}$									$\widehat{\tau}_{2,n_{1,\kappa},n_{2,\kappa}}$					
		$p=1,\ q=1$			$p=2,\ q=1$			$p=2,\ q=2$			$p=3,\ q=2$			$p=3,\ q=3$		
$n_1=n_2$	λ_2	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
200	2.00	0.155	0.082	0.024	0.193	0.113	0.031	0.269	0.182	0.069	0.137	0.079	0.023	0.117	0.069	0.026
	1.90	0.209	0.130	0.054	0.249	0.161	0.058	0.270	0.195	0.089	0.149	0.078	0.028	0.126	0.076	0.026
	1.80	0.315	0.224	0.101	0.349	0.250	0.125	0.413	0.320	0.177	0.190	0.119	0.042	0.153	0.092	0.029
	1.70	0.514	0.391	0.221	0.561	0.449	0.266	0.562	0.484	0.315	0.256	0.174	0.065	0.195	0.122	0.049
	1.60	0.723	0.614	0.374	0.746	0.667	0.448	0.743	0.661	0.495	0.346	0.247	0.101	0.284	0.199	0.082
	1.50	0.871	0.811	0.608	0.886	0.827	0.675	0.882	0.834	0.710	0.442	0.330	0.167	0.333	0.228	0.108
	1.40	0.972	0.941	0.853	0.955	0.931	0.845	0.960	0.943	0.877	0.545	0.427	0.241	0.445	0.338	0.163
	1.30	0.993	0.990	0.950	0.998	0.993	0.966	0.987	0.979	0.947	0.699	0.591	0.348	0.546	0.424	0.217
	1.20	1.000	0.998	0.989	0.998	0.996	0.988	0.996	0.993	0.983	0.801	0.701	0.482	0.645	0.536	0.320
	1.10	1.000	0.999	0.998	1.000	1.000	0.999	1.000	1.000	0.999	0.875	0.803	0.596	0.770	0.658	0.433
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.945	0.894	0.744	0.851	0.780	0.574
	0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.994	0.967	0.980	0.951	0.851
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.982
1000	2.00	0.121	0.072	0.012	0.141	0.079	0.020	0.169	0.098	0.025	0.106	0.053	0.012	0.111	0.054	0.016
	1.90	0.301	0.211	0.089	0.366	0.260	0.096	0.378	0.296	0.158	0.157	0.085	0.017	0.130	0.074	0.020
	1.80	0.725	0.619	0.375	0.776	0.688	0.464	0.787	0.702	0.483	0.227	0.144	0.044	0.194	0.123	0.033
	1.70	0.970	0.946	0.840	0.983	0.962	0.870	0.973	0.944	0.872	0.369	0.253	0.101	0.272	0.172	0.066
	1.60	0.998	0.998	0.992	1.000	0.998	0.986	0.999	0.998	0.989	0.582	0.436	0.217	0.372	0.268	0.128
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.738	0.621	0.388	0.573	0.440	0.235
	1.40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.866	0.771	0.550	0.689	0.592	0.363
	1.30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.956	0.914	0.778	0.819	0.719	0.507
	1.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.988	0.978	0.916	0.918	0.839	0.643
	1.10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.995	0.975	0.968	0.935	0.812
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.992	0.981	0.918
	0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 1: Rejection Rates for Two-Sample Test of Equivalent Means using Asymptotic Normality (continued)

		$\widehat{\tau}_{1,n_1,n_2}$									$\widehat{\tau}_{2,n_{1,\kappa},n_{2,\kappa}}$					
		$p=1,\ q=1$			$p=2,\ q=1$			$p=2,\ q=2$			$p=3,\ q=2$			$p=3,\ q=3$		
$n_1=n_2$	λ_2	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
10000	2.00	0.112	0.066	0.012	0.113	0.063	0.011	0.133	0.074	0.015	0.101	0.045	0.008	0.110	0.049	0.005
	1.90	0.965	0.938	0.824	0.975	0.956	0.860	0.981	0.952	0.875	0.223	0.136	0.045	0.148	0.086	0.019
	1.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.549	0.412	0.204	0.315	0.210	0.081
	1.70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.841	0.751	0.521	0.542	0.414	0.195
	1.60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.958	0.930	0.833	0.798	0.678	0.448
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.992	0.978	0.923	0.865	0.684
	1.40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.983	0.969	0.892
	1.30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.995	0.973
	1.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996
	1.10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
20000	2.00	0.093	0.043	0.006	0.099	0.050	0.007	0.114	0.056	0.014	0.088	0.032	0.006	0.095	0.048	0.016
	1.90	0.999	0.999	0.996	1.000	1.000	0.995	0.999	0.999	0.996	0.295	0.196	0.065	0.162	0.094	0.028
	1.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.664	0.571	0.328	0.377	0.274	0.109
	1.70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.942	0.901	0.743	0.696	0.565	0.311
	1.60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.995	0.968	0.884	0.794	0.577
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.975	0.960	0.870
	1.40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.994	0.971
	1.30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	1.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	1.10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 2: Rejection Rates for Two-Sample Test of Equivalent Means using Bootstrap

$n_1 = n_2$	λ_2	$\widehat{\tau}_{1,n_1,n_2}$									$\widehat{\tau}_{2,n_{1,\kappa},n_{2,\kappa}}$					
		$p=1,\ q=1$			$p=2,\ q=1$			$p=2,\ q=2$			$p=3,\ q=2$			$p=3,\ q=3$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
50	2.00	0.093	0.042	0.013	0.148	0.088	0.031	0.270	0.194	0.101	0.113	0.068	0.017	0.173	0.107	0.046
	1.90	0.086	0.048	0.013	0.158	0.088	0.035	0.280	0.209	0.116	0.105	0.058	0.017	0.160	0.096	0.042
	1.80	0.122	0.068	0.018	0.175	0.123	0.052	0.303	0.232	0.129	0.121	0.071	0.023	0.179	0.113	0.039
	1.70	0.140	0.072	0.017	0.223	0.156	0.060	0.371	0.285	0.180	0.138	0.072	0.023	0.205	0.132	0.049
	1.60	0.202	0.118	0.037	0.290	0.195	0.081	0.444	0.355	0.223	0.176	0.113	0.036	0.239	0.168	0.061
	1.50	0.319	0.199	0.069	0.379	0.270	0.122	0.514	0.428	0.292	0.235	0.162	0.057	0.277	0.194	0.085
	1.40	0.398	0.274	0.127	0.484	0.384	0.203	0.576	0.493	0.347	0.298	0.192	0.079	0.317	0.217	0.098
	1.30	0.524	0.382	0.198	0.594	0.489	0.287	0.685	0.614	0.482	0.377	0.259	0.104	0.399	0.295	0.122
	1.20	0.615	0.503	0.286	0.710	0.616	0.398	0.764	0.702	0.555	0.486	0.346	0.170	0.483	0.359	0.165
	1.10	0.760	0.647	0.414	0.795	0.700	0.543	0.850	0.800	0.687	0.585	0.457	0.212	0.586	0.457	0.218
	1.00	0.849	0.773	0.537	0.883	0.822	0.685	0.893	0.859	0.763	0.693	0.569	0.299	0.664	0.515	0.264
	0.75	0.990	0.968	0.861	0.989	0.975	0.924	0.980	0.971	0.941	0.909	0.837	0.584	0.869	0.788	0.490
	0.50	1.000	0.999	0.997	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.981	0.895	0.983	0.936	0.761
100	2.00	0.054	0.020	0.008	0.119	0.069	0.020	0.199	0.137	0.045	0.033	0.013	0.002	0.046	0.025	0.008
	1.90	0.067	0.032	0.004	0.137	0.074	0.021	0.231	0.147	0.064	0.042	0.019	0.001	0.049	0.020	0.004
	1.80	0.119	0.057	0.014	0.175	0.106	0.043	0.243	0.176	0.083	0.059	0.031	0.006	0.070	0.037	0.008
	1.70	0.173	0.099	0.031	0.283	0.198	0.076	0.371	0.288	0.145	0.084	0.033	0.007	0.115	0.058	0.009
	1.60	0.301	0.186	0.067	0.410	0.290	0.145	0.464	0.372	0.231	0.153	0.077	0.020	0.144	0.085	0.025
	1.50	0.435	0.315	0.125	0.545	0.417	0.225	0.620	0.519	0.349	0.217	0.116	0.033	0.213	0.129	0.045
	1.40	0.603	0.471	0.230	0.703	0.595	0.354	0.730	0.654	0.480	0.296	0.178	0.058	0.310	0.185	0.056
	1.30	0.760	0.637	0.386	0.823	0.751	0.573	0.850	0.795	0.629	0.481	0.316	0.113	0.438	0.268	0.093
	1.20	0.888	0.808	0.598	0.920	0.866	0.727	0.922	0.873	0.766	0.626	0.461	0.208	0.531	0.391	0.151
	1.10	0.965	0.922	0.767	0.965	0.935	0.838	0.973	0.946	0.879	0.769	0.611	0.324	0.675	0.527	0.236
	1.00	0.991	0.973	0.896	0.995	0.983	0.952	0.994	0.987	0.959	0.898	0.799	0.475	0.800	0.667	0.325
	0.75	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.982	0.879	0.984	0.940	0.733
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996	1.000	0.999	0.955

Table 2: Rejection Rates for Two-Sample Test of Equivalent Means using Bootstrap (continued)

		$\widehat{\tau}_{1,n_1,n_2}$									$\widehat{\tau}_{2,n_{1,\kappa},n_{2,\kappa}}$					
		$p=1,\ q=1$			$p=2,\ q=1$			$p=2,\ q=2$			$p=3,\ q=2$			$p=3,\ q=3$		
$n_1 = n_2$	λ_2	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
200	2.00	0.046	0.020	0.002	0.089	0.034	0.009	0.163	0.091	0.036	0.007	0.001	0.000	0.017	0.006	0.001
	1.90	0.092	0.046	0.009	0.129	0.066	0.018	0.178	0.109	0.032	0.019	0.007	0.000	0.012	0.006	0.001
	1.80	0.170	0.094	0.029	0.215	0.136	0.045	0.299	0.206	0.090	0.034	0.017	0.002	0.023	0.010	0.001
	1.70	0.316	0.204	0.073	0.390	0.288	0.119	0.458	0.357	0.180	0.068	0.027	0.005	0.039	0.009	0.002
	1.60	0.522	0.350	0.157	0.614	0.478	0.240	0.643	0.549	0.358	0.133	0.055	0.009	0.093	0.045	0.007
	1.50	0.739	0.597	0.327	0.795	0.692	0.466	0.820	0.736	0.563	0.258	0.135	0.027	0.176	0.074	0.016
	1.40	0.913	0.841	0.606	0.914	0.857	0.694	0.938	0.897	0.770	0.453	0.273	0.083	0.325	0.153	0.038
	1.30	0.976	0.948	0.822	0.989	0.968	0.884	0.977	0.959	0.895	0.697	0.495	0.153	0.492	0.288	0.074
	1.20	0.997	0.988	0.940	0.993	0.990	0.962	0.991	0.984	0.964	0.850	0.714	0.330	0.680	0.487	0.173
	1.10	0.999	0.998	0.994	1.000	0.999	0.997	1.000	0.999	0.992	0.956	0.875	0.557	0.856	0.695	0.316
	1.00	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.965	0.800	0.955	0.863	0.514
	0.75	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	0.998	0.951
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
1000	2.00	0.039	0.008	0.001	0.061	0.023	0.004	0.077	0.032	0.003	0.000	0.000	0.000	0.000	0.000	0.000
	1.90	0.143	0.082	0.014	0.187	0.099	0.021	0.272	0.177	0.064	0.001	0.000	0.000	0.000	0.000	0.000
	1.80	0.517	0.344	0.131	0.624	0.471	0.249	0.672	0.518	0.303	0.007	0.000	0.000	0.000	0.000	0.000
	1.70	0.904	0.813	0.547	0.947	0.880	0.678	0.934	0.889	0.737	0.071	0.013	0.002	0.016	0.002	0.000
	1.60	0.995	0.989	0.933	0.996	0.985	0.946	0.998	0.990	0.966	0.335	0.110	0.007	0.102	0.025	0.005
	1.50	1.000	1.000	0.996	1.000	1.000	0.999	1.000	1.000	0.998	0.794	0.509	0.067	0.414	0.140	0.016
	1.40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.978	0.885	0.381	0.790	0.473	0.070
	1.30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.994	0.831	0.978	0.868	0.303
	1.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.992	0.999	0.990	0.714
	1.10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.958
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 2: Rejection Rates for Two-Sample Test of Equivalent Means using Bootstrap (continued)

		$\widehat{\tau}_{1,n_1,n_2}$									$\widehat{\tau}_{2,n_{1,\kappa},n_{2,\kappa}}$					
$n_1 = n_2$	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	0.040	0.011	0.000	0.044	0.012	0.000	0.061	0.022	0.002	0.000	0.000	0.000	0.000	0.000	0.000
	1.90	0.899	0.796	0.544	0.932	0.863	0.629	0.943	0.896	0.711	0.000	0.000	0.000	0.000	0.000	0.000
	1.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.131	0.013	0.000	0.001	0.000	0.000
	1.70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.992	0.822	0.060	0.189	0.013	0.000
	1.60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.954	0.979	0.710	0.012
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.670
	1.40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
20000	2.00	0.017	0.006	0.001	0.035	0.008	0.000	0.049	0.017	0.003	0.000	0.000	0.000	0.000	0.000	0.000
	1.90	0.999	0.993	0.947	0.999	0.994	0.960	0.999	0.997	0.983	0.000	0.000	0.000	0.000	0.000	0.000
	1.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.714	0.145	0.001	0.000	0.000	0.000
	1.70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.742	0.792	0.165	0.001
	1.60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.274
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998
	1.40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3: Rejection Rates for Returns to Scale Test using Asymptotic Normality

		$\widehat{\tau}_{3,n}$												$\widehat{\tau}_{4,n}$		
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	0.207	0.133	0.041	0.300	0.211	0.081	0.313	0.228	0.120	0.301	0.232	0.120	0.214	0.147	0.063
	0.1	0.227	0.157	0.062	0.325	0.231	0.088	0.301	0.226	0.111	0.319	0.234	0.125	0.216	0.155	0.067
	0.2	0.385	0.266	0.113	0.402	0.284	0.145	0.382	0.289	0.147	0.378	0.298	0.177	0.237	0.166	0.079
	0.3	0.592	0.440	0.196	0.608	0.475	0.253	0.508	0.403	0.225	0.522	0.432	0.251	0.306	0.240	0.106
	0.4	0.717	0.513	0.224	0.698	0.514	0.242	0.716	0.525	0.256	0.691	0.547	0.296	0.381	0.285	0.136
	0.6	0.827	0.648	0.300	0.817	0.660	0.328	0.831	0.666	0.364	0.826	0.658	0.384	0.423	0.326	0.153
	0.8	0.884	0.702	0.357	0.874	0.685	0.363	0.875	0.715	0.410	0.889	0.730	0.408	0.440	0.346	0.162
	1.0	0.891	0.739	0.399	0.905	0.727	0.394	0.919	0.740	0.411	0.906	0.776	0.470	0.479	0.388	0.190
	1.2	0.907	0.752	0.429	0.903	0.750	0.410	0.924	0.753	0.417	0.930	0.770	0.482	0.478	0.380	0.188
	1.4	0.911	0.779	0.477	0.923	0.767	0.454	0.935	0.783	0.468	0.948	0.818	0.524	0.495	0.401	0.199
100	0.0	0.168	0.103	0.031	0.253	0.170	0.060	0.276	0.199	0.094	0.286	0.217	0.102	0.152	0.091	0.032
	0.1	0.260	0.169	0.049	0.284	0.173	0.068	0.261	0.188	0.078	0.294	0.215	0.103	0.153	0.095	0.033
	0.2	0.446	0.326	0.143	0.432	0.327	0.128	0.398	0.296	0.148	0.350	0.255	0.134	0.194	0.120	0.036
	0.3	0.776	0.652	0.362	0.687	0.564	0.325	0.605	0.479	0.259	0.573	0.474	0.290	0.285	0.188	0.078
	0.4	0.847	0.670	0.324	0.829	0.658	0.343	0.826	0.657	0.346	0.827	0.653	0.375	0.380	0.276	0.112
	0.6	0.909	0.781	0.474	0.925	0.786	0.444	0.915	0.763	0.473	0.931	0.790	0.495	0.412	0.321	0.157
	0.8	0.936	0.817	0.531	0.945	0.806	0.493	0.935	0.816	0.509	0.952	0.826	0.535	0.427	0.337	0.171
	1.0	0.923	0.828	0.550	0.935	0.821	0.541	0.948	0.830	0.554	0.953	0.841	0.583	0.455	0.385	0.164
	1.2	0.926	0.813	0.562	0.930	0.815	0.539	0.943	0.828	0.547	0.947	0.839	0.573	0.466	0.376	0.164
	1.4	0.933	0.818	0.573	0.935	0.815	0.562	0.945	0.819	0.557	0.962	0.842	0.596	0.447	0.358	0.163

Table 3: Rejection Rates for Returns to Scale Test using Asymptotic Normality (continued)

		$\widehat{\tau}_{3,n}$												$\widehat{\tau}_{4,n}$		
		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
n	δ	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
200	0.0	0.141	0.069	0.015	0.207	0.133	0.037	0.256	0.167	0.074	0.264	0.189	0.086	0.167	0.103	0.028
	0.1	0.216	0.132	0.035	0.244	0.147	0.048	0.267	0.184	0.077	0.267	0.201	0.095	0.170	0.106	0.026
	0.2	0.539	0.393	0.185	0.467	0.343	0.164	0.448	0.338	0.175	0.390	0.290	0.162	0.222	0.133	0.039
	0.3	0.914	0.849	0.609	0.822	0.728	0.476	0.779	0.669	0.468	0.714	0.606	0.390	0.364	0.256	0.091
	0.4	0.929	0.811	0.516	0.939	0.827	0.529	0.940	0.831	0.533	0.934	0.821	0.555	0.393	0.300	0.134
	0.6	0.945	0.841	0.616	0.956	0.864	0.627	0.966	0.871	0.653	0.967	0.877	0.653	0.418	0.335	0.184
	0.8	0.945	0.853	0.665	0.959	0.852	0.655	0.968	0.869	0.668	0.965	0.883	0.675	0.437	0.349	0.198
	1.0	0.943	0.838	0.624	0.948	0.839	0.626	0.957	0.847	0.638	0.960	0.860	0.648	0.430	0.338	0.172
	1.2	0.938	0.862	0.651	0.944	0.857	0.650	0.952	0.860	0.665	0.958	0.877	0.671	0.448	0.362	0.196
	1.4	0.952	0.847	0.629	0.948	0.857	0.639	0.963	0.868	0.636	0.964	0.872	0.648	0.417	0.337	0.174
1000	0.0	0.105	0.060	0.010	0.164	0.089	0.017	0.213	0.125	0.037	0.223	0.151	0.062	0.137	0.074	0.016
	0.1	0.274	0.172	0.053	0.270	0.171	0.050	0.284	0.196	0.071	0.295	0.193	0.065	0.142	0.075	0.014
	0.2	0.925	0.855	0.665	0.806	0.690	0.475	0.716	0.590	0.365	0.607	0.484	0.294	0.212	0.123	0.026
	0.3	1.000	1.000	1.000	1.000	1.000	0.997	0.996	0.994	0.975	0.998	0.989	0.942	0.523	0.388	0.176
	0.4	0.949	0.892	0.750	0.963	0.905	0.759	0.967	0.910	0.772	0.978	0.923	0.789	0.429	0.326	0.202
	0.6	0.957	0.900	0.764	0.961	0.907	0.772	0.972	0.907	0.773	0.977	0.920	0.785	0.382	0.303	0.190
	0.8	0.958	0.906	0.779	0.969	0.910	0.779	0.968	0.906	0.793	0.972	0.913	0.792	0.402	0.318	0.210
	1.0	0.955	0.899	0.768	0.955	0.905	0.766	0.960	0.900	0.760	0.959	0.904	0.761	0.382	0.298	0.198
	1.2	0.958	0.905	0.764	0.965	0.907	0.760	0.969	0.909	0.788	0.976	0.913	0.770	0.396	0.325	0.211
	1.4	0.953	0.904	0.791	0.959	0.903	0.772	0.959	0.894	0.772	0.954	0.905	0.766	0.404	0.319	0.203

Table 3: Rejection Rates for Returns to Scale Test using Asymptotic Normality (continued)

n	δ	$\widehat{\tau}_{3,n}$												$\widehat{\tau}_{4,n}$		
		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
10000	0.0	0.113	0.061	0.013	0.151	0.078	0.009	0.155	0.087	0.022	0.178	0.107	0.029	0.117	0.055	0.015
	0.1	0.786	0.675	0.394	0.626	0.484	0.235	0.513	0.378	0.159	0.419	0.306	0.139	0.148	0.079	0.022
	0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.991	0.434	0.298	0.123
	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	0.956	0.914	0.760
	0.4	0.964	0.908	0.819	0.965	0.908	0.812	0.976	0.923	0.845	0.982	0.930	0.845	0.345	0.257	0.140
	0.6	0.974	0.921	0.823	0.970	0.919	0.827	0.976	0.930	0.841	0.984	0.941	0.843	0.343	0.240	0.138
	0.8	0.961	0.916	0.816	0.960	0.916	0.813	0.965	0.919	0.823	0.966	0.919	0.822	0.317	0.242	0.138
	1.0	0.970	0.933	0.841	0.975	0.929	0.844	0.975	0.931	0.850	0.974	0.937	0.850	0.340	0.243	0.141
	1.2	0.967	0.924	0.834	0.970	0.921	0.831	0.969	0.922	0.842	0.972	0.919	0.829	0.333	0.233	0.140
	1.4	0.972	0.914	0.832	0.969	0.918	0.835	0.970	0.924	0.838	0.973	0.924	0.834	0.317	0.221	0.134
20000	0.0	0.103	0.049	0.010	0.106	0.057	0.014	0.160	0.079	0.024	0.176	0.096	0.026	0.125	0.065	0.021
	0.1	0.960	0.926	0.755	0.835	0.720	0.447	0.683	0.538	0.273	0.585	0.442	0.212	0.176	0.091	0.029
	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	0.547	0.392	0.178
	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	0.948
	0.4	0.974	0.922	0.833	0.975	0.922	0.837	0.977	0.930	0.838	0.982	0.952	0.859	0.324	0.243	0.129
	0.6	0.975	0.945	0.870	0.977	0.945	0.872	0.980	0.945	0.877	0.985	0.955	0.879	0.331	0.233	0.143
	0.8	0.970	0.920	0.834	0.970	0.930	0.841	0.976	0.926	0.854	0.976	0.934	0.847	0.284	0.218	0.115
	1.0	0.972	0.933	0.846	0.969	0.934	0.847	0.971	0.934	0.865	0.971	0.938	0.851	0.319	0.234	0.133
	1.2	0.966	0.935	0.853	0.970	0.930	0.851	0.972	0.927	0.858	0.972	0.933	0.859	0.316	0.232	0.137
	1.4	0.976	0.940	0.848	0.976	0.924	0.844	0.979	0.922	0.834	0.977	0.944	0.840	0.312	0.231	0.130

Table 4: Rejection Rates for Returns to Scale Test using Bootstrap

		$\widehat{\tau}_{3,n}$												$\widehat{\tau}_{4,n}$		
		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
n	δ	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
50	0.0	0.113	0.050	0.008	0.188	0.100	0.034	0.230	0.157	0.062	0.243	0.167	0.085	0.156	0.078	0.016
	0.1	0.133	0.063	0.011	0.216	0.106	0.035	0.224	0.153	0.059	0.248	0.164	0.085	0.157	0.073	0.020
	0.2	0.242	0.127	0.026	0.262	0.171	0.058	0.293	0.201	0.069	0.303	0.228	0.125	0.182	0.096	0.022
	0.3	0.429	0.267	0.071	0.472	0.327	0.113	0.423	0.301	0.136	0.448	0.348	0.175	0.251	0.135	0.037
	0.4	0.664	0.354	0.089	0.664	0.399	0.149	0.683	0.472	0.200	0.667	0.516	0.268	0.521	0.352	0.151
	0.6	0.798	0.536	0.160	0.824	0.561	0.229	0.827	0.633	0.307	0.825	0.659	0.378	0.678	0.474	0.243
	0.8	0.890	0.605	0.212	0.888	0.616	0.276	0.897	0.701	0.389	0.894	0.737	0.427	0.795	0.578	0.307
	1.0	0.930	0.673	0.276	0.938	0.687	0.317	0.934	0.737	0.376	0.936	0.801	0.511	0.853	0.630	0.321
	1.2	0.951	0.704	0.295	0.953	0.720	0.339	0.962	0.757	0.417	0.961	0.807	0.534	0.875	0.678	0.377
	1.4	0.969	0.731	0.369	0.973	0.742	0.389	0.982	0.793	0.458	0.981	0.840	0.568	0.895	0.708	0.399
100	0.0	0.090	0.038	0.002	0.154	0.080	0.013	0.202	0.120	0.043	0.219	0.138	0.057	0.064	0.022	0.003
	0.1	0.144	0.063	0.008	0.146	0.080	0.020	0.176	0.109	0.029	0.224	0.147	0.060	0.073	0.023	0.003
	0.2	0.274	0.165	0.037	0.286	0.158	0.044	0.295	0.184	0.070	0.266	0.180	0.082	0.086	0.033	0.003
	0.3	0.641	0.471	0.171	0.558	0.409	0.169	0.508	0.358	0.161	0.499	0.378	0.204	0.208	0.106	0.015
	0.4	0.850	0.579	0.203	0.838	0.594	0.260	0.835	0.629	0.308	0.826	0.650	0.379	0.604	0.444	0.210
	0.6	0.942	0.726	0.389	0.948	0.751	0.387	0.943	0.778	0.470	0.944	0.821	0.554	0.764	0.591	0.285
	0.8	0.979	0.797	0.466	0.980	0.795	0.469	0.984	0.846	0.550	0.985	0.872	0.627	0.855	0.691	0.352
	1.0	0.984	0.813	0.507	0.994	0.826	0.524	0.993	0.862	0.588	0.997	0.891	0.678	0.926	0.771	0.402
	1.2	0.981	0.808	0.530	0.997	0.821	0.544	0.993	0.861	0.612	0.997	0.890	0.669	0.933	0.811	0.425
	1.4	0.987	0.812	0.561	0.998	0.821	0.578	0.997	0.855	0.613	0.999	0.892	0.692	0.949	0.824	0.454

Table 4: Rejection Rates for Returns to Scale Test using Bootstrap (continued)

n	δ	$\widehat{\tau}_{3,n}$												$\widehat{\tau}_{4,n}$		
		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
200	0.0	0.050	0.017	0.000	0.119	0.058	0.004	0.170	0.089	0.032	0.189	0.125	0.038	0.038	0.013	0.000
	0.1	0.098	0.041	0.006	0.126	0.066	0.010	0.182	0.104	0.028	0.206	0.132	0.051	0.028	0.012	0.000
	0.2	0.340	0.202	0.043	0.319	0.192	0.048	0.334	0.217	0.097	0.297	0.208	0.088	0.061	0.019	0.002
	0.3	0.835	0.715	0.359	0.719	0.562	0.281	0.675	0.548	0.298	0.635	0.497	0.273	0.197	0.080	0.014
	0.4	0.965	0.795	0.441	0.966	0.810	0.494	0.965	0.835	0.567	0.952	0.854	0.617	0.684	0.517	0.282
	0.6	0.994	0.835	0.609	0.997	0.879	0.652	0.997	0.899	0.711	0.994	0.922	0.740	0.879	0.723	0.404
	0.8	0.985	0.856	0.664	0.998	0.876	0.693	1.000	0.910	0.741	1.000	0.927	0.770	0.937	0.800	0.496
	1.0	0.992	0.836	0.630	0.998	0.865	0.675	1.000	0.881	0.707	1.000	0.919	0.761	0.939	0.818	0.500
	1.2	0.987	0.860	0.673	0.998	0.870	0.684	1.000	0.898	0.751	1.000	0.922	0.780	0.956	0.842	0.516
	1.4	0.990	0.847	0.638	0.997	0.878	0.685	1.000	0.907	0.722	1.000	0.921	0.775	0.940	0.820	0.516
1000	0.0	0.044	0.011	0.001	0.070	0.020	0.001	0.121	0.054	0.007	0.161	0.081	0.025	0.000	0.000	0.000
	0.1	0.133	0.054	0.004	0.145	0.063	0.007	0.190	0.106	0.019	0.204	0.100	0.029	0.003	0.000	0.000
	0.2	0.815	0.672	0.303	0.655	0.503	0.178	0.586	0.429	0.168	0.495	0.369	0.159	0.016	0.001	0.000
	0.3	1.000	1.000	0.996	1.000	0.999	0.949	0.994	0.983	0.918	0.989	0.967	0.869	0.409	0.158	0.016
	0.4	0.991	0.893	0.780	0.995	0.918	0.800	1.000	0.939	0.839	1.000	0.952	0.874	0.912	0.776	0.506
	0.6	0.994	0.904	0.793	1.000	0.920	0.821	1.000	0.935	0.845	0.999	0.949	0.866	0.933	0.840	0.613
	0.8	0.993	0.909	0.809	0.998	0.919	0.826	1.000	0.934	0.853	1.000	0.946	0.861	0.932	0.852	0.623
	1.0	0.988	0.898	0.793	0.995	0.920	0.811	1.000	0.923	0.820	1.000	0.935	0.849	0.928	0.852	0.640
	1.2	0.993	0.911	0.788	1.000	0.918	0.809	1.000	0.931	0.842	1.000	0.947	0.848	0.931	0.840	0.600
	1.4	0.992	0.908	0.813	0.992	0.911	0.820	1.000	0.926	0.830	1.000	0.931	0.856	0.931	0.834	0.614

Table 4: Rejection Rates for Returns to Scale Test using Bootstrap (continued)

		$\widehat{\tau}_{3,n}$												$\widehat{\tau}_{4,n}$		
		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
n	δ	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
10000	0.0	0.039	0.015	0.000	0.059	0.011	0.000	0.079	0.035	0.004	0.116	0.051	0.008	0.000	0.000	0.000
	0.1	0.611	0.396	0.111	0.446	0.272	0.066	0.367	0.214	0.037	0.316	0.193	0.051	0.000	0.000	0.000
	0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.989	0.998	0.995	0.972	0.059	0.002	0.000
	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	1.000	1.000	0.882
	0.4	0.991	0.911	0.837	0.995	0.914	0.845	1.000	0.944	0.880	1.000	0.958	0.899	0.891	0.802	0.598
	0.6	0.991	0.924	0.847	0.998	0.929	0.864	1.000	0.947	0.891	1.000	0.963	0.907	0.912	0.816	0.606
	0.8	0.988	0.915	0.841	0.996	0.926	0.850	1.000	0.938	0.873	1.000	0.948	0.884	0.907	0.811	0.624
	1.0	0.989	0.937	0.866	0.994	0.936	0.875	1.000	0.952	0.896	1.000	0.957	0.901	0.901	0.807	0.588
	1.2	0.991	0.924	0.860	0.996	0.933	0.863	1.000	0.940	0.887	1.000	0.949	0.886	0.903	0.805	0.611
	1.4	0.993	0.919	0.849	0.997	0.930	0.865	1.000	0.942	0.880	1.000	0.946	0.890	0.889	0.780	0.584
20000	0.0	0.035	0.010	0.000	0.049	0.017	0.000	0.074	0.031	0.004	0.102	0.050	0.005	0.000	0.000	0.000
	0.1	0.898	0.756	0.396	0.684	0.475	0.166	0.521	0.343	0.107	0.468	0.283	0.089	0.000	0.000	0.000
	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	0.230	0.019	0.000
	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	1.000	1.000	1.000
	0.4	0.988	0.923	0.847	0.996	0.928	0.858	0.999	0.958	0.881	0.997	0.969	0.912	0.886	0.801	0.617
	0.6	0.997	0.949	0.888	0.999	0.952	0.895	1.000	0.960	0.909	1.000	0.970	0.924	0.899	0.797	0.611
	0.8	0.994	0.922	0.849	0.999	0.936	0.869	1.000	0.945	0.884	1.000	0.962	0.896	0.879	0.805	0.609
	1.0	0.988	0.937	0.864	0.997	0.943	0.870	1.000	0.949	0.896	1.000	0.956	0.902	0.891	0.793	0.584
	1.2	0.994	0.938	0.871	0.997	0.938	0.871	1.000	0.950	0.894	1.000	0.961	0.901	0.897	0.808	0.607
	1.4	0.992	0.943	0.861	0.997	0.940	0.867	1.000	0.948	0.885	1.000	0.958	0.897	0.882	0.803	0.591

Table 5: Rejection Rates for Convexity Test using Asymptotic Normality

n	δ	$\widehat{\tau}_{5,n}$						$\widehat{\tau}_{6,n}$								
		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
50	1.4*	0.125	0.066	0.026	0.093	0.061	0.022	0.067	0.047	0.014	0.022	0.013	0.004	0.048	0.029	0.009
	0.0	0.329	0.242	0.123	0.314	0.248	0.137	0.154	0.108	0.037	0.077	0.052	0.024	0.120	0.083	0.040
	0.1	0.372	0.266	0.135	0.339	0.259	0.153	0.172	0.110	0.048	0.102	0.067	0.026	0.112	0.078	0.046
	0.2	0.419	0.320	0.174	0.361	0.292	0.178	0.180	0.108	0.046	0.109	0.072	0.032	0.143	0.103	0.047
	0.3	0.530	0.437	0.273	0.412	0.334	0.221	0.214	0.148	0.060	0.138	0.093	0.040	0.176	0.112	0.056
	0.4	0.728	0.636	0.459	0.534	0.454	0.315	0.248	0.171	0.076	0.162	0.108	0.045	0.183	0.122	0.062
	0.6	0.893	0.858	0.747	0.726	0.658	0.528	0.373	0.301	0.149	0.237	0.176	0.089	0.259	0.183	0.101
	0.8	0.964	0.948	0.897	0.837	0.784	0.674	0.492	0.404	0.260	0.352	0.257	0.143	0.363	0.282	0.157
	1.0	0.988	0.981	0.955	0.913	0.890	0.815	0.652	0.569	0.386	0.477	0.378	0.223	0.481	0.386	0.250
	1.2	0.996	0.993	0.974	0.948	0.928	0.885	0.730	0.654	0.496	0.590	0.487	0.300	0.618	0.525	0.374
	1.4	0.999	0.997	0.993	0.972	0.960	0.924	0.783	0.710	0.533	0.697	0.601	0.438	0.658	0.586	0.413
100	1.4*	0.104	0.056	0.019	0.110	0.066	0.017	0.075	0.048	0.011	0.041	0.019	0.002	0.044	0.026	0.005
	0.0	0.282	0.182	0.070	0.364	0.267	0.142	0.163	0.105	0.035	0.094	0.062	0.020	0.086	0.053	0.017
	0.1	0.308	0.210	0.092	0.368	0.279	0.157	0.158	0.101	0.035	0.090	0.059	0.018	0.083	0.051	0.018
	0.2	0.422	0.315	0.157	0.409	0.311	0.191	0.175	0.113	0.048	0.091	0.061	0.021	0.099	0.058	0.023
	0.3	0.611	0.512	0.335	0.521	0.449	0.280	0.204	0.134	0.053	0.108	0.057	0.024	0.115	0.067	0.022
	0.4	0.821	0.756	0.596	0.661	0.573	0.437	0.259	0.185	0.085	0.120	0.087	0.033	0.124	0.080	0.036
	0.6	0.961	0.946	0.901	0.879	0.846	0.745	0.470	0.359	0.199	0.234	0.149	0.078	0.240	0.163	0.067
	0.8	0.992	0.989	0.975	0.957	0.938	0.889	0.629	0.535	0.347	0.391	0.279	0.148	0.395	0.299	0.167
	1.0	0.999	0.999	0.996	0.985	0.978	0.952	0.797	0.712	0.513	0.546	0.442	0.243	0.554	0.447	0.275
	1.2	1.000	1.000	0.997	0.998	0.996	0.990	0.856	0.782	0.629	0.660	0.544	0.364	0.664	0.560	0.369
	1.4	1.000	1.000	1.000	0.998	0.995	0.985	0.890	0.848	0.719	0.735	0.630	0.428	0.735	0.641	0.466

*The production set Ψ is strictly convex.

Table 5: Rejection Rates for Convexity Test using Asymptotic Normality (continued)

n	δ	$\widehat{\tau}_{5,n}$						$\widehat{\tau}_{6,n}$								
		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
200	1.4*	0.113	0.063	0.020	0.114	0.064	0.024	0.053	0.024	0.003	0.054	0.030	0.010	0.042	0.027	0.006
	0.0	0.255	0.165	0.062	0.338	0.261	0.136	0.141	0.079	0.024	0.102	0.060	0.016	0.070	0.044	0.013
	0.1	0.271	0.181	0.066	0.354	0.274	0.139	0.140	0.097	0.034	0.111	0.066	0.020	0.073	0.046	0.008
	0.2	0.430	0.312	0.150	0.426	0.332	0.201	0.171	0.114	0.032	0.125	0.075	0.027	0.087	0.055	0.015
	0.3	0.733	0.643	0.459	0.609	0.516	0.341	0.219	0.140	0.050	0.138	0.088	0.024	0.097	0.063	0.022
	0.4	0.955	0.923	0.848	0.837	0.764	0.607	0.305	0.222	0.113	0.204	0.134	0.051	0.137	0.082	0.019
	0.6	0.997	0.994	0.987	0.978	0.960	0.915	0.568	0.450	0.254	0.379	0.282	0.144	0.249	0.170	0.085
	0.8	1.000	1.000	1.000	0.999	0.999	0.994	0.796	0.702	0.486	0.597	0.470	0.294	0.452	0.327	0.178
	1.0	1.000	1.000	1.000	0.999	0.998	0.995	0.894	0.834	0.643	0.758	0.656	0.455	0.633	0.512	0.297
	1.2	1.000	1.000	1.000	1.000	1.000	1.000	0.962	0.922	0.796	0.838	0.757	0.596	0.740	0.626	0.393
	1.4	1.000	1.000	1.000	1.000	1.000	1.000	0.962	0.934	0.824	0.901	0.840	0.667	0.814	0.720	0.523
1000	1.4*	0.104	0.043	0.010	0.131	0.080	0.027	0.080	0.045	0.011	0.069	0.046	0.016	0.075	0.040	0.015
	0.0	0.176	0.099	0.024	0.352	0.245	0.104	0.147	0.087	0.022	0.101	0.055	0.022	0.093	0.055	0.017
	0.1	0.272	0.170	0.036	0.381	0.272	0.142	0.136	0.080	0.019	0.111	0.070	0.021	0.091	0.056	0.020
	0.2	0.686	0.580	0.344	0.589	0.461	0.275	0.182	0.104	0.028	0.127	0.059	0.021	0.098	0.062	0.021
	0.3	0.991	0.977	0.924	0.919	0.874	0.722	0.306	0.192	0.069	0.200	0.121	0.043	0.120	0.078	0.030
	0.4	1.000	1.000	1.000	0.996	0.993	0.980	0.510	0.372	0.179	0.338	0.221	0.091	0.219	0.131	0.057
	0.6	1.000	1.000	1.000	1.000	1.000	1.000	0.860	0.745	0.520	0.684	0.549	0.313	0.451	0.327	0.156
	0.8	1.000	1.000	1.000	1.000	1.000	1.000	0.960	0.928	0.791	0.892	0.804	0.561	0.719	0.588	0.364
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	0.991	0.976	0.922	0.966	0.931	0.767	0.875	0.769	0.531
	1.2	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.993	0.970	0.984	0.963	0.877	0.958	0.890	0.711
	1.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.985	0.991	0.980	0.921	0.979	0.937	0.795

*The production set Ψ is strictly convex.

Table 5: Rejection Rates for Convexity Test using Asymptotic Normality (continued)

		$\widehat{\tau}_{5,n}$						$\widehat{\tau}_{6,n}$								
		$p = 1$			$p = 2$			$p = 3$			$p = 4$			$p = 5$		
n	δ	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
10000	1.4*	0.108	0.055	0.010	0.151	0.079	0.018	0.103	0.059	0.016	0.099	0.055	0.018	0.092	0.058	0.014
	0.0	0.129	0.066	0.018	0.279	0.183	0.057	0.144	0.081	0.034	0.140	0.079	0.017	0.112	0.069	0.022
	0.1	0.450	0.314	0.122	0.429	0.269	0.114	0.139	0.086	0.032	0.140	0.074	0.019	0.123	0.068	0.013
	0.2	0.997	0.994	0.976	0.937	0.887	0.745	0.211	0.127	0.048	0.176	0.113	0.033	0.123	0.070	0.024
	0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.475	0.353	0.150	0.318	0.206	0.069	0.200	0.126	0.037
	0.4	1.000	1.000	1.000	1.000	1.000	1.000	0.821	0.705	0.444	0.598	0.448	0.237	0.402	0.281	0.117
	0.6	1.000	1.000	1.000	1.000	1.000	1.000	0.992	0.988	0.909	0.928	0.857	0.645	0.770	0.623	0.343
	0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.992	0.992	0.982	0.907	0.950	0.885	0.662
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.978	0.990	0.971	0.869
	1.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.999	0.995	0.961
	1.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.980
20000	1.4*	0.097	0.048	0.007	0.163	0.081	0.017	0.098	0.048	0.017	0.109	0.066	0.022	0.088	0.055	0.017
	0.0	0.135	0.058	0.009	0.292	0.186	0.053	0.136	0.077	0.024	0.121	0.082	0.029	0.102	0.065	0.023
	0.1	0.587	0.463	0.220	0.476	0.341	0.127	0.151	0.099	0.029	0.123	0.080	0.032	0.111	0.063	0.024
	0.2	1.000	1.000	1.000	0.988	0.972	0.911	0.254	0.145	0.047	0.172	0.109	0.045	0.143	0.087	0.032
	0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.529	0.390	0.168	0.330	0.226	0.093	0.228	0.147	0.057
	0.4	1.000	1.000	1.000	1.000	1.000	1.000	0.880	0.785	0.547	0.656	0.498	0.242	0.452	0.315	0.142
	0.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.959	0.955	0.903	0.682	0.833	0.704	0.418
	0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.994	0.934	0.974	0.933	0.747
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.985	0.998	0.985	0.916
	1.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	0.996	0.972
	1.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996

Table 6: Rejection Rates for Convexity Test using Bootstrap

n	δ	$\widehat{\tau}_{5,n}$						$\widehat{\tau}_{6,n}$								
		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
50	1.4*	0.063	0.035	0.010	0.065	0.034	0.009	0.037	0.024	0.014	0.043	0.025	0.018	0.050	0.039	0.028
	0.0	0.235	0.141	0.046	0.264	0.187	0.071	0.140	0.104	0.062	0.138	0.107	0.088	0.152	0.137	0.108
	0.1	0.258	0.168	0.060	0.276	0.207	0.087	0.145	0.099	0.060	0.136	0.099	0.086	0.130	0.110	0.085
	0.2	0.308	0.210	0.067	0.304	0.235	0.096	0.162	0.114	0.067	0.134	0.105	0.092	0.155	0.131	0.114
	0.3	0.424	0.312	0.140	0.347	0.280	0.124	0.174	0.124	0.067	0.162	0.124	0.092	0.156	0.135	0.118
	0.4	0.631	0.497	0.289	0.473	0.384	0.230	0.210	0.153	0.082	0.175	0.138	0.119	0.174	0.150	0.125
	0.6	0.848	0.768	0.596	0.671	0.593	0.409	0.298	0.206	0.080	0.244	0.185	0.158	0.216	0.172	0.142
	0.8	0.944	0.913	0.798	0.791	0.718	0.572	0.438	0.326	0.110	0.303	0.209	0.173	0.304	0.246	0.202
	1.0	0.977	0.963	0.914	0.888	0.848	0.729	0.570	0.438	0.137	0.403	0.283	0.237	0.351	0.248	0.197
	1.2	0.990	0.983	0.937	0.931	0.908	0.824	0.646	0.513	0.159	0.421	0.316	0.253	0.415	0.300	0.248
	1.4	0.995	0.994	0.980	0.961	0.943	0.884	0.693	0.562	0.193	0.407	0.293	0.248	0.464	0.320	0.264
100	1.4*	0.054	0.027	0.005	0.075	0.035	0.006	0.019	0.009	0.005	0.029	0.014	0.006	0.034	0.016	0.011
	0.0	0.180	0.090	0.022	0.287	0.203	0.088	0.095	0.045	0.020	0.137	0.102	0.071	0.102	0.067	0.042
	0.1	0.216	0.120	0.038	0.309	0.221	0.094	0.099	0.042	0.018	0.108	0.070	0.045	0.090	0.056	0.037
	0.2	0.317	0.201	0.059	0.343	0.256	0.119	0.105	0.051	0.017	0.129	0.081	0.064	0.107	0.075	0.051
	0.3	0.507	0.392	0.200	0.468	0.351	0.185	0.153	0.073	0.023	0.152	0.104	0.070	0.119	0.083	0.061
	0.4	0.754	0.651	0.452	0.594	0.508	0.339	0.210	0.117	0.034	0.191	0.125	0.087	0.147	0.102	0.079
	0.6	0.947	0.917	0.817	0.855	0.805	0.640	0.456	0.263	0.091	0.284	0.198	0.144	0.255	0.160	0.117
	0.8	0.989	0.981	0.950	0.942	0.922	0.834	0.607	0.394	0.131	0.400	0.274	0.201	0.353	0.238	0.168
	1.0	0.999	0.996	0.988	0.981	0.969	0.937	0.793	0.579	0.235	0.494	0.352	0.273	0.437	0.294	0.238
	1.2	1.000	0.997	0.997	0.996	0.993	0.983	0.846	0.659	0.289	0.533	0.408	0.319	0.515	0.352	0.260
	1.4	1.000	1.000	0.997	0.996	0.992	0.979	0.899	0.684	0.316	0.561	0.405	0.325	0.546	0.385	0.303

*The production set Ψ is strictly convex.

Table 6: Rejection Rates for Convexity Test using Bootstrap (continued)

		$\widehat{\tau}_{5,n}$						$\widehat{\tau}_{6,n}$								
		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
n	δ	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
200	1.4*	0.064	0.026	0.005	0.081	0.036	0.013	0.007	0.004	0.001	0.030	0.009	0.002	0.027	0.009	0.005
	0.0	0.159	0.091	0.019	0.283	0.184	0.086	0.063	0.021	0.008	0.064	0.036	0.018	0.096	0.058	0.034
	0.1	0.181	0.099	0.026	0.299	0.199	0.082	0.075	0.026	0.005	0.070	0.035	0.014	0.093	0.057	0.035
	0.2	0.312	0.197	0.072	0.353	0.272	0.141	0.094	0.037	0.011	0.087	0.044	0.021	0.105	0.073	0.046
	0.3	0.642	0.528	0.291	0.540	0.434	0.254	0.132	0.049	0.013	0.108	0.052	0.024	0.128	0.078	0.048
	0.4	0.925	0.879	0.735	0.787	0.689	0.525	0.244	0.113	0.024	0.165	0.106	0.041	0.189	0.115	0.076
	0.6	0.994	0.991	0.971	0.965	0.939	0.890	0.551	0.335	0.087	0.378	0.221	0.083	0.322	0.209	0.137
	0.8	1.000	1.000	0.994	0.999	0.998	0.976	0.823	0.638	0.225	0.593	0.452	0.161	0.478	0.336	0.230
	1.0	1.000	1.000	1.000	0.999	0.996	0.992	0.917	0.792	0.396	0.759	0.602	0.200	0.571	0.405	0.308
	1.2	1.000	1.000	1.000	1.000	1.000	0.999	0.946	0.843	0.488	0.798	0.681	0.228	0.619	0.463	0.345
	1.4	1.000	1.000	1.000	1.000	1.000	1.000	0.967	0.875	0.531	0.814	0.684	0.238	0.604	0.505	0.438
1000	1.4*	0.043	0.014	0.000	0.088	0.049	0.008	0.008	0.001	0.000	0.013	0.005	0.000	0.019	0.005	0.000
	0.0	0.097	0.042	0.005	0.273	0.174	0.051	0.031	0.009	0.000	0.037	0.014	0.006	0.050	0.023	0.005
	0.1	0.167	0.073	0.010	0.301	0.196	0.070	0.040	0.010	0.001	0.043	0.018	0.003	0.049	0.022	0.004
	0.2	0.579	0.415	0.183	0.490	0.374	0.187	0.063	0.013	0.002	0.055	0.017	0.005	0.080	0.035	0.007
	0.3	0.977	0.946	0.858	0.884	0.797	0.614	0.177	0.061	0.006	0.109	0.033	0.008	0.123	0.053	0.016
	0.4	1.000	1.000	1.000	0.995	0.988	0.967	0.526	0.267	0.039	0.347	0.146	0.031	0.239	0.123	0.038
	0.6	1.000	1.000	1.000	1.000	1.000	0.999	0.919	0.782	0.341	0.776	0.514	0.163	0.563	0.365	0.146
	0.8	1.000	1.000	1.000	1.000	1.000	1.000	0.988	0.958	0.689	0.938	0.792	0.360	0.774	0.621	0.239
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.992	0.829	0.977	0.892	0.522	0.837	0.726	0.274
	1.2	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.992	0.915	0.988	0.919	0.611	0.871	0.789	0.298
	1.4	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.922	0.989	0.913	0.655	0.884	0.812	0.270

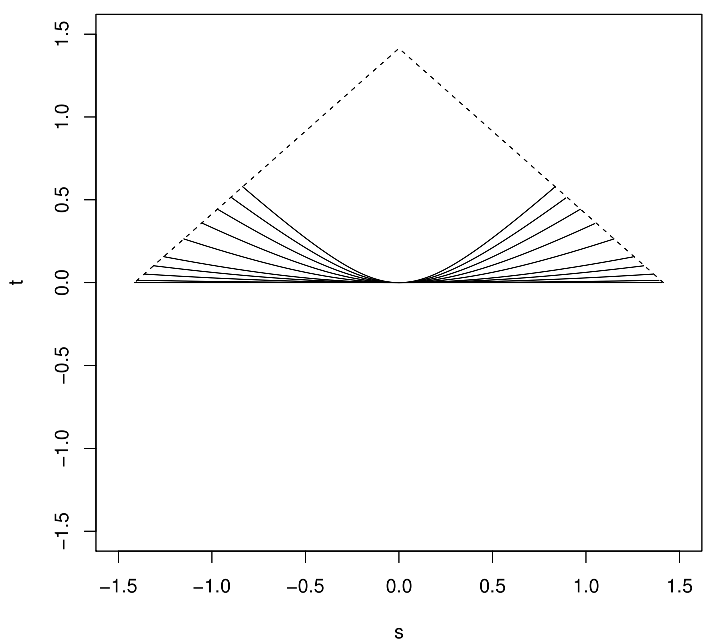
*The production set Ψ is strictly convex.

Table 6: Rejection Rates for Convexity Test using Bootstrap (continued)

n	δ	$\widehat{\tau}_{5,n}$						$\widehat{\tau}_{6,n}$								
		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
10000	1.4*	0.048	0.015	0.002	0.083	0.035	0.001	0.004	0.001	0.000	0.008	0.001	0.000	0.016	0.005	0.000
	0.0	0.053	0.024	0.004	0.194	0.103	0.023	0.021	0.007	0.000	0.026	0.008	0.001	0.047	0.010	0.002
	0.1	0.295	0.156	0.034	0.289	0.162	0.044	0.015	0.007	0.000	0.021	0.002	0.001	0.055	0.013	0.001
	0.2	0.990	0.978	0.926	0.893	0.826	0.606	0.057	0.022	0.000	0.055	0.015	0.002	0.064	0.021	0.002
	0.3	1.000	1.000	1.000	1.000	1.000	0.999	0.466	0.156	0.013	0.231	0.081	0.008	0.159	0.066	0.009
	0.4	1.000	1.000	1.000	1.000	1.000	1.000	0.977	0.825	0.277	0.744	0.425	0.063	0.473	0.218	0.045
	0.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.971	0.994	0.962	0.601	0.911	0.707	0.221
	0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	0.995	0.884	0.968	0.888	0.453
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.997	0.930	0.987	0.898	0.597
	1.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	0.999	0.964	0.979	0.905	0.660
	1.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.959	0.984	0.912	0.683
20000	1.4*	0.041	0.015	0.001	0.090	0.026	0.005	0.002	0.001	0.000	0.009	0.005	0.000	0.010	0.001	0.000
	0.0	0.052	0.015	0.000	0.190	0.087	0.017	0.004	0.000	0.000	0.029	0.009	0.000	0.032	0.008	0.000
	0.1	0.442	0.262	0.066	0.352	0.202	0.058	0.009	0.003	0.000	0.028	0.009	0.000	0.036	0.014	0.000
	0.2	1.000	1.000	0.997	0.973	0.945	0.814	0.040	0.008	0.000	0.058	0.016	0.001	0.061	0.021	0.000
	0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.555	0.227	0.015	0.240	0.090	0.008	0.178	0.065	0.012
	0.4	1.000	1.000	1.000	1.000	1.000	1.000	0.994	0.952	0.465	0.840	0.515	0.083	0.595	0.284	0.055
	0.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.986	0.674	0.958	0.820	0.292
	0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.919	0.982	0.946	0.625
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.962	0.989	0.946	0.706
	1.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.974	0.987	0.949	0.751
	1.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.988	0.992	0.964	0.796

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

(s,t)-space



(x,y)-space

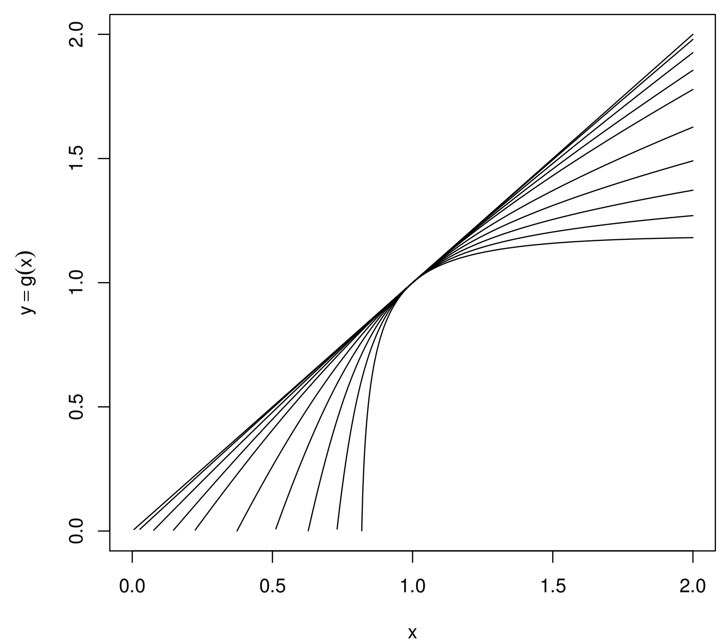


Figure 2: Function $g(\cdot)$ for Tests of Convexity

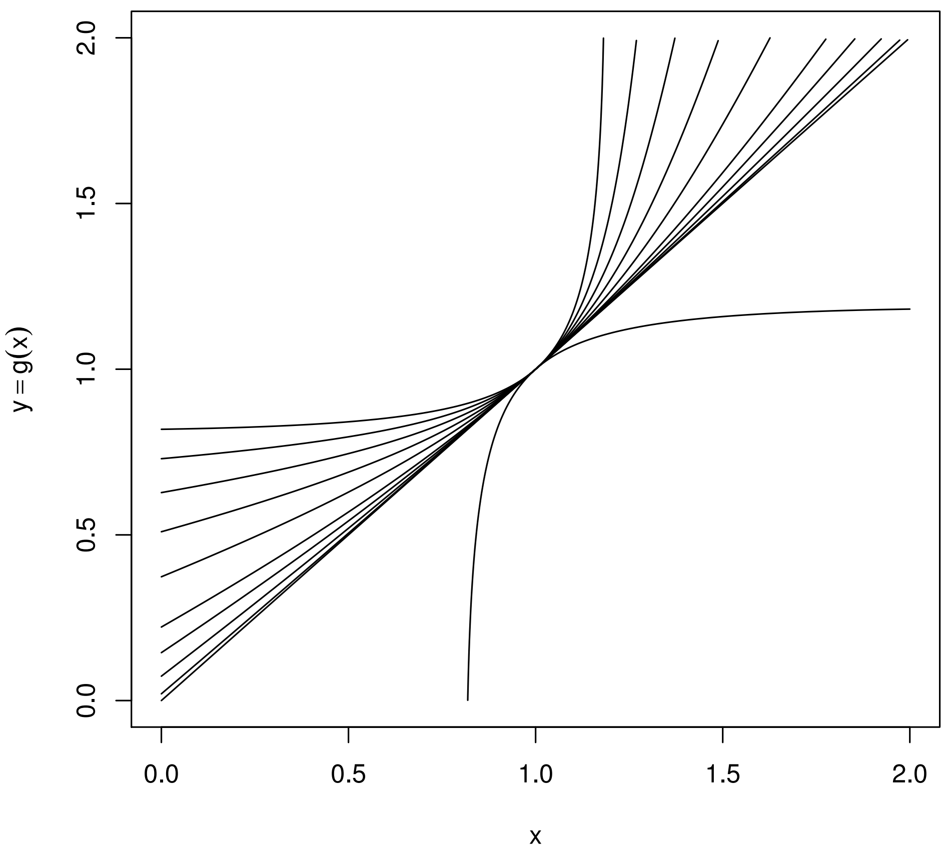


Figure 3: Rejection Rates for Tests of Convexity using Asymptotic Normality ($\delta = 0.0$)

$n = 50$ $n = 100$

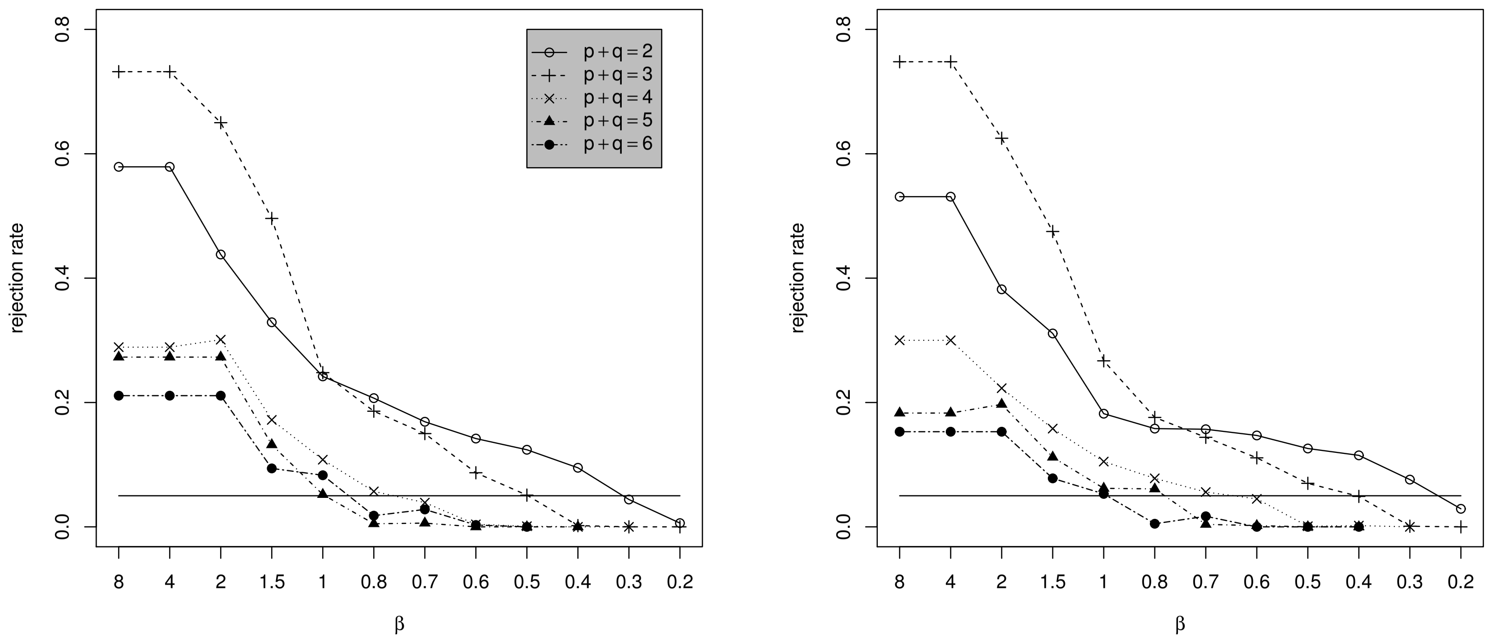
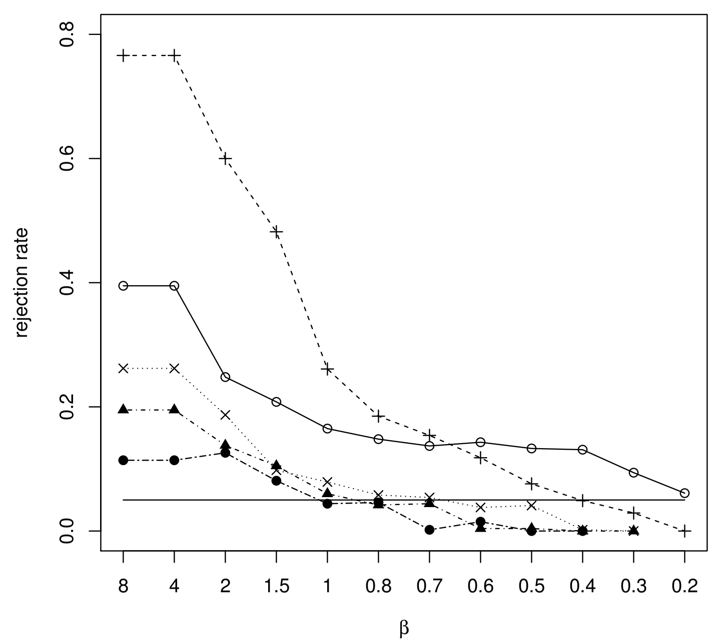


Figure 3: Rejection Rates for Tests of Convexity using Asymptotic Normality (continued)

$n = 200$



$n = 1000$

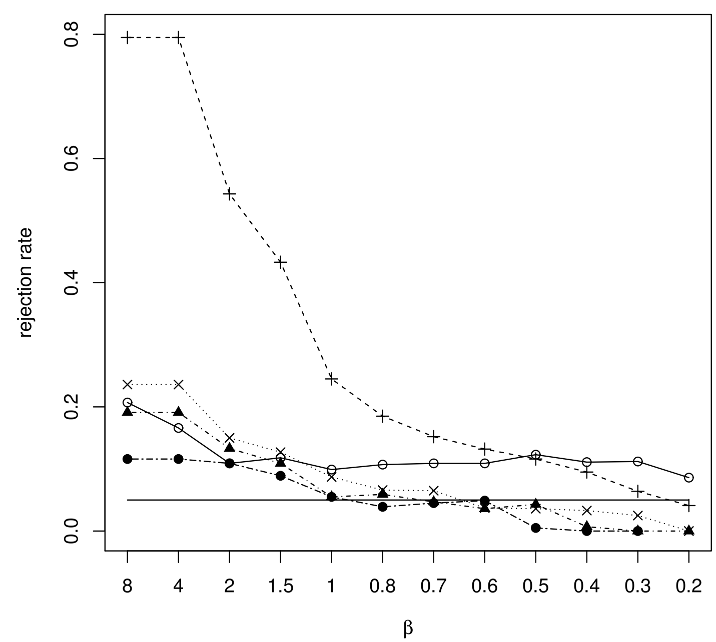
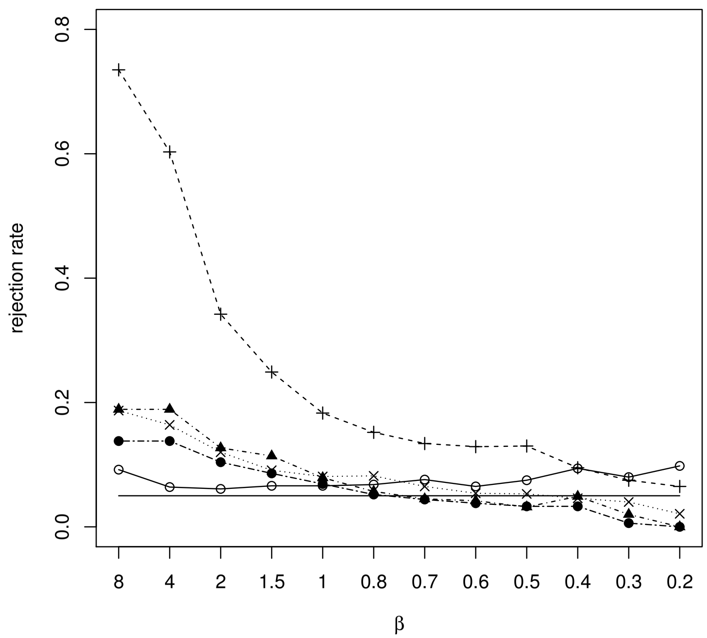


Figure 3: Rejection Rates for Tests of Convexity using Asymptotic Normality (continued)

$n = 10000$



$n = 20000$