Research Division
Federal Reserve Bank of St. Louis
Working Paper Series

Non-parametric, Unconditional Quantile Estimation for Efficiency Analysis with an Application to Federal Reserve Check Processing Operations

David C. Wheelock
and
Paul W. Wilson

Working Paper 2005-027B
http://research.stlouisfed.org/wp/2005/2005-027.pdf

April 2005
Revised February 2007

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

Non-parametric, Unconditional Quantile Estimation for Efficiency Analysis with an Application to Federal Reserve Check Processing Operations

David C. Wheelock Paul W. Wilson*

December 2007

Abstract

This paper examines the technical efficiency of U.S. Federal Reserve check processing offices over 1980–2003. We extend results from Park et al. (2000) and Daouia and Simar (2007) to develop an unconditional, hyperbolic, α-quantile estimator of efficiency. Our new estimator is fully non-parametric and robust with respect to outliers; when used to estimate distance to quantiles lying close to the full frontier, it is strongly consistent and converges at rate root-n, thus avoiding the curse of dimensionality that plagues data envelopment analysis (DEA) estimators. Our methods could be used by policymakers to compare inefficiency levels across offices or by managers of individual offices to identify peer offices.

*Wheelock: Research Department, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166–0442; wheelock@stls.frb.org. Wilson: The John E. Walker Department of Economics, 222 Sirrine Hall, Clemson University, Clemson, South Carolina 29634–1309, USA; email pww@clemson.edu. This research was conducted while Wilson was a visiting scholar in the Research Department of the Federal Reserve Bank of St. Louis. We have benefited from discussions with Léopold Simar and Jean-Pierre Florens, as well as research assistance by Heidi Beyer, Joshua Ulrich, and Neil Wiggins; of course, any remaining errors are ours, not theirs. The views expressed in this paper do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System. *JEL* classification nos.: C14, G21, L32. Keywords: payments system, check processing, productivity, efficiency, quantile estimation.

1 Introduction

Benchmarking has become a wildly popular idea in management, finance, economics, education, public policy, and other arenas; the Google internet search engine currently returns on the order of 22,500,000 hits for the keyword "benchmarking." The Oxford English Dictionary defines benchmarking as "the action or practice of comparing something to a benchmark; evaluation against an established standard," suggesting that while an established standard is important for benchmarking, there may be more than one such standard. Benchmarking may involve detailed evaluation and comparison of a particular unit's operating procedures with those of a competitor, perhaps using standard accounting ratios such as return-on-assets or other measures. Efficiency analysis is a more formal approach, wherein a statistical model of a production process with a well-defined benchmark for purposes of comparison is specified and then estimated, allowing possibilities for statistical inference.

The performance of firms and other decision-making units (DMUs) in terms of technical efficiency, as well as allocative, cost, and other efficiencies, has received widespread attention in the economics, statistics, management science, and related literatures. In the case of private firms, estimates of inefficiency have been used to explain insolvency rates and merger activities, the effects of changes in regulatory environments, and overall industry performance.[1] In the case of public and non-profit entities, estimates of inefficiency are intrinsically interesting because these entities do not face a market test, and inefficiency estimates often provide the only objective criteria for gauging performance. Measuring the performance of public entities may be important for allocating scarce public resources, for deciding which to eliminate during periods of consolidation, *etc.* In particular, identifying inefficient entities is a critical first step in any attempt to improve performance.

Both parametric and non-parametric approaches have been used to estimate inefficiency. A popular parametric approach based on the ideas of Aigner et al. (1977) and Meeusen and van den Broeck (1977) involves the estimation of a specific response function with a composite error term. Often, studies specify a translog response function, which is regarded as a flexible form. Researchers have found, however, that the translog function is often a mis-

[1] See, for example, Berger and Humphrey (1997) for a survey, and Gilbert and Wilson (1998) and Wheelock and Wilson (1995, 2000) for specific applications involving commercial banks.

specification when DMUs are of widely varying size.[2] In an attempt to increase flexibility, some researchers have augmented translog specifications with trigonometric terms along the lines of Gallant (1981, 1982). In order to maximize log-likelihoods, however, the number of additional terms must usually be restricted severely, and in most cases is probably far less than the number that would minimize criteria such as asymptotic mean integrated square error.[3]

Non-parametric approaches, by contrast, are popular because they avoid having to specify *a priori* a particular functional relationship to be estimated; the data are allowed to speak for themselves. Non-parametric methods usually involve the estimation of a production set or some other set by either the free-disposal hull (FDH) of sample observations, or the convex hull of the FDH. Methods based on the convex hull of the FDH are collectively referred to as data envelopment analysis (DEA). DEA is well known and has been applied widely: as of early 2004, DEA had been used in more than 1,800 articles published in some 490 refereed journals (Gattoufi *et al.*, 2004). The statistical properties of DEA estimators have been established, and methods are now available for making statistical inferences about efficiency based on DEA.[4]

Despite their popularity, DEA estimators have some obvious drawbacks. Although these estimators avoid the need for a priori specification of functional forms, they do impose (eventually) non-increasing returns to scale; i.e., they do not allow increasing returns to scale everywhere. Moreover, it has long been recognized that DEA estimates of inefficiency are sensitive to outliers in the data. Perhaps even more problematic, at least in many applications, is that DEA estimators suffer from the well-known *curse of dimensionality*

[2]See Cooper and McLaren (1996), Banks et al. (1997), Wheelock and Wilson (2001), and Wilson and Carey (2004) for empirical examples. For Monte Carlo evidence, see Guilkey et al. (1983) and Chalfant and Gallant (1985).

[3]Published papers using this approach have typically not optimized the number of terms according to such criteria. In addition, Barnett et al. (1991) note that "the basis functions with which Gallant's model seeks to span the neoclassical function space are sines and cosines, despite the fact that such trigonometric functions are periodic and hence are far from neoclassical. In other words, the basis functions, which should be dense in the space to be spanned do not themselves even lie within that space." Instead of trigonometric functions, one could use as the basis functions members of a family of orthogonal polynomials (e.g., Laguerre or Legendre polynomials), but the problems of determining the optimal number of terms, and using these in a non-linear, maximum-likelihood framework, remain.

[4]See Simar and Wilson (2000b) for a survey, and Kneip et al. (2007) for more recent results, on the statistical properties of DEA estimators. See Simar and Wilson (1998, 2000a) and Kneip et al. (2007) for details about the use of bootstrap methods to make inferences based on DEA.

that often plagues non-parametric estimators. The number of observations required to obtain meaningful estimates of inefficiency increases dramatically with the number of production inputs and outputs; for a given sample size, adding dimensions results in more observations falling on the *estimated* frontier. In many applications, including the one in this paper, there are simply too few observations available to obtain meaningful estimates of inefficiency using DEA.[5]

Recently, two alternative non-parametric approaches that avoid some of the problems with the traditional DEA and FDH methods have been developed. Both the "order-m" approach of Cazals et al. (2002) and the "order-α" approach of Daouia (2003), Aragon et al. (2005) and Daouia and Simar (2007) are based on the idea of estimating "partial frontiers" that lie close to the "full frontier" (i.e., the boundary of the production set). Both approaches allow one to interpret estimators of the partial frontiers as estimators of the full frontier by viewing the order (either m or α) as a sequence of appropriate order in the sample size n. Both approaches involve some conditioning with the result that for a fixed order (again, either m or α), the input-oriented partial frontiers are different from the output-oriented partial frontiers of the same fixed order. While this is of little or no consequence when the partial frontier estimators are viewed as estimators of the full frontier, it might be troubling where the partial frontiers themselves are used as benchmarks against which efficiency is measured. As discussed below, a compelling reason for using partial frontiers of fixed order as benchmarks is that they can be estimated with root-n consistency, whereas this feature is lost if the partial frontier estimators are viewed as estimators of the full frontier.

In this paper, we extend results obtained by Daouia and Simar (2007) for conditional (input or output) order-α quantiles to estimate *unconditional, hyperbolic* order-α quantiles, allowing estimation of efficiency along a hyperbolic path where inputs and outputs are adjusted simultaneously, rather than in either strictly an input or an output direction.[6] As discussed below, use of a hyperbolic path avoids some of the ambiguity in choosing between an input or output-orientation, even if our partial frontier estimator is viewed as an estimator

[5]One can find numerous published applications of DEA to datasets with 50–150 observations and 5 or more dimensions in the input-output space. Inefficiency estimates from such studies are likely meaningless in a statistical sense due to the curse of dimensionality problem (see Simar and Wilson (2000b) for discussion).

[6]In the real world, there are undoubtedly situations where inputs and outputs cannot be adjusted simultaneously. This presents no problem here, however, because our method does not involve estimating behavioral relationships.

of the full frontier. As with the conditional estimators of Daouia and Simar, our hyperbolic distance function estimator is fully non-parametric, robust with respect to outliers, and is strongly consistent, converging at the classical, parametric root-n rate when the partial (rather than full) frontier is estimated. Further, as discussed below, the measurement of inefficiency along a hyperbolic path results in near-automatic identification of relevant peers that managers of a DMU might study to learn how to improve their own operations. This contrasts with the traditional DEA and FDH approaches, where a manager might learn which DMUs are more efficient than others, but not necessarily which might be useful to emulate.[7]

We use the unconditional quantile estimator described below to examine the technical efficiency of Federal Reserve check-processing offices. Check processing is a logistic operation similar to processing and delivery operations encountered in numerous industries (e.g., postal and package-delivery services, military supply operations, *etc.*). Fed offices receive checks from the depository institutions (hereafter "banks") that cash the checks or receive them from their depositors. Fed offices sort the checks, credit the accounts of depositing banks, and forward the checks to the banks upon which they are drawn. Federal Reserve offices are required by the Monetary Control Act of 1980 to charge a price for clearing checks that recovers the Fed's costs plus a "private sector adjustment factor" that represents a rate of return and taxes that a private firm would have to pay. Paper check volume has declined markedly in recent years as electronic payments media have become increasingly popular (Gerdes and Walton, 2002). The Fed has eliminated check processing at some locations to reduce costs, and further declines in volume would intensify pressure on the Federal Reserve to reduce costs so as to remain in compliance with the Monetary Control Act. This paper offers a methodology that could assist Fed officials in identifying possible inefficiencies in check operations that could help them achieve their objectives.

The small number of Federal Reserve offices that process checks (49 offices as of 2003) and relatively high dimension of our four-input, two-output model of check processing imply that we are unlikely to obtain statistically meaningful estimates of inefficiency using DEA. This problem is not unusual in situations where researchers seek to evaluate the performance

[7] Simply finding a DMU that is more efficient than one's own may not be particularly useful to a manager if the more efficient DMU operates at a much different scale, or produces a very different mix of outputs using a very different mix of inputs.

of producers. Because of its rapid convergence rate, the unconditional quantile estimator is well-suited to investigating the efficiency of Fed check offices, as well as in other applications where the curse of dimensionality is likely to be an issue.

The paper proceeds as follows: Section 2 briefly discusses traditional non-parametric efficiency estimators and their drawbacks. Section 3 introduces the non-parametric, unconditional quantile estimator, and presents an example illustrating the problems with the conditional approaches and how these are overcome by the unconditional approach. Asymptotic results that are important for interpreting estimates are presented in Section 4; proofs are deferred to the Appendix. Section 5 describes a model of check processing operations, Section 6 presents estimation results, Section 7 offers policy recommendations, and Section 8 concludes.

2 Problems with Traditional Efficiency Estimators

Standard microeconomic theory of the firm introduces the notion of a production possibilities set

$$\mathcal{P} \equiv \{(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \text{ can produce } \boldsymbol{y}\} \subset \mathbb{R}_+^{p+q}, \tag{2.1}$$

where $\boldsymbol{x} \in \mathbb{R}_+^p$ and $\boldsymbol{y} \in \mathbb{R}_+^q$ denote vectors of inputs and outputs, respectively. The upper boundary of $\mathcal{P}$, denoted $\mathcal{P}^\partial$, is sometimes referred to as the *technology* or the *production frontier.* In economics, management science, and other disciplines, the goal is often to estimate distance from an arbitrary point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{p+q}$ to the boundary $\mathcal{P}^\partial$ along a particular path. Shephard (1970) defined input and output distance functions given by

$$\theta(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) \equiv \sup \left\{\theta > 0 \mid (\theta^{-1}\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}\right\} \tag{2.2}$$

and

$$\lambda(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) \equiv \inf \left\{\lambda > 0 \mid (\boldsymbol{x}, \lambda^{-1}\boldsymbol{y}) \in \mathcal{P}\right\}, \tag{2.3}$$

respectively. The input distance function $\theta(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ measures distance from $(\boldsymbol{x}, \boldsymbol{y})$ to $\mathcal{P}^\partial$ in a direction orthogonal to $\boldsymbol{y}$, while the output distance function $\lambda(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ measures distance from the same point to $\mathcal{P}^\partial$ in a direction orthogonal to $\boldsymbol{x}$.

Under constant returns to scale (CRS), $\theta(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) = \lambda(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})^{-1}$. However, with variable returns to scale (VRS), the choice of orientation (either input or output) can have a large impact on measured efficiency. With VRS, a large firm could conceivably lie close to the frontier $\mathcal{P}^\partial$ in the output direction, but far from $\mathcal{P}^\partial$ in the input direction. Similarly, a small firm might lie close to $\mathcal{P}^\partial$ in the input direction, but far from $\mathcal{P}^\partial$ in the output direction. Such differences are related to the slope and curvature of $\mathcal{P}^\partial$.

As an alternative to the Shephard (1970) input and output measures, Färe et al. (1985) proposed measuring efficiency along a hyperbolic path from the point of interest to $\mathcal{P}^\partial$. The hyperbolic-graph distance function given by

$$\gamma(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) \equiv \sup\left\{\gamma > 0 \mid (\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y}) \in \mathcal{P}\right\} \tag{2.4}$$

measures distance from the fixed point $(\boldsymbol{x}, \boldsymbol{y})$ to $\mathcal{P}^\partial$ along the hyperbolic path $(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y})$, $\gamma \in \mathbb{R}^1_{++}$. Note that for $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$, $\theta(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) \geq 1$, $\lambda(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) \leq 1$, and $\gamma(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) \geq 1$ by construction.[8]

The distance functions in (2.2)–(2.4) are defined in terms of the *unknown, true* production set $\mathcal{P}$, and must be *estimated* from a set $\mathcal{S}_n = \{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^n$ of observed input/output combinations. Traditional, non-parametric approaches typically assume $\Pr((\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathcal{P}) = 1 \ \forall\ i = 1, \ \ldots,\ n$ and replace $\mathcal{P}$ in (2.2)–(2.4) with an estimator of the production set to obtain estimators of the Shephard input- and output-oriented distance functions. Deprins et al. (1984) proposed the free-disposal hull (FDH) of the observations in $\mathcal{S}_n$, i.e.,

$$\widehat{\mathcal{P}}_{\text{FDH}}(\mathcal{S}_n) = \bigcup_{(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathcal{S}_n} \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{p+q}_+ \mid \boldsymbol{y} \leq \boldsymbol{y}_i,\ \boldsymbol{x} \geq \boldsymbol{x}_i\}. \tag{2.5}$$

DEA estimators, assuming VRS, are obtained by replacing $\mathcal{P}$ in (2.2)–(2.4) with the convex hull of $\widehat{\mathcal{P}}_{\text{FDH}}$, given by

$$\widehat{\mathcal{P}}_{\text{DEA}}(\mathcal{S}_n) = \Big\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{p+q}_+ \mid \boldsymbol{y} \leq \sum_{i=1}^n \kappa_i \boldsymbol{y}_i,\ \boldsymbol{x} \geq \sum_{i=1}^n \kappa_i \boldsymbol{x}_i,$$
$$\sum_{i=1}^n \kappa_i = 1,\ \kappa_i \geq 0 \ \forall\ i = 1,\ \ldots,\ n \Big\}, \tag{2.6}$$

while DEA estimators incorporating CRS are obtained by replacing $\mathcal{P}$ with the convex cone of $\widehat{\mathcal{P}}_{\text{DEA}}(\mathcal{S}_n)$ or $\widehat{\mathcal{P}}_{\text{FDH}}(\mathcal{S}_n)$, obtained by dropping the constraint $\sum_{i=1}^n \kappa_i = 1$ in (2.6). DEA

[8]The Shephard (1970) input and output distance functions defined in (2.2) and (2.3) are reciprocals of the corresponding Farrell (1957) measures. Färe et al. (1985) defined a Farrell-type hyperbolic measure that is the reciprocal of the measure defined in (2.4).

estimates of input or output distance functions are obtained by solving the resulting familiar linear programs, while FDH estimates based on (2.5) are obtained more easily using simple numerical computations. Estimators of the hyperbolic distance function in (2.4) under VRS have not been used in practice, apparently due to the fact that when $\widehat{\mathcal{P}}_{\text{DEA}}(\mathcal{S}_n)$ replaces $\mathcal{P}$ in (2.4), the resulting estimator cannot be expressed as a linear program, although it can be computed using numerical methods.

Asymptotic properties of estimators of the input and output distance functions in (2.2)–(2.3) based on $\widehat{\mathcal{P}}_{\text{FDH}}(\mathcal{S}_n)$ and $\widehat{\mathcal{P}}_{\text{DEA}}(\mathcal{S}_n)$, as well as the assumptions needed to establish consistency of the estimators, are summarized in Simar and Wilson (2000b). Unfortunately, however, despite their widespread use, both DEA and FDH estimators suffer from a number of vexing problems:

1. Convergence rates are pathologically slow when there are more than a few inputs and outputs. DEA (VRS) estimators of $\theta(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ and $\lambda(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ defined in (2.2) and (2.3) converge at the rate $n^{-2/(p+q+1)}$ (Kneip et al., 1998), while FDH estimators converge at an even slower rate, $n^{-1/(p+q)}$ (Park et al., 2000).[9] The convergence rates become worse as dimensionality $(p + q)$ increases, becoming quite slow for commonly used numbers of inputs and outputs. Although many published applications of DEA estimators have used 100–200 observations with 5–10 dimensions, the slow convergence rate of DEA estimators means that the results of such studies are largely meaningless from a statistical viewpoint.

2. In addition to slow convergence rates, DEA and FDH estimators are extremely sensitive to outliers. Although several methods have been proposed for detecting outliers in high-dimensional data (e.g., Wilson, 1993, 1995; Kuntz and Scholtes, 2000; Simar, 2003; and

[9] The faster convergence rate of DEA estimators results from convexification of the free-disposal hull of the data, but DEA estimators are inconsistent if the true production set is not convex, while FDH estimators remain consistent. Convergence rates of DEA and FDH estimators of the hyperbolic measure $\gamma(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ defined in (2.4) have not been established previously. However, rates of convergence of DEA and FDH input- and output-efficiency estimators are similar to the rates of convergence derived by Korostelev et al. (1995a, 1995b) for DEA and FDH *frontier* estimators. Unfortunately, as noted by Kneip et al. (1998), there appears to be no straightforward way to adapt the results obtained by Korostelev et al. to the problem of determining convergence rates of efficiency estimators. Nonetheless, since DEA and FDH estimators of the hyperbolic measure $\gamma(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ measure distance from a fixed point $(\boldsymbol{x}, \boldsymbol{y})$ to the boundary of $\widehat{\mathcal{P}}_{\text{DEA}}$ or $\widehat{\mathcal{P}}_{\text{FDH}}$, as do DEA and FDH estimators of $\theta(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ and $\lambda(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$, one might speculate that the convergence rates of DEA and FDH estimators of the hyperbolic efficiency measure might be similar to their input- and output-oriented counterparts. This is confirmed below in Section 4 for the FDH case.

Porembski et al., 2005), some subjective interpretation is required. Moreover, merely because an outlier is found does not mean it should be deleted. Outliers are atypical observations; an observation might be atypical because it has low probability of being observed. In this case, the outlier might be the most interesting part of the data. On the other hand, outliers can also result from measurement errors, coding errors, or other mistakes. When data have been corrupted by such errors, they should be repaired or deleted if correction is not possible. In applied work, however, it is often difficult to identify why an observation is atypical.

3. Under VRS, some portions of the boundaries of $\widehat{\mathcal{P}}_{\text{FDH}}(\mathcal{S}_n)$ and $\widehat{\mathcal{P}}_{\text{DEA}}(\mathcal{S}_n)$ will necessarily be parallel to the output axes, while others will necessarily be parallel to the input axes. For observations lying near the boundaries of $\widehat{\mathcal{P}}_{\text{FDH}}(\mathcal{S}_n)$ and $\widehat{\mathcal{P}}_{\text{DEA}}(\mathcal{S}_n)$ where these phenomena occur, estimates of input- and output-efficiency can differ greatly. For example, an observation just below the boundary of $\widehat{\mathcal{P}}_{\text{DEA}}(\mathcal{S}_n)$ in the region where this boundary is parallel to the input axes might appear very efficient in an output-orientation, but very inefficient in an input orientation. The problem could be avoided by using $\widehat{\mathcal{P}}_{\text{FDH}}(\mathcal{S}_n)$ or $\widehat{\mathcal{P}}_{\text{DEA}}(\mathcal{S}_n)$ to estimate the hyperbolic measure $\gamma(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ defined in (2.4), but the other problems would remain.

The conditional α-quantile estimators developed by Daouia (2003), Aragon et al. (2005) and Daouia and Simar (2007), as well as the unconditional hyperbolic α-quantile estimator developed in the next section, are more robust with respect to outliers than standard DEA and FDH estimators, and achieve root-n convergence when partial frontiers (or efficiency measured relative to partial frontiers) are estimated. However, the conditional α-quantile estimators do not address the third problem listed above; moreover, if efficiency is measured relative to an estimate of a partial frontier, in order to retain the root-n rate of convergence, the problem is compounded by the existence of different partial frontiers for either the input- or output-orientations. Using a hyperbolic measure of efficiency avoids this problem, and as noted previously, results in near-automatic identification of relevant peers for meaningful comparisons among firms.

3 Quantile Estimation for Efficiency Analysis

3.1 The Statistical Model

Together with the definition in (2.1), the following assumptions, similar to those found in Park et al. (2000), define a statistical model:

Assumption 3.1. *The production set* $\mathcal{P}$ *is compact and free disposal, i.e., if* $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$, $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \in \mathcal{P}$, *and* $\widetilde{\boldsymbol{x}} \geq \boldsymbol{x}$, *then* $(\widetilde{\boldsymbol{x}}, \widetilde{\widetilde{\boldsymbol{y}}}) \in \mathcal{P} \; \forall \; \mathbf{0} \leq \widetilde{\widetilde{\boldsymbol{y}}} \leq \boldsymbol{y}$.

Assumption 3.2. *The sample observations* $\mathcal{S}_n = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{n}$ *are realizations of identically, independently distributed (iid) random variables with probability density function* $f(\boldsymbol{x}, \boldsymbol{y})$ *with support over* $\mathcal{P}$.

A point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$ is said to be on the *frontier* of $\mathcal{P}$, denoted $\mathcal{P}^{\partial}$, if $(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y}) \notin \mathcal{P}$ for any $\gamma > 1$; let $(\boldsymbol{x}_0^{\partial}, \boldsymbol{y}_0^{\partial}) \in \mathcal{P}^{\partial}$ denote such a point.

Assumption 3.3. *At the frontier, the density* f *is strictly positive, i.e.,* $f_0 = f(\boldsymbol{x}_0^{\partial}, \boldsymbol{y}_0^{\partial}) > 0$, *and sequentially Lipschitz continuous, i.e., for all sequences* $(\boldsymbol{x}_n, \boldsymbol{y}_n) \in \mathcal{P}$ *converging to* $(\boldsymbol{x}_0^{\partial}, \boldsymbol{y}_0^{\partial})$, $|f(\boldsymbol{x}_n, \boldsymbol{y}_n) - f(\boldsymbol{x}_0^{\partial}, \boldsymbol{y}_0^{\partial})| \leq c_1 ||(\boldsymbol{x}_n, \boldsymbol{y}_n) - (\boldsymbol{x}_0^{\partial}, \boldsymbol{y}_0^{\partial})||$ *for some positive constant* c_1.

Now let y^k denote the kth element of $\boldsymbol{y}$, $k = 1, \ldots, q$, and let $\boldsymbol{y}^{(k)} = \left[y^1 \;\; \ldots \;\; y^{k-1} \;\; y^{k+1} \;\; \ldots \;\; y^q\right]$ denote the vector $\boldsymbol{y}$ with the kth element deleted. In addition, let $\boldsymbol{y}^{(k)}(\eta) = \left[y^1 \;\; \ldots \;\; y^{k-1} \;\; \eta \;\; y^{k+1} \;\; \ldots \;\; y^q\right]$ denote a vector similar to $\boldsymbol{y}$, but with η substituted for the kth element of $\boldsymbol{y}$. For each $k = 1, \ldots, 1$ define a function

$$g_{\mathcal{P}}^{k}\left(\boldsymbol{x}, \boldsymbol{y}^{(k)}\right) \equiv \max\left\{\eta \mid \left(\boldsymbol{x}, \boldsymbol{y}^{(k)}(\eta)\right) \in \mathcal{P}\right\}. \tag{3.1}$$

As discussed in Park et al. (2000), the production set $\mathcal{P}$ can be defined in terms of any of the functions $g_{\mathcal{P}}^{k}$. Along the lines of Park et al., the following analysis is presented in terms of $g_{\mathcal{P}}^{q}$, denoted simply as g.

Assumption 3.4. *At the frontier,* $g(\cdot, \cdot)$ *is (i) positive, i.e.,* $g(\boldsymbol{x}_0^{\partial}, \boldsymbol{y}_0^{\partial}) > 0$*; (ii) continuously differentiable; and (iii) the first derivative is Lipschitz continuous, i.e., for all* $(\boldsymbol{x}, \boldsymbol{y})$, $|g(\boldsymbol{x}, \boldsymbol{y}^{(q)}) - g(\boldsymbol{x}_0^{\partial}, \boldsymbol{y}_0^{\partial(q)}) - \nabla g(\boldsymbol{x}_0^{\partial}, \boldsymbol{y}_0^{\partial(q)})'((\boldsymbol{x}, \boldsymbol{y}^{(q)}) - (\boldsymbol{x}_0^{\partial}, \boldsymbol{y}_0^{\partial(q)}))| \leq c_2 ||(\boldsymbol{x}, \boldsymbol{y}^{(q)}) - (\boldsymbol{x}_0, \boldsymbol{y}_0^{(q)}))||^2$ *for some positive constant* c_2, *and for* $k = 1, \ldots, p$ *and* $\ell = 1, \ldots, q-1$, $\frac{\partial}{\partial x^k} g(\boldsymbol{x}, \boldsymbol{y}^{(q)})|_{(\boldsymbol{x}, \boldsymbol{y}^{(q)}) = (\boldsymbol{x}_0^{\partial}, \boldsymbol{y}_0^{\partial(q)})} > 0$ *and* $\frac{\partial}{\partial y^\ell} g(\boldsymbol{x}, \boldsymbol{y}^{(q)})|_{(\boldsymbol{x}, \boldsymbol{y}^{(q)}) = (\boldsymbol{x}_0^{\partial}, \boldsymbol{y}_0^{\partial(q)})} < 0$.

The density $f(\boldsymbol{x}, \boldsymbol{y})$ introduced in Assumption 3.3 implies a probability function

$$H(\boldsymbol{x}_0, \boldsymbol{y}_0) = \Pr(\boldsymbol{x} \leq \boldsymbol{x}_0, \boldsymbol{y} \geq \boldsymbol{y}_0). \tag{3.2}$$

This is a non-standard probability distribution function, given the direction of the inequality for $\boldsymbol{y}$; nonetheless, it is well-defined. This function gives the probability of drawing an observation from $f(\boldsymbol{x}, \boldsymbol{y})$ that weakly *dominates* the DMU operating at $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{P}$; an observation $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})$ weakly dominates $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ if $\widetilde{\boldsymbol{x}} \leq \boldsymbol{x}_0$ and $\widetilde{\boldsymbol{y}} \geq \boldsymbol{y}_0$. Clearly, $H(\boldsymbol{x}_0, \boldsymbol{y}_0)$ is monotone, nondecreasing in $\boldsymbol{x}_0$ and monotone, non-increasing in $\boldsymbol{y}_0$.

Using $H(\cdot, \cdot)$, the hyperbolic distance function in (2.4) can be written as

$$\gamma(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}) = \sup\left\{\gamma > 0 \mid H(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y}) > 0\right\}. \tag{3.3}$$

Alternatively, define the hyperbolic α-quantile distance function

$$\gamma_\alpha(\boldsymbol{x}, \boldsymbol{y}) = \sup\left\{\gamma > 0 \mid H(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y}) > (1-\alpha)\right\} \tag{3.4}$$

for $\alpha \in (0, 1]$. If $\alpha = 1$, then $\gamma_\alpha(\boldsymbol{x}, \boldsymbol{y}) = \gamma(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$. For $0 < \alpha < 1$ and a fixed point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{p+q}$, $\gamma_\alpha(\boldsymbol{x}, \boldsymbol{y}) >$ (respectively, $<$) 1 gives the proportionate, simultaneous decrease (increase) in inputs and increase (decrease) in outputs required to move from $(\boldsymbol{x}, \boldsymbol{y})$ along a path $(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y})$, $\gamma > 0$, to a point with $(1-\alpha)$ probability of being weakly dominated. The hyperbolic α-quantile frontier is defined by

$$\mathcal{P}_\alpha^\partial = \left\{(\gamma_\alpha(\boldsymbol{x}, \boldsymbol{y})^{-1}\boldsymbol{x}, \gamma_\alpha(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{y}) \mid (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}\right\}. \tag{3.5}$$

Using Assumption 3.3 and the fact that $H(\boldsymbol{x}_0, \boldsymbol{y}_0)$ is monotone, nondecreasing in $\boldsymbol{x}_0$ and monotone, non-increasing in $\boldsymbol{y}_0$, it is easy to show that $\mathcal{P}_\alpha^\partial$ is monotone in the sense that if $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{P}_\alpha^\partial$, $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \in \mathcal{P}_\alpha^\partial$, and $\widetilde{\boldsymbol{x}} \geq \boldsymbol{x}_0$, then $\widetilde{\boldsymbol{y}} \geq \boldsymbol{y}_0$.

The probabilistic formulation used here is closely related to the work of Daouia and Simar (2007), which builds on earlier work by Daouia (2003) and Aragon et al. (2005). Daouia and Simar decompose the distribution function given in (3.2) to obtain

$$\begin{aligned} H(\boldsymbol{x}_0, \boldsymbol{y}_0) &= \Pr(\boldsymbol{x} \leq \boldsymbol{x}_0 \mid \boldsymbol{y} \geq \boldsymbol{y}_0)\Pr(\boldsymbol{y} \geq \boldsymbol{y}_0) = F_{x|y}(\boldsymbol{x}_0 \mid \boldsymbol{y}_0)S_y(\boldsymbol{y}_0) \\ &= \Pr(\boldsymbol{y} \geq \boldsymbol{y}_0 \mid \boldsymbol{x} \leq \boldsymbol{x}_0)\Pr(\boldsymbol{x} \leq \boldsymbol{x}_0) = S_{y|x}(\boldsymbol{y}_0 \mid \boldsymbol{x}_0)F_x(\boldsymbol{x}_0) \end{aligned} \tag{3.6}$$

(the terms on the right-hand side of (3.6) also appear in Cazals et al., 2002, and Daraio and Simar, 2005). Working in a Farrell-type framework, they define conditional quantile-based efficiency scores that are equivalent to the reciprocals of the Shephard-type input- and output-oriented conditional α-quantile distance functions given by

$$\theta_\alpha(\boldsymbol{x}, \boldsymbol{y}) = \sup\left\{\theta \geq 0 \mid F_{x|y}(\theta^{-1}\boldsymbol{x} \mid \boldsymbol{y}) > (1-\alpha)\right\} \tag{3.7}$$

and

$$\lambda_\alpha(\boldsymbol{x}, \boldsymbol{y}) = \inf\left\{\lambda \geq 0 \mid S_{y|x}(\lambda\boldsymbol{y} \mid \boldsymbol{x}) > (1-\alpha)\right\}. \tag{3.8}$$

For $\alpha \in (0,1)$, $\theta_\alpha(\boldsymbol{x}, \boldsymbol{y}) < \theta(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ and $\lambda_\alpha(\boldsymbol{x}, \boldsymbol{y}) > \lambda(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ by construction. These implicitly define input and output conditional α-quantile frontiers given by

$$\mathcal{P}^\partial_{x,\alpha} = \left\{(\theta_\alpha(\boldsymbol{x}, \boldsymbol{y})^{-1}\boldsymbol{x}, \boldsymbol{y}) \mid (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}\right\} \tag{3.9}$$

and

$$\mathcal{P}^\partial_{y,\alpha} = \left\{(\boldsymbol{x}, \lambda_\alpha(\boldsymbol{x}, \boldsymbol{y})^{-1}\boldsymbol{y}) \mid (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}\right\}, \tag{3.10}$$

respectively.

Figure 1 provides an illustration for the simple case where $p = q = 1$, and $f(x, y)$ is uniform over the unit triangle with corners at (0,0), (1,0), and (1,1). The solid line shows $\mathcal{P}^\partial$. Simple algebra leads to expressions for the terms appearing in (3.6), making it easy to compute $\gamma_\alpha(x, y)$ as well as $\theta_\alpha(x, y)$ and $\lambda_\alpha(x, y)$ for a variety of pairs (x, y). Hence, for a given value of $\alpha \in (0, 1)$, the hyperbolic α-quantile frontier and the two conditional α-quantile frontiers can be traced. This has been done in Figure 1 for $\alpha = 0.95$, where $\mathcal{P}^\partial_\alpha$ is illustrated by the dashed line, while the conditional α-quantile functions are shown by the dotted lines. The steeper of the two dotted lines shows the input-oriented conditional α-quantile frontier; the other shows the output-oriented conditional α-quantile frontier. For a fixed $\alpha \in (0, 1)$, these frontiers differ from one another, although the difference diminishes as $\alpha \to 1$. The input frontier $\mathcal{P}^\partial_{x,\alpha}$ will necessarily have steeper slope than $\mathcal{P}^\partial$ along a ray $\lambda\boldsymbol{x}$, $\lambda > 0$, while the output frontier $\mathcal{P}^\partial_{y,\alpha}$ will have less steep slope than $\mathcal{P}^\partial$ along a ray $\lambda\boldsymbol{x}$, $\lambda > 0$.

The hyperbolic α-quantile shown in Figure 1 parallels the full frontier $\mathcal{P}^\partial$ due to the uniform distribution of x and y; if $f(x, y)$ were not uniform, then $\mathcal{P}^\partial_\alpha$ might be more or less steep than $\mathcal{P}^\partial$, depending on the shape of $f(x, y)$ near the full frontier. For example, if $f(x, y)$ concentrated probability mass near the full frontier for large x and was relatively disperse near the full frontier for small x, the partial frontier $\mathcal{P}_{\alpha}^{\partial}$ would tend to be steeper than the full frontier $\mathcal{P}^{\partial}$. Nonetheless, it would remain true that the conditional α-quantile frontiers $\mathcal{P}_{x,\alpha}^{\partial}$ and $\mathcal{P}_{y,\alpha}^{\partial}$ would have different slopes (i.e., different from the slopes of $\mathcal{P}_{\alpha}^{\partial}$ and $\mathcal{P}^{\partial}$, as well as different from each other's slopes).

Measurement of efficiency requires comparison against a benchmark. While traditional nonparametric approaches have used the full frontier $\mathcal{P}^{\partial}$ as a benchmark, efficiency can also be measured in terms of other benchmarks, such as partial frontiers. Measuring efficiency relative to the hyperbolic α-quantile $\mathcal{P}_{\alpha}^{\partial}$ may well result in different efficiency measures than one would obtain using the full frontier as a benchmark, requiring a different interpretation (this is discussed in greater detail below in Section 6). However, as will be seen in Sections 3.2 and 4 below, it is "easier" to estimate $\mathcal{P}_{\alpha}^{\partial}$ in the sense that $\mathcal{P}_{\alpha}^{\partial}$ can be estimated with root-n consistency, while estimating $\mathcal{P}^{\partial}$ incurs the curse of dimensionality, as well as difficulties for inference. Moreover, it is common to evaluate performance in terms of quantiles (often expressed in terms of percentiles for such purposes); e.g., student performance is often expressed in terms of test-score percentiles; child height and weight are often described in terms of percentiles or quantiles of a distribution, etc. Thus, while technical efficiency has traditionally been benchmarked against the full frontier $\mathcal{P}^{\partial}$, applied researchers as well as consumers of policy studies should find comparisons based on the partial frontier $\mathcal{P}_{\alpha}^{\partial}$ intuitive and familiar.

3.2 Nonparametric, Hyperbolic Estimators

Estimation of $\gamma_{\alpha}(\boldsymbol{x}, \boldsymbol{y})$, and hence $\mathcal{P}_{\alpha}^{\partial}$, is straightforward. The empirical analog of the distribution function defined in (3.2) is given by

$$\widehat{H}_n(\boldsymbol{x}_0, \boldsymbol{y}_0) = n^{-1} \sum_{i=1}^{n} I(\boldsymbol{x}_i \leq \boldsymbol{x}_0, \boldsymbol{y}_i \geq \boldsymbol{y}_0), \tag{3.11}$$

where $I(\cdot)$ denotes the indicator function. Then an estimator of $\gamma_{\alpha}(\boldsymbol{x}, \boldsymbol{y})$ is obtained by replacing $H(\cdot, \cdot)$ in (3.4) with $\widehat{H}_n(\cdot, \cdot)$ to obtain

$$\widehat{\gamma}_{\alpha,n}(\boldsymbol{x}, \boldsymbol{y}) = \sup \left\{ \gamma > 0 \mid \widehat{H}_n(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y}) > (1-\alpha) \right\}. \tag{3.12}$$

Computing $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x}, \boldsymbol{y})$ is essentially a univariate problem. An exact solution is possible using methods similar to those used by Daouia and Simar to compute estimators of the input- and output-conditional α-quantile distance functions defined in (3.7)–(3.8).

Given $\mathcal{S}_n$ and the fixed point of interest $(\boldsymbol{x}_0, \boldsymbol{y}_0)$, for each $i = 1, \ldots, n$, define $\chi_i = \min_{j=1,\ldots,p} \frac{x_0^j}{x_i^j}$ and $\psi_i = \min_{j=1,\ldots,q} \frac{y_i^j}{y_0^j}$. Let $\mathcal{A} = \{i \mid \boldsymbol{y}_0\chi_i \leq \boldsymbol{y}_i\}$, $\mathcal{B} = \{i \mid \boldsymbol{x}_0\psi_i^{-1} \geq \boldsymbol{x}_i\}\backslash\mathcal{A}$, and $\mathcal{C} = \{\omega_j\} = \{\chi_i \mid i \in \mathcal{A}\} \bigcup \{\psi_i \mid i \in \mathcal{B}\}$.[10] Then $\#\mathcal{C} = n$. A nonparametric estimator of the hyperbolic α-quantile distance function $\gamma_\alpha(\boldsymbol{x}_0, \boldsymbol{y}_0)$ defined in (3.4) is given by

$$\widehat{\gamma}_{\alpha,n}(\boldsymbol{x}_0, \boldsymbol{y}_0) = \begin{cases} \omega_{(\alpha n)} & \text{if } \alpha n \in \mathbb{N}_{++}, \\ \omega_{([\alpha n]+1)} & \text{otherwise;} \end{cases} \tag{3.13}$$

where $[\alpha n]$ denotes the integer part of αn, $\mathbb{N}_{++}$ denotes the set of strictly positive integers, and $\omega_{(j)}$ denotes the jth largest element of the set $\mathcal{C}$, i.e., $\omega_{(1)} \leq \omega_{(2)} \leq \ldots \leq \omega_{(n)}$.

An estimator of distance to the full frontier, $\mathcal{P}^\alpha$, is obtained by setting $\alpha = 1$ in (3.13); denote the resulting estimator by $\widehat{\gamma}_n(\boldsymbol{x}_0, \boldsymbol{y}_0)$. This amounts to replacing $\mathcal{P}$ in (2.4) by the FDH estimator of $\mathcal{P}$ defined in (2.5).

Alternatively, given the point of interest $(\boldsymbol{x}_0, \boldsymbol{y}_0)$, it is easy to find initial values γ_a, γ_b that bracket the solution so that $\widehat{H}_n(\gamma_a^{-1}\boldsymbol{x}_0, \gamma_a\boldsymbol{y}_0) > (1-\alpha)$ and $\widehat{H}_n(\gamma_b^{-1}\boldsymbol{x}_0, \gamma_b\boldsymbol{y}_0) < (1-\alpha)$, and then solve for $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ using the bisection method. This method can be made accurate to an arbitrarily small degree. The following algorithm describes the procedure:

[1] Set $\gamma_1 := 1$, $\gamma_b := 1$.

[2] If $\widehat{H}_n(\gamma_a^{-1}\boldsymbol{x}, \gamma_a\boldsymbol{y}) \leq (1-\alpha)$ then set $\gamma_a := 0.5 \times \gamma_a$.

[3] Repeat step [2] until $\widehat{H}_n(\gamma_a^{-1}\boldsymbol{x}, \gamma_a\boldsymbol{y}) > (1-\alpha)$.

[4] If $\widehat{H}_n(\gamma_b^{-1}\boldsymbol{x}, \gamma_b\boldsymbol{y}) \geq (1-\alpha)$ then set $\gamma_b := 2 \times \gamma_b$.

[5] Repeat step [4] until $\widehat{H}_n(\gamma_b^{-1}\boldsymbol{x}, \gamma_b\boldsymbol{y}) < (1-\alpha)$.

[6] Set $\gamma_c := (\gamma_a + \gamma_b)/2$ and compute $\widehat{H}_n(\gamma_c^{-1}\boldsymbol{x}, \gamma_c\boldsymbol{y})$.

[7] If $\widehat{H}_n(\gamma_c^{-1}\boldsymbol{x}, \gamma_c\boldsymbol{y}) \leq (1-\alpha)$ then set $\gamma_b := \gamma_c$; otherwise set $\gamma_a := \gamma_c$.

[10] Removing elements of $\mathcal{A}$ in the definition of $\mathcal{B}$ is required in the event of a tie, which can occur when the path $(\gamma^{-1}\boldsymbol{x}_0, \gamma\boldsymbol{y}_0)$, $\gamma > 0$, passes through one of the the observations in $\mathcal{S}_n$.

[8] If $(\gamma_b - \gamma_a) > \epsilon$, where ϵ is a suitably small tolerance value, repeat steps [6]–[7].

[9] If $\widehat{H}_n(\gamma_c^{-1}\boldsymbol{x}, \gamma_c \boldsymbol{y}) \leq (1-\alpha)$ set $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x}, \boldsymbol{y}) := \gamma_a$; otherwise set $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x}, \boldsymbol{y}) := \gamma_c$.

Note that finding γ_a first reduces the computational burden. Given γ_a, γ_b can be found using only the subset of sample observations that dominate the point $(\gamma_a^{-1}\boldsymbol{x}, \gamma_a \boldsymbol{y})$. Moreover, only this same subset of observations need be used in the first pass through steps [6]–[7]. Upon reaching step [8], the relevant subset of observations can be further reduced each time γ_a is reset in step [7]. Setting the convergence tolerance ϵ in step [8] to 10^{-6} will yield solutions accurate to 5 decimal places, which is likely to be sufficient for most applications.[11] Due to storage requirements, sorting, and the large number of logical comparisons required by the exact method, computing $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x}, \boldsymbol{y})$ using the bisection method is much faster than the exact method.[12] Applied researchers can use the `hquan` command in versions 1.1 (and later) of Wilson's (2007) *FEAR* library of routines for efficiency analysis to compute estimates $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x}, \boldsymbol{y})$ using the numerical procedure described above.

3.3 A Simple Illustration

Figure 2 illustrates the hyperbolic distance function estimator $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x}, \boldsymbol{y})$ for the simple one input, one output case ($p = q = 1$). Twenty observations are represented by the solid diamonds, and the DEA estimate of the frontier is shown by the piecewise-linear curve passing through the five observations at (2, 2), (3, 3), (4, 4), (5, 4.2), and (8, 4.4). Letting (x_0, y_0) represent the point (8, 4.4), the dotted curve passing through the observation at (8, 4.4) shows the path $\gamma^{-1}x, \gamma y \ \forall \ \gamma > 0$. For $\alpha = 0.8$, $\widehat{\gamma}_{\alpha,20}(x_0, y_0) = 0.9090\overline{90}$ and $\widehat{\gamma}_{\alpha,20}(x_0, y_0)^{-1} = 1.1$.

The open circle in Figure 2 indicates the point $(\widehat{\gamma}_{\alpha,20}(x_0, y_0)^{-1}x, \ \widehat{\gamma}_{\alpha,20}(x_0, y_0)y) = (8.8, 4)$. Since $n = 20$ and $\alpha = 0.8$ in this example, exactly $(1-\alpha)n = (1-0.8)20 = 4$ sample

[11] Note that many rational decimal fractions become irrational numbers in the base-2 representation used by modern digital computers; *e.g.,* the base-10 fraction 0.95 has no exact representation in base-2. To avoid problems with the logical comparisons in steps [7] and [9], comparisons should be made against $(1-\alpha-\nu)$ instead of $(1-\alpha)$, where ν is the smallest positive real number that can be represented on the computer architecture in use that yields the result $1-\nu \neq 1$. For machines using 64-bit IEEE arithmetic, this number is $2^{-53} \approx 1.110223 \times 10^{-16}$.

[12] The hyperbolic α-quantile distance function estimates and corresponding bootstrap confidence interval estimates shown in Table 3 took longer by a factor of roughly 70 using the exact method as opposed to the bisection method using a 64-bit Intel T7400, 2.17GHz Core-2 Duo processor.

observations weakly dominate the point (8.8, 4). Three of these dominating points—at (4, 4), (5, 4.2), and the original point at (8, 4.4)—lie on the DEA estimate of the frontier $\mathcal{P}^{\partial}$, while the fourth lies at (6, 4).

Note that the original observation (x_0, y_0) at (8, 4.4) is not dominated by any other observations in the sample. However, by projecting this point onto the α-quantile at the point represented by the open circle, three other observations are found that produce *almost* as much output as the DMU of interest, but use considerably less input. Thus, the manager of the DMU at (8, 4.4) might obtain useful information for managing his own DMU by studying the operations of the three DMUs at (4, 4), (5, 4.2), and (6, 4).[13]

Of course, with only two dimensions, one could always plot the data as in Figure 2. But with multiple inputs and multiple outputs, visualizing the data is problematic. DEA yields estimates of efficiency, but little else. For an inefficient DMU, one could look for DMUs that weakly dominate the DMU of interest to find relevant practices that might be emulated to improve efficiency, but with high dimensions, it is sometimes the case that there are few, if any, dominance relationships in the data.[14] The example illustrated by Figure 2 shows how the hyperbolic α-quantile estimator reveals relevant, practically useful comparisons among DMUs, even for DMUs that DEA would indicate are ostensibly efficient.

By contrast, the estimators described by Daouia and Simar (2007) of the conditional α-quantile distance functions in (3.7) and 3.8) yield $\widehat{\theta}_{\alpha,20}(x_0, y_0) = 1$ and $\widehat{\lambda}_{\alpha,20}(x_0, y_0) = 1.1$ given the observations shown in Figure 2 and $\alpha = 0.8$ (by coincidence, $\widehat{\lambda}_{\alpha,20}(x_0, y_0) = \widehat{\gamma}_{\alpha,20}(x_0, y_0)^{-1}$). Thus the input-oriented, conditional α-quantile partial frontier $\mathcal{P}^{\partial}_{x,\alpha}$ corre-

[13] Perhaps the DMU at (8, 4.4) is operating at an inefficient scale size, beyond the region of CRS. Or, because DEA estimates of the true frontier $\mathcal{P}^{\partial}$ are biased downward, i.e., $\widehat{\mathcal{P}}_{\text{DEA}} \subseteq \mathcal{P}$ (see Simar and Wilson, 2000a, for discussion), it may well be possible for the DMU at (8,4.4) to significantly reduce its input-usage without reducing output. It is important to remember that the observations on the piecewise-linear DEA frontier in Figure 2 are only *ostensibly* efficient. The DEA frontier estimate is nothing more than a *biased* estimate of the true, but unobserved, frontier $\mathcal{P}^{\partial}$. Using DEA and the bootstrap methods of Simar and Wilson (1998, 2000a) would necessarily yield estimates of confidence intervals for $\theta(x_0, y_0 \mid \mathcal{P})$ and $\lambda(x_0, y_0 \mid \mathcal{P})$ defined in (2.2)–(2.3) that do not include 1; e.g., the lower bound of an estimated confidence interval for the input-distance function $\theta(x_0, y_0 \mid \mathcal{P})$ will necessarily lie to the right of 1. This reflects the fact that the true production set boundary, $\mathcal{P}^{\partial}$, necessarily lies to the left and above the piecewise-linear DEA frontier estimate shown in Figure 2.

[14] Wheelock and Wilson (2003) analyzed three cross-sections of observations on U.S. commercial banks in 1984, 1993, and 2002, with $p = q = 5$, and 13,845, 10,661, and 7,561 observations (respectively). In each cross section, *all* observations were found to lie on the FDH frontier estimate. Consequently, there is no observation in any of the three cross sections that is dominated by another observation in the same cross section.

sponds with the DEA estimator of the full frontier at the point (8, 4.4), while the output-oriented conditional α-quantile partial frontier $\mathcal{P}^{\partial}_{y,\alpha}$ lies below the DEA frontier estimator at (8, 4.4), as is apparent from (3.10) and the earlier discussion regarding Figure 1.

4 Asymptotic Results

Park et al. (2000) derive asymptotic results for FDH distance function estimators of the input- and output-oriented distance functions defined in (2.2) and (2.3). Here, some similar results are obtained for the FDH estimator of the hyperbolic distance function $\gamma(\boldsymbol{x}_0, \boldsymbol{y}_0)$ using techniques similar to those used by Park et al.; these results are useful for establishing properties of the hyperbolic α-quantile distance function estimator. Proofs are given in the Appendix.

Theorem 4.1. *Under Assumptions 3.1–3.4, for all* $\epsilon > 0$,

$$\Pr\left[n^{1/(p+q)}\left(\gamma(\boldsymbol{x},\boldsymbol{y}) - \widehat{\gamma}_n(\boldsymbol{x},\boldsymbol{y})\right) \leq \epsilon\right] = 1 - \exp\left[-\left(\mu_{\mathcal{H},0}\epsilon\right)^{p+q}\right] + o(1)$$

where $\mu_{\mathcal{H},0}$ *is a constant given in Definition A.2 in the Appendix.*

Analogous to Corollary 3.2 in Park et al. (2000), the next results follow directly from Theorem 4.1; in particular, (ii) is obtained by exploiting the relation between exponential and Weibull distributions.

Corollary 4.1. *Under Assumptions 3.1–3.4, (i)* $\widehat{\gamma}_n(\boldsymbol{x},\boldsymbol{y})$ *converges to* $\gamma(\boldsymbol{x},\boldsymbol{y})$ *with* $O_p\left(n^{-1/(p+q)}\right)$*, and there is no* $\delta > \frac{1}{p+q}$ *such that* $\gamma(\boldsymbol{x},\boldsymbol{y}) - \widehat{\gamma}_n(\boldsymbol{x},\boldsymbol{y}) = O_p\left(n^{-\delta}\right)$*; and (ii)* $n^{1/(p+q)}\left(\gamma(\boldsymbol{x},\boldsymbol{y}) - \widehat{\gamma}_n(\boldsymbol{x},\boldsymbol{y})\right) \xrightarrow{d}$ *Weibull*$\left(\mu^{p+q}_{\mathcal{H},0},\ p+q\right)$.

These results establish that the FDH estimator $\widehat{\gamma}_n(\boldsymbol{x},\boldsymbol{y})$ of the hyperbolic distance function defined in (2.4) is consistent, converges at the rate $n^{-1/(p+q)}$, and asymptotically is distributed Weibull, confirming the earlier speculation in footnote 9. Asymptotic results in Daouia and Simar (2007) for the conditional α-quantile estimators can be extended to the case of the hyperbolic quantile estimator $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y})$ by adapting their analysis to allow both input and output quantities to be adjusted.

Theorem 4.2. *For* $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{P}$, $\alpha \in (0,1]$, *and for every* $\epsilon > 0$,

$$\Pr\left(\left|\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) - \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})\right| > \epsilon\right) \leq 2e^{-2n\omega^2/(2-\alpha)^2} \quad \forall\, n \geq 1 \tag{4.14}$$

where

$$\omega = H\left((\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})-\epsilon)^{-1}\boldsymbol{x}, (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})-\epsilon)\boldsymbol{y}\right). \tag{4.15}$$

Theorem 4.2 implies that $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y})$ converges completely to $\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})$, denoted by $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) \xrightarrow{c} \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})$. Hence, $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y})$ is a strongly consistent estimator of $\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})$.[15]

Theorem 4.3. *In addition to Assumptions 3.1–3.4, assume $H(\gamma^{-1}\boldsymbol{x},\gamma\boldsymbol{y})$ is differentiable with respect to γ near $\gamma = \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})$. Then*

$$\sqrt{n}\left(\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) - \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})\right) \xrightarrow{d} N\left(0, \sigma_\alpha^2(\boldsymbol{x},\boldsymbol{y})\right) \tag{4.16}$$

where

$$\sigma_\alpha(\boldsymbol{x},\boldsymbol{y})^2 = \alpha(1-\alpha)\left[\frac{\partial H\left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})^{-1}\boldsymbol{x}, \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})\boldsymbol{y}\right)}{\partial \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})}\right]^{-2}. \tag{4.17}$$

Thus, the estimator $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y})$ of the hyperbolic distance function $\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})$ is asymptotically normal, with convergence at the classical, parametric rate of $O\left(n^{-1/2}\right)$. Consequently, the hyperbolic efficiency estimator $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y})$ does not suffer the ill-effects of the curse-of-dimensionality that plagues DEA and FDH estimators, since its convergence rate depends solely on the sample size n and involves neither p nor q. These results are not surprising, given that similar results were obtained by Daouia and Simar (2007) for estimators of the input and output conditional α-quantile distance functions defined in (3.7)–(3.8).

In principle, the asymptotic distributional results in (4.16)–(4.17) could be used for statistical inference-making when $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y})$ is used to estimate $\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})$ for $\alpha \in (0,1)$. However, this would require estimation of the derivative of the probability function $H(\cdot,\cdot)$, which seems difficult. Alternatively, it is easy to implement a bootstrap procedure for purposes of testing and inference. Since quantiles (rather than the boundary of the production set) are estimated, a naive bootstrap based on resampling from the empirical distribution of the inputs and outputs could be implemented for inference-making purposes. However, although valid asymptotically, the naive bootstrap may result in poor coverages because of the discrete nature of data in small samples. In Section 6 below we implement a smooth bootstrap procedure that overcomes this problem.

[15] A sequence of random variables $\{\zeta_n\}_{n=1}^{\infty}$ converges completely to a random variable ζ, denoted by $\zeta_n \xrightarrow{c} \zeta$, if $\lim_{n\to\infty}\sum_{j=1}^{n}\Pr\left(|\zeta_j - \zeta| \geq \epsilon\right) < \infty \ \forall\ \epsilon > 0$. This type of convergence was introduced by Hsu and Robbins (1947). Complete convergence implies, and is a stronger form of convergence than almost-sure convergence.

Daouia and Simar (2007) show that for their conditional α-quantile distance function estimators, letting α be a sequence in n and allowing $\alpha \to 1$ as $n \to \infty$ allows their estimators to be interpreted as robust estimators of distance to the full frontier $\mathcal{P}^{\partial}$, rather than of distance to the conditional α-quantiles $\mathcal{P}^{\partial}_{x,\alpha}$ and $\mathcal{P}^{\partial}_{y,\alpha}$. As shown next, similar results hold for the hyperbolic α-quantile distance function estimator.

Lemma 4.1. *Under Assumptions 3.2–3.4, for any* $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$,

$$n^{1/(p+q)} \left(\widehat{\gamma}_n(\boldsymbol{x}, \boldsymbol{y}) - \widehat{\gamma}_{\alpha(n),n}(\boldsymbol{x}, \boldsymbol{y})\right) \xrightarrow{a.s.} 0$$

as $n \to \infty$*, where the order of* $\alpha(n) > 0$ *is such that* $n^{(p+q+1)/(p+q)}(1 - \alpha(n)) \to 0$ *as* $n \to \infty$.

Analogous to the decomposition given in Daouia and Simar (2007), it is clear that

$$\begin{aligned} n^{1/(p+q)} \left(\gamma(\boldsymbol{x}, \boldsymbol{y}) - \widehat{\gamma}_{\alpha(n),n}(\boldsymbol{x}, \boldsymbol{y})\right) &= n^{1/(p+q)} \left(\widehat{\gamma}(\boldsymbol{x}, \boldsymbol{y}) - \widehat{\gamma}_n(\boldsymbol{x}, \boldsymbol{y})\right) \\ &\quad + n^{1/(p+q)} \left(\widehat{\gamma}_n(\boldsymbol{x}, \boldsymbol{y}) - \widehat{\gamma}_{\alpha(n),n}(\boldsymbol{x}, \boldsymbol{y})\right). \end{aligned} \tag{4.18}$$

Applying Lemma 4.1 and Corollary 4.1, part (ii) to (4.18) yields the following result.

Theorem 4.4. *Under Assumptions 3.2–3.4 and with the order of* $\alpha(n) > 0$ *such that* $n^{(p+q+1)/(p+q)}(1 - \alpha(n)) \to 0$ *as* $n \to \infty$*, for any* $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$,

$$n^{1/(p+q)} \left(\gamma(\boldsymbol{x}, \boldsymbol{y}) - \widehat{\gamma}_{\alpha(n),n}(\boldsymbol{x}, \boldsymbol{y})\right) \xrightarrow{d} \text{Weibull}\left(\mu_{\mathcal{H},0}^{p+q}, p+q\right)$$

where $\mu_{\mathcal{H},0}$ *is a constant.*

The constant $\mu_{\mathcal{H},0}$ that appears here is the same as the one appearing in Theorem 4.1, and as noted earlier, an expression for $\mu_{\mathcal{H},0}$ is given in the Appendix.

To summarize, the results obtained in this section reveal that as with the conditional α-quantile estimators of Daouia and Simar (2007), the unconditional, hyperbolic α-quantile estimator has two interpretations—as an estimator of distance to a partial frontier (when α is fixed), or as an estimator of distance to the full frontier (when α is a sequence in n of appropriate order). In the first case, the estimator is root-n consistent and asymptotically normal. In the latter case, the root-n convergence and asymptotic normality properties in Theorem 4.3 are lost, and the curse of dimensionality re-appears. There are no free lunches here. However, compared to the FDH estimator $\widehat{\gamma}_n(\boldsymbol{x}, \boldsymbol{y})$ of $\gamma(\boldsymbol{x}, \boldsymbol{y})$, the estimator $\widehat{\gamma}_{\alpha(n),n}(\boldsymbol{x}, \boldsymbol{y})$ will be far less sensitive to outliers. Moreover, under either interpretation, the hyperbolic estimator avoids problem no. 3 listed near the end of Section 2.

5 Federal Reserve Check Processing

We use the methods described above to examine the efficiency of Federal Reserve check-processing operations. Roughly half of all checks written in the United States are cashed or deposited at banks other than those upon which they are drawn. Many of these checks are processed by the Federal Reserve, the world's largest volume processor. In 2003, Federal Reserve offices processed 15.8 billion checks (Board of Governors of the Federal Reserve System, 2003, p. 118). Check volume has declined since 1999, when Federal Reserve offices processed 17.1 billion checks, as electronic payments media have become increasingly popular. The decline in volume has put considerable pressure on the Fed to reduce costs, and has increased interest in the efficiency with which the Fed provides payments services. The methods we propose could be used to rank offices in terms of efficiency and to help check office managers identify peer offices whose operations they might wish to study for ideas about how to improve the efficiency of their own facilities.

Recent studies of the Fed's efficiency conclude that Federal Reserve check operations suffer from considerable cost, or "x-", inefficiency. Using quarterly data for 1979-90, and both parametric and non-parametric methods, Bauer and Hancock (1993) found evidence of considerable cost inefficiency at Fed check offices both before and after the Fed began to charge for payments services in 1982. In a related study, Bauer and Ferrier (1996) found both high average cost inefficiency and considerable dispersion of inefficiency across Fed offices during 1990-94. Further, Bauer and Ferrier (1996) detected evidence of technological regress in check processing during the early 1990s, which they associated with declining processing volume at some sites, migration of "high-quality" check business (e.g., Social Security and payroll checks) to electronic media, and the implementation of new check services (e.g., application of magnetic ink character recognition to paper checks).

Gilbert et al. (2004) used DEA to investigate the productivity of Fed check offices by pooling observations across time and estimating a single production frontier for the entire sample. That study found that the median Fed office became more productive during the 1990s, after experiencing a decline in productivity during the 1980s. The authors did not investigate efficiency, however, because their sample size was too small to yield reliable estimates of the production frontier for each period.

Wheelock and Wilson (2004) applied the "order-m" estimator of Cazals et al. (2002) to estimate the input-technical efficiency of Fed check offices. Unlike DEA, the order-m estimator has a root-n convergence rate when used to estimate distance to a partial frontier, and thus is useful for small sample applications. Wheelock and Wilson (2004) found evidence of a small improvement in the efficiency of the median Fed office during the 1990s, but a decrease in efficiency and increased dispersion across Fed offices during 2000-03. Wheelock and Wilson (2004) did not compare efficiency levels across offices, nor identify relevant peer groups for individual offices.

The processing of checks is a standard logistic operation. "Forward item" processing involves receiving checks from depositing banks, sorting them, crediting the accounts of the depositing banks, and delivering the checks to the banks upon which they are drawn. Some Fed offices process Federal Government checks and postal money orders, as well as commercial checks. Federal Reserve offices also process "return items" (which include checks returned on account of insufficient funds) and provide various electronic check services, such as imaging and truncation. Fed check offices also incur costs in making adjustments necessitated by processing and other errors. Following the convention of other studies, e.g., Bauer and Hancock (1993), Bauer and Ferrier (1996), Gilbert et al. (2004), and Wheelock and Wilson (2004), we focus here on the forward processing of commercial and Federal Government check items.

Gilbert et al. (2004) and Wheelock and Wilson (2004) model Federal Reserve check processing as consisting of two jointly-produced outputs measured by (i) the number of forward items processed and (ii) the number of endpoints served. An endpoint is an office of a depository institution to which a Fed office delivers checks. The number of endpoints is a measure of the level of service provided by a check office — an office serving many endpoints, all else equal, provides a higher level of service than an office serving fewer endpoints. In this sense, check processing is analogous to the delivery of mail by a post office. The output of a post office is not simply the number of items it delivers, but also the number of addresses to which it delivers mail. Presumably, a post office that delivers mail to a single address provides less service than a post office that delivers an identical quantity of mail to several addresses.[16]

Federal Reserve check facilities use a variety of inputs to process checks and deliver them to paying banks. We follow Bauer and Hancock (1993) and subsequent studies by defining four distinct categories of inputs used in the processing of forward items: (1) personnel, (2) materials, software, equipment and support, (3) transit services, and (4) facilities. Estimation of technical efficiency requires estimates of the physical quantities used of each input, rather than total expenditures. Table 1 describes our method of constructing measures of the four inputs for each Fed check office using expense data for forward items processing. Our sample consists of quarterly data on all Fed offices that processed checks at any time between 1980 and 2003. Offices that did not process in a particular quarter, or for which data are unavailable, are omitted from estimation only for that quarter. Table 2 gives summary statistics for both inputs and outputs observed in the third quarter of the first, middle and ending years of our sample. Because of significant changes in the Fed's accounting system after 1994 and 2000, data and results are not directly comparable across the three periods 1980–1994, 1995–2000, and 2001–2003.

6 Estimation Results

First, we use the DEA and FDH estimators to estimate technical efficiency in both the input and output orientations for each Fed check office in each period of our sample. We use observations on every office operating in a given quarter (45–48 offices depending on the quarter) to estimate the production frontier $\mathcal{P}^{\partial}$ for that quarter. DEA estimates suggest a moderate level of technical inefficiency among Fed check-processing offices over the period of our study; in the input orientation, mean technical efficiency estimates for each of 95 quarters range from 1.0 to 1.431, with an overall mean of 1.118. Hence, we estimate an average level of inefficiency of 11.8 percent. Results are similar in the output orientation: mean inverse (output) technical efficiency estimates for each period range from 1.0 to 1.373, with an overall mean of 1.071, implying an average level of inefficiency of 7.1 percent (here and throughout, we report inverses of output-oriented efficiency estimates, so that in all orientations, increasing values of efficiency estimates correspond to increasing *in*efficiency).

[16] Gilbert et al. (2004) present a statistical test of the hypothesis that the number of endpoints served by a Fed check office constitutes a distinct output.

By contrast, when we use the FDH estimator, we estimate average inefficiencies of 1.0170 and 1.019 in the input and output orientations, respectively, or less than 2 percent inefficiency. Recall that the DEA estimator of $\mathcal{P}$ defined in (2.6) is merely the convex hull of the FDH estimator defined in (2.5). Hence the assumption of convexity accounts for approximately 75 percent of the apparent inefficiency detected by the DEA estimation. This assumption might be reasonable, but its large impact on estimates of efficiency could give one pause.

Our principal reason for doubting either the DEA or FDH efficiency estimates, however, is the relatively small size of our sample and the slow convergence rates of the DEA and FDH estimators. We simply have too few observations to obtain statistically meaningful estimates of technical efficiency using either estimator. This is a common problem in applications; however, in many cases the problem has been ignored.

The hyperbolic quantile estimator discussed in Section 3.2, by contrast, has a root-n convergence rate when used to estimate distance to a partial frontier and, unlike DEA, does not impose convexity. We use the estimator to obtain estimates of technical inefficiency for every Fed office operating in each quarter. Of course, specific estimates depend on the choice of α, which determines the location of the unique α–quantile frontier. In general, the larger the value of α, the larger the estimates $\widehat{\gamma}$. Figure 3 plots our hyperbolic α-quantile efficiency estimates for four values of α—0.8, 0.85, 0.9, and 0.95—against each other; each panel compares estimates for a pair of values for α. The similar ranking of offices across different values of α is reflected in the fact that most points fall on or near a straight line. For $\alpha = 0.8$, 0.85, or 0.9, the rankings are broadly similar, and somewhat similar for $\alpha = 0.95$ Since our sample size is rather small ($n = 44$) the estimates begin to show some instability as α approaches 1. At $\alpha = 0.95$, only three observations (the original and two others) will weakly dominate the projection of a particular observation onto the α-quantile frontier. Similar results were obtained for other quarters represented in our data. As with the example provided by Daouia and Simar (2007) using data on mutual funds to estimate conditional α-quantile efficiencies, the results seem to be rather robust with respect to the choice of α.

Results for the third quarter of 2003, the last quarter in our sample, are presented in Table 3, alongside corresponding DEA and FDH estimates for comparison. We have sorted offices in order from most to least efficient according to point estimates of efficiency ($\widehat{\gamma}$) for

α=0.9, and report both point estimates and 95-percent confidence intervals obtained using a smooth bootstrap.[17]

Our estimates reveal considerable variation across offices in estimated efficiency. We estimate that the most efficient office used just 34 percent of the input amount and produced nearly three times (1/0.34) more output than an office (perhaps hypothetical) located on the α=0.9 quantile frontier along a hyperbolic path from the first office. The least efficient office in that quarter, by contrast, used 105.8 percent of the input and produced $100/1.0584 = 94.5$ percent of the output produced by an office on the α-quantile frontier. The 95 percent confidence intervals indicate that differences in efficiency estimates across many offices are statistically significant.

We find a correspondence between the rankings of offices by α-quantile, DEA and FDH efficiency estimates. In general, offices found to be least efficient, as indicated by input or inverse output distance function estimates greater than 1.0, using either the DEA or FDH estimators correspond to offices identified as least efficient by our quantile estimator. Many of the DEA and FDH estimates for individual offices are equal to 1.0, however, indicating that the office is located on the estimated frontier and, thus, estimated to have no inefficiency. Across the 4,405 quarterly observations on Fed check offices, 50.1 percent of all DEA distance function estimates and 88.2 percent of FDH estimates are equal to 1.0.[18] Such large numbers of DMUs located on the DEA and FDH frontiers reflects the relatively small size of our sample as compared to the number of inputs and outputs in our check processing model. In contrast to hyperbolic α-quantile estimation, DEA and FDH provide only a partial ordering of offices

[17] In principle, since the boundary of support of $f(\boldsymbol{x}, \boldsymbol{y})$ is not estimated, a naive bootstrap (where bootstrap samples are drawn from the empirical distribution function of the sample data) could be used to construct confidence interval estimates. With very small samples, as in the present case, however, the discrete nature of the observed data cause the naive bootstrap to yield bootstrap distributions that are also discrete. Although the problem disappears asymptotically, using a smooth bootstrap, where pseudo values γ_i^* are drawn from a kernel density estimate of the original distance function estimates $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $i = 1, \ldots, n$ seems to give more sensible results. To implement the smooth bootstrap, each observation $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ is projected onto the estimated α-quantile and then randomly projected away form the α-quantile by computing $(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*) = \left(\gamma_i^* \widehat{\gamma}_{\alpha,n}(\boldsymbol{x}_i, \boldsymbol{y}_i)^{-1}\boldsymbol{x},\ \gamma_i^{*-1}\widehat{\gamma}_{\alpha,n}(\boldsymbol{x}_i, \boldsymbol{y}_i)\boldsymbol{y}\right)$ for each $i = 1, \ldots, n$. Then bootstrap distance function estimates for each of the original sample observations are computed relative to the α-quantile of the bootstrap data $(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)$.

[18] The convexity assumption of the DEA estimator explains why there are fewer offices located on the DEA frontier than on the FDH frontier. Because the FDH frontier necessarily lies above the α-quantile frontier, the high percentage of offices located on the FDH frontier explains why the distance function estimates $\widehat{\gamma}$ are less than 1.0 for most offices.

in terms of efficiency.[19] Moreover, as noted above, although the quantile efficiency results presented in Table 3 are for a value of $\alpha = 0.9$, we obtained similar results for different values of α.

Table 4 presents efficiency estimates for the 44 offices represented in Table 3 in the same period, namely the third quarter of 2003. Table 4 contains 2 groups of three columns, corresponding to $\alpha = 0.9$ and $\alpha = 0.95$; the three columns within each group give estimates of the hyperbolic α-quantile distance function, the input conditional order-α quantile distance function, and the inverse of the output conditional order-α quantile distance function. Hence, for $\alpha = 0.9$, the column labeled "$\widehat{\gamma}_\alpha$" in Table 4 is identical to the similarly labeled column in Table 3. Comparing the hyperbolic estimates corresponding to the two values of α confirms the impression from Figure 3; i.e., while there are some changes in the rankings, the results obtained with the two values of α are broadly similar. However, when comparing either of the sets of hyperbolic estimates with estimates of the input and output conditional α-quantile estimates, one sees substantial differences. With $\alpha = 0.9$, a number of the conditional estimates are equal to unity, while an even greater number are equal to unity when $\alpha = 0.95$. Moreover, it is apparent that in many cases the conditional estimates are sensitive to the choice of input or output orientation, as one might expect given the discussion in Section 2.

Figures 4–6 plot the mean, median and variance of hyperbolic α-quantile efficiency estimates (with $\alpha = 0.9$) across offices by quarter from 1980:1 to 2003:3. Similar to the findings obtained by Gilbert et al. (2004) for productivity, and Wheelock and Wilson (2004) for input-technical efficiency, our estimates indicate an increase in mean and median inefficiency after Fed offices began to charge for payments services in 1982. Inefficiency continued to increase until about 1990, then it declined until about 1993, and finally increased from the mid-1990s through the end of our sample. Variation across offices was relatively high when the pricing regime was first implemented, relatively low in the late 1990s, and then high again during the period of declining volume after 1999.

[19]The failure of DEA and FDH to provide a complete ranking has long been noted in the literature. Andersen and Petersen (1993) proposed a "super-efficiency" score, based on omitting the observation under consideration from the reference set when computing a DEA efficiency score for a particular observation. Although this approach provides a complete ranking, it confounds estimation and interpretation of the results by implicitly assuming a different technology for each sample observation.

7 Policy Recommendations

The estimates reported in Table 3 and illustrated in Figures 4–6 indicate considerable variation across Federal Reserve offices in terms of technical efficiency. Because our methods provide a ranking of offices in terms of efficiency, as well as an indication of whether differences across offices are statistically significant, System policymakers could use our estimator to identify high- or low-performing offices that might require additional attention.

Managers of individual offices might find it more useful to use our methods to identify high-performing peer offices with characteristics that are similar to those of the manager's own office. Such offices would probably be more relevant for the manager to study than other high performing offices that operate at considerably different scale or mix of inputs and outputs than the manager's own office.

Recalling the example from Section 3.3, it is easy to use our quantile estimator to identify relevant peers for comparison with a particular office. Peers consist of all offices that weakly dominate in terms of efficiency the projection of a given office onto the α-quantile frontier.

Table 5 presents a list of peer offices for each Fed check processing site, where sites are listed from most to least efficient based on estimation of the α=0.9 quantile using data for 2003:3, i.e., in the same order as in Table 3. Typically, the number of peers will equal $(1-\alpha)n$, i.e., $(1-0.9)(44) = 4.4$ offices. We identify five peers for three Fed offices, however, because a few offices produce equal amounts of one output (number of endpoints). We also list in the table the number of times a given office appears in the relevant peer group of other offices. For example, #11 is among the peers for 16 offices. Not surprisingly, more efficient offices tend to be in relevant peer groups more frequently than less efficient offices. Relatively inefficient offices rarely dominate the projections of other offices onto the α-quantile frontier.

Our main policy recommendation is that policymakers and managers of entities such as Federal Reserve check facilities use techniques such as those presented in this paper to identify differences in performance across decision-making units and for choosing relevant peer groups for individual DMUs. Although we focus exclusively on technical efficiency in this paper, the methods proposed here could be applied easily to examine other types of efficiency to obtain a more complete picture of performance. Used in this way, statistical analysis can assist policymakers and managers of individual offices in identifying operations that merit more in-depth study, and help guide the allocation of scarce public resources.

8 Summary and Conclusions

This paper proposes a new unconditional quantile estimator for estimating the efficiency of individual firms or offices that use a common set of inputs to produce a common set of outputs. Like DEA and FDH estimators, our quantile estimator is fully non-parametric. However, unlike DEA and FDH, our estimator is robust with respect to outliers in the data, and when used to estimate distance to a partial frontier, is strongly consistent with the rapid root-n convergence rate typical of parametric estimators.

Although our quantile estimator has some similarities with the *conditional* α quantile estimators of Daouia (2003), Aragon et al. (2005) and Daouia and Simar (2007), we avoid problems in comparing efficiency across DMUs of different sizes that arise with the use of those estimators. Furthermore, our approach enables the identification of relevant peer groups that managers of individual DMUs might study to learn how to improve their own operations.

We apply our methods to study the efficiency of Federal Reserve System facilities that process paper checks. Declining check volume has put considerable pressure on Fed offices to reduce costs and inefficiencies, and has caused the Fed to eliminate check processing operations at some of its facilities. Further volume declines are anticipated, and thus policymakers will likely be forced to consider further eliminations. The methods proposed here for estimating efficiency could be used to help identify best practices as well as under-performing offices. Although we examine only technical efficiency in this paper, our methods could also be used to examine other types of efficiency, such as cost or scale efficiency, to obtain a more complete picture of the performance of individual offices. Further, the methods are applicable not only to investigating the efficiency of check processing operations, but to any setting in which policymakers or operations managers are interested in evaluating the performance of individual facilities or comparing performance across facilities within an organization.

A Appendix

The terms defined next are needed in the proofs of Lemma A.1 and Theorem 4.1.

Definition A.1. *For* $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{P}$,

$$\mathcal{H}(\boldsymbol{x}_0, \boldsymbol{y}_0) \equiv \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{p+q} \mid \boldsymbol{x} \leq \boldsymbol{x}_0,\ \boldsymbol{y} \geq \boldsymbol{y}_0\} \bigcap \mathcal{P}, \tag{A.1}$$

$$\mathcal{H}_0(\xi) \equiv \mathcal{H}\left(\boldsymbol{x}_0^\partial (1-\xi)^{-1},\ \boldsymbol{y}_0^\partial (1-\xi)\right) \ \textit{for}\ \xi > 0, \tag{A.2}$$

$$(\boldsymbol{x}_0^\partial, \boldsymbol{y}_0^\partial) \equiv \left(\gamma(\boldsymbol{x}_0, \boldsymbol{y}_0)^{-1}\boldsymbol{x}_0, \gamma(\boldsymbol{x}_0, \boldsymbol{y}_0)\boldsymbol{y}_0\right) \in \mathcal{P}^\partial. \tag{A.3}$$

Definition A.2.

$$g_1 \equiv \prod_{k=1}^{p} \frac{\partial}{\partial x^k} g(\boldsymbol{x}, \boldsymbol{y}^{(q)})|_{(\boldsymbol{x}, \boldsymbol{y}^{(q)}) = (\boldsymbol{x}_0^\partial, \boldsymbol{y}_0^{\partial(q)})} \prod_{\ell=1}^{q} \frac{\partial}{\partial y^\ell} g(\boldsymbol{x}, \boldsymbol{y}^{(q)})|_{(\boldsymbol{x}, \boldsymbol{y}^{(q)}) = (\boldsymbol{x}_0^\partial, \boldsymbol{y}_0^{\partial(q)})}, \tag{A.4}$$

$$\overline{x}_0 \equiv \sum_{k=1}^{p} \frac{\partial}{\partial x^k} g(\boldsymbol{x}, \boldsymbol{y}^{(q)})|_{(\boldsymbol{x}, \boldsymbol{y}^{(q)}) = (\boldsymbol{x}_0^\partial, \boldsymbol{y}_0^{\partial(q)})} x_0^{\partial k}, \tag{A.5}$$

$$\overline{y}_0 \equiv \sum_{\ell=1}^{q-1} \frac{\partial}{\partial y^\ell} g(\boldsymbol{x}, \boldsymbol{y}^{(q)})|_{(\boldsymbol{x}, \boldsymbol{y}^{(q)}) = (\boldsymbol{x}_0^\partial, \boldsymbol{y}_0^{\partial(q)})} y_0^\ell + y_0^q, \tag{A.6}$$

$$\mu_{\mathcal{H},0} \equiv \left(\frac{f_0}{|g_1|} \frac{\overline{x}_0^p \overline{y}_0^q}{(p+q)!}\right)^{1/(p+q)}. \tag{A.7}$$

The following lemma is an analog to Lemma A.3 in Park et al. (2000), and is needed for the proof of Theorem 4.1.

Lemma A.1. *For any sequence* ξ *that goes to zero,*

$$\rho_{\mathcal{H},0}(\xi) = (\mu_{\mathcal{H},0}\xi)^{p+q} + O\left(\xi^{p+q+1}\right).$$

Proof. By Assumption 3.3,

$$\rho_{\mathcal{H},0}(\xi) = f_0 + O(\xi))\mathcal{L}[\mathcal{H}_0(\xi) \bigcap \mathcal{P}] \tag{A.8}$$

where $f_0 = f(\boldsymbol{x}_0^\partial, \boldsymbol{y}_0^\partial)$ and $\mathcal{L}$ denotes the Lebesgue measure. Since $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$ iff $\boldsymbol{y}^q \leq g(\boldsymbol{x}, \boldsymbol{y}^{(q)})$,

$$\mathcal{L}[\mathcal{H}_0(\xi) \bigcap \mathcal{P}] = \int_\Omega I\left(y^q \leq g(\boldsymbol{x}, \boldsymbol{y}^{(q)})\right) dx\, dy$$

where $I(\cdot)$ is the indicator function and $\Omega = (-\infty, \boldsymbol{x}_0^\partial(1-\xi)^{-1}] \times [\boldsymbol{y}_0^\partial(1-\xi), \infty)$. Let $d_0(\boldsymbol{x}, \boldsymbol{y}) = \left((\boldsymbol{x} - \boldsymbol{x}_0^\partial),\ (\boldsymbol{y} - \boldsymbol{y}_0^\partial)^{(q)}\right)$ and $\nabla g_0 = \nabla g(\boldsymbol{x}_0^\partial, \boldsymbol{y}_0^{\partial(q)})$. Note $g(\boldsymbol{x}_0^\partial, \boldsymbol{y}_0^{\partial(q)}) = y_0^{\partial q}$ and $\|d_0(\boldsymbol{x}, \boldsymbol{y})\| = O(\xi)$ for any $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{H}_0(\xi) \bigcap \mathcal{P}$. By Assumption 3.4,

$$g(\boldsymbol{x}, \boldsymbol{y}^{(q)}) = y_0^{\partial q} + \nabla g_0' d_0(\boldsymbol{x}, \boldsymbol{y}) + O(\xi^2).$$

Hence

$$\mathcal{L}[\mathcal{H}_0(\xi)\bigcap\mathcal{P}] = \int_\Omega I\left[\nabla g_0' d_0(\boldsymbol{x},\boldsymbol{y}) \geq y^q - y_0^{\partial q} + O(\xi^2)\right]\, dx\, dy. \tag{A.9}$$

Now transform the coordinate system by shifting the origin to $\left(\boldsymbol{x}_0^{\partial}(1-\xi)^{-1},\ \boldsymbol{y}_0^{\partial}(1-\xi)\right)$ and reversing the direction of the input axes; in addition, rescale the axes by the absolute values of the corresponding partial derivatives. In order to maintain an additive structure, let $1-\xi = 1+\eta \geq 1$ for $\xi > 0$, implying $\eta = \frac{\xi}{1-\xi} > 0$ and $O(\xi) = O(\eta)$. Specifically, (i) transform y^ℓ to $\widetilde{y}^\ell = |\frac{\partial}{\partial y^\ell} g(\boldsymbol{x},\boldsymbol{y}^{(q)})|_{(\boldsymbol{x},\boldsymbol{y}^{(q)})=(\boldsymbol{x}_0^{\partial},\boldsymbol{y}_0^{\partial(q)})}\left(y^\ell - y^{\partial\ell}(1-\xi)\right)\ \forall\ 1 \leq \ell \leq q-1$; (ii) transform y^q to $\widetilde{y}^q = y^q - y_0^{\partial q}(1-\xi)$; and (iii) transform x^k to $\widetilde{x}^k = \frac{\partial}{\partial x^k} g(\boldsymbol{x},\boldsymbol{y}^{(q)})|_{(\boldsymbol{x},\boldsymbol{y}^{(q)})}\left(x^{\partial k}(1+\eta) - x^k\right)\ \forall\ 1 \leq k \leq p$ (note that in the right-hand side of the last expression, the x^k axis has been "flipped," so that the orientation of $\widetilde{x}^k$ is opposite that of x^k). Then (A.9) can be rewritten as

$$\mathcal{L}[\mathcal{H}_0(\xi)\bigcap\mathcal{P}] = |g_1|^{-1}\int_{[0,\infty)^{p+q}} I\left[\sum_{k=1}^{p}\widetilde{x}^k + \sum_{\ell=1}^{q}\widetilde{y}^\ell \leq \xi(\overline{y}_0 + (1-\xi)^{-1}\overline{x}_0 + O(\xi))\right] d\widetilde{x}\, d\widetilde{y} \tag{A.10}$$

where expressions for g_1, $\overline{x}_0$, and $\overline{y}_0$ are given in Definition A.2.

The integral in (A.11) is the volume of a $(p+q)$-dimensional simplex with edges of length $\xi(\overline{y}_0 + (1-\xi)^{-1}\overline{x}_0 + O(\xi))$, or equivalently, $\xi(\overline{y}_0 + O(\xi)) + \eta\overline{x}_0$. The result follows from this and (A.8). □

Proof of Theorem 4.1. Consider $(\boldsymbol{x}_0,\boldsymbol{y}_0) \in \mathcal{P}$ and let $\gamma_0 = \gamma(\boldsymbol{x}_0,\boldsymbol{y}_0)$ and $\widehat{\gamma}_n = \widehat{\gamma}_n(\boldsymbol{x}_0,\boldsymbol{y}_0)$. By construction, $\gamma_0 \geq \widehat{\gamma}_0$ and hence $\Pr(\gamma_0 - \widehat{\gamma}_0 \leq \epsilon) = 0\ \forall\ \epsilon \leq 0$. For $\epsilon > 0$, necessarily

$$\begin{aligned}
\Pr(\gamma_0 - \widehat{\gamma}_0 \leq \epsilon) &= \Pr(-\widehat{\gamma}_0 \leq \epsilon - \gamma_0)\\
&= \Pr(\widehat{\gamma}_0 \geq \gamma_0 - \epsilon)\\
&= 1 - \Pr(\widehat{\gamma}_0 < \gamma_0 - \epsilon)\\
&= 1 - \Pr(\omega_{(n)} < \gamma_0 - \epsilon)\\
&= 1 - \left[\Pr(\omega_{(i)} < \gamma_0 - \epsilon)\right]^n
\end{aligned} \tag{A.11}$$

due to Assumption 3.2. The right-hand side of the last expression converges to unity as $n \to \infty$ for any $\epsilon > 0$, and hence $\widehat{\gamma}_0$ is weakly consistent.

For an arbitrarily small $\epsilon > 0$, rewrite (A.11) as

$$\Pr(\gamma_0 - \widehat{\gamma}_0 \leq \epsilon) = 1 - \left[1 - \Pr\left(\boldsymbol{x}_i \leq \boldsymbol{x}_0(\gamma_0-\epsilon)^{-1} \text{ and } \boldsymbol{y}_i \geq \boldsymbol{y}_0(\gamma_0-\epsilon)\right)\right]^n. \tag{A.12}$$

Let $\boldsymbol{x}_0(\gamma_0 - \epsilon)^{-1} = \boldsymbol{x}_0^{\partial}(1-\xi)^{-1}$ and $\boldsymbol{y}_0(\gamma_0 - \epsilon) = \boldsymbol{y}_0^{\partial}(1-\xi)$; clearly, $\xi > 0$ for $\epsilon > 0$. The probability expression on the right-hand side of (A.12) is related to the set $\mathcal{H}_0(\xi)$ given in Definition A.1.

Let $\rho_{\mathcal{H},0}(\xi) = \Pr\left((\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathcal{H}_0(\xi)\right)$. Similar to the reasoning in Park et al. (2000), this probability contains all information about the distribution of $\widehat{\gamma}_0$; moreover, $\rho_{\mathcal{H},0}(\xi)$ depends on the density f and the curvature of g. The distance $\gamma_0 - \widehat{\gamma}_0$ goes to 0 as $n \to \infty$; hence, only the properties of f and g in a neighborhood of $(\boldsymbol{x}_0^{\partial}, \boldsymbol{y}_0^{\partial})$ are needed to derive the asymptotic distribution of $\Pr(\gamma_0 - \widehat{\gamma}_0 \leq \epsilon)$.

Analogous to the reasoning in Park et al. (2000), consider an arbitrarily small ξ. Locally, g is linear and f is constant. Hence, $\mathcal{H}_0(\xi)$ may be approximated by a $(p+q)$-dimensional simplex where the angles depend on the partial derivatives of g at $(\boldsymbol{x}_0^{\partial}, \boldsymbol{y}_0^{\partial})$, and $\rho_{\mathcal{H},0}(\xi)$ is approximately proportional to ξ^{p+q}.

To obtain a sequence ξ in terms of the sample size n, define $\xi = n^{-1/(p+q)}\epsilon$ for any $\epsilon > 0$. Since $\lim_{n\to\infty}\left(1 - \frac{\rho}{n}\right)^n = e^{-\rho}$, the theorem follows from (A.11) and Lemma A.1. □

Proof of Theorem 4.2. Let $\epsilon > 0$. Then

$$\begin{aligned}\Pr\left(|\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) - \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})| > \epsilon\right) = \; & \Pr\left(\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) > \gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \epsilon\right) \\ & + \Pr\left(\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) < \gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \epsilon\right).\end{aligned}$$

Note that for $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) \; (\gtreqless) \; \gamma$, the fact that $H(\boldsymbol{x},\boldsymbol{y})$ is monotone nondecreasing with $\boldsymbol{x}$ and monotone nonincreasing with $\boldsymbol{y}$ implies $\widehat{H}(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y}) \; (\gtreqless) \; 1-\alpha$. Let

$$\widetilde{V}_i = I\left(\boldsymbol{x}_i \leq (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \epsilon)^{-1}\boldsymbol{x}, \boldsymbol{y}_i \geq (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \epsilon)\,\boldsymbol{y}\right)$$

and $V_i = \widetilde{V}_i - (1-\alpha)$. Then

$$\begin{aligned}
&\Pr\left[\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) > \gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \epsilon\right] \\
&\qquad \le \Pr\left[\widehat{H}\left((\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})+\epsilon)^{-1}\boldsymbol{x}, (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})+\epsilon)\boldsymbol{y}\right) > 1-\alpha\right] \\
&\qquad = \Pr\left[n^{-1}\sum_{i=1}^{n}\widetilde{V}_i > 1-\alpha\right] \\
&\qquad = \Pr\left[\sum_{i=1}^{n} V_i > 0\right] \\
&\qquad = \Pr\left[\sum_{i=1}^{n} V_i - \sum_{i=1}^{n} E(V_i) > n\omega_1\right],
\end{aligned}$$

where $\omega_1 = -E(V_1)$; the last step holds due to independent sampling. By definition, $\Pr(\alpha - 1 \le V_i \le 1) = 1 \; \forall \; i = 1, \ldots, n$. Using Hoeffding's (1963, Theorem #2) inequality gives

$$\Pr\left[\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) > \gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \epsilon\right] \le e^{-2n\omega_1^2/(2-\alpha)^2}. \tag{A.13}$$

Similarly, let

$$\widetilde{W}_i = I\left(\boldsymbol{x}_i \le (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) - \epsilon)^{-1}\boldsymbol{x}, \boldsymbol{y}_i \ge (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) - \epsilon)\boldsymbol{y}\right)$$

and $W_i = \widetilde{W}_i - (1-\alpha)$. Then

$$\begin{aligned}
&\Pr\left[\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) < \gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \epsilon\right] \\
&\qquad \le \Pr\left[\widehat{H}\left((\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})+\epsilon)^{-1}\boldsymbol{x}, (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})+\epsilon)\boldsymbol{y}\right) > 1-\alpha\right] \\
&\qquad = \Pr\left[n^{-1}\sum_{i=1}^{n}\widetilde{W}_i \le 1-\alpha\right] \\
&\qquad = \Pr\left[\sum_{i=1}^{n} W_i \le 0\right] \\
&\qquad = \Pr\left[\sum_{i=1}^{n} W_i - \sum_{i=1}^{n} E(W_i) \le n\omega_2\right],
\end{aligned}$$

where $\omega_2 = -E(W_1)$. Again using Hoeffding's inequality,

$$\Pr\left[\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) < \gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) - \epsilon\right] \le e^{-2n\omega_2^2/(2-\alpha)^2}. \tag{A.14}$$

The result in (4.14) is obtained by setting $\omega = \min\{\omega_1, \omega_2\}$.

To obtain the result in (4.15), note

$$\begin{aligned}\omega_1 &= -E(V_1)\\ &= -E\left[I\left(\boldsymbol{x}_1 \le (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})+\epsilon)^{-1}\boldsymbol{x}, \boldsymbol{y}_1 \ge (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})+\epsilon)\,\boldsymbol{y}\right) - (1-\alpha)\right]\\ &= (1-\alpha) - H\left((\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})+\epsilon)^{-1}\boldsymbol{x}, (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})+\epsilon)\,\boldsymbol{y}\right),\end{aligned}$$

and

$$\begin{aligned}\omega_2 &= -E(W_1)\\ &= -E\left[I\left(\boldsymbol{x}_1 \le (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})-\epsilon)^{-1}\boldsymbol{x}, \boldsymbol{y}_1 \ge (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})-\epsilon)\,\boldsymbol{y}\right) - (1-\alpha)\right]\\ &= (1-\alpha) - H\left((\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})-\epsilon)^{-1}\boldsymbol{x}, (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})-\epsilon)\,\boldsymbol{y}\right).\end{aligned}$$

Also note that

$$-H\left((\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})+\epsilon)^{-1}\boldsymbol{x}, (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})+\epsilon)\,\boldsymbol{y}\right) \le -H\left((\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})-\epsilon)^{-1}\boldsymbol{x}, (\gamma_\alpha(\boldsymbol{x},\boldsymbol{y})-\epsilon)\,\boldsymbol{y}\right)$$

for any $\epsilon > 0$. Hence $\omega_1 \ge \omega_2$ and $\omega = \min\{\omega_1, \omega_2\} = \omega_2$. □

Proof of Theorem 4.3. Define random variables

$$V_n = \sqrt{n}\left(\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) - \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})\right) \tag{A.15}$$

and

$$W_i = (1-\alpha) - I\left(\boldsymbol{x}_i \le \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})^{-1}\boldsymbol{x}, \boldsymbol{y}_i \ge \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})\boldsymbol{y}\right). \tag{A.16}$$

Clearly, $E(W_i) = (1-\alpha) - H\left(\boldsymbol{x}_i \le \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})^{-1}\boldsymbol{x}, \boldsymbol{y}_i \ge \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})\boldsymbol{y}\right) = (1-\alpha) - (1-\alpha) = 0$ and

$$\begin{aligned}\mathrm{VAR}(W_i) &= E(W_i^2)\\ &= E\Big[(1-\alpha)^2 + I\left(\boldsymbol{x}_i \le \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})^{-1}\boldsymbol{x}, \boldsymbol{y}_i \ge \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})\boldsymbol{y}\right)^2\\ &\qquad -2(1-\alpha)I\left(\boldsymbol{x}_i \le \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})^{-1}\boldsymbol{x}, \boldsymbol{y}_i \ge \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})\boldsymbol{y}\right)\Big]\\ &= (1-\alpha)^2 + (1-\alpha) - 2(1-\alpha)^2\\ &= \alpha(1-\alpha).\end{aligned}$$

Define

$$W_n = \frac{1}{\sqrt{n}G'(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}))}\sum_{i=1}^{n} W_i. \tag{A.17}$$

Then

$$\begin{aligned}\text{VAR}(W_n) &= \frac{1}{nG'(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}))^2} \times n\text{VAR}(W_i) \\ &= \frac{\alpha(1-\alpha)}{G'(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}))^2}.\end{aligned}$$

By the Lindeberg-Levy central limit theorem, $W_n \xrightarrow{d} N(0, \sigma_\alpha^2(\boldsymbol{x},\boldsymbol{y}))$ where

$$\sigma_\alpha^2(\boldsymbol{x},\boldsymbol{y}) = \alpha(1-\alpha)G'(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}))^{-2}.$$

Now define $R_n = V_n - W_n$. It suffices to show that $R_n \xrightarrow{p} 0$. By definition, $\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) \geq \gamma$ iff $\widehat{H}(\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y}) > (1-\alpha)$ for $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{P}$. Then for any real t,

$$\begin{aligned}V_n \geq t &\Leftrightarrow \sqrt{n}\left(\widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) - \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})\right) \geq t \\ &\Leftrightarrow \widehat{\gamma}_{\alpha,n}(\boldsymbol{x},\boldsymbol{y}) \geq -\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}} \\ &\Leftrightarrow \widehat{H}\left(\left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right)^{-1}\boldsymbol{x}, \left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right)\boldsymbol{y}\right) > 1-\alpha \\ &\Leftrightarrow Z_{t,n} > T_n \qquad \text{(A.18)}\end{aligned}$$

where

$$\begin{aligned}T_{t,n} &= \frac{\sqrt{n}}{G'(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}))}\left[G\left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right) - (1-\alpha)\right], \\ Z_{t,n} &= \frac{\sqrt{n}}{G'(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}))}\left[G\left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right)\right. \\ &\qquad \left. -\widehat{H}\left(\left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right)^{-1}\boldsymbol{x}, \left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right)\boldsymbol{y}\right)\right].\end{aligned}$$

Recall that $G'(\gamma) < 0$; using the definition of the first derivative of a function, it is clear that

$$G\left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right) - (1-\alpha) = \frac{t}{\sqrt{n}}G'\left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right) + \frac{t}{\sqrt{n}}o(1). \qquad \text{(A.19)}$$

Then by (A.19),

$$T_{t,n} \xrightarrow{p} t \text{ as } n \to \infty. \qquad \text{(A.20)}$$

Turning to $Z_{t,n}$, it is clear that $E(Z_{t,n}) = 0$ and hence $E(Z_{t,n} - W_n) = 0$ since $E(W_n) = 0$. Therefore $\text{VAR}((Z_{t,n} - W_n) = E\left[(Z_{t,n} - W_n)^2\right]$. Note that

$$\begin{aligned}
Z_{t,n} &= \frac{\sqrt{n}}{G'(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}))}\left[G\left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right) - \right. \\
&\quad \left. \frac{1}{n}\sum_{i=1}^{n} I\left(\boldsymbol{x}_i \leq \left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right)^{-1}\boldsymbol{x}, \boldsymbol{y}_i \geq \left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right)\boldsymbol{y}\right)\right] \\
&= \frac{\sqrt{n}}{G'(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}))}\frac{1}{n}\sum_{i=1}^{n}\left[G\left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right)\right. \\
&\quad \left. I\left(\boldsymbol{x}_i \leq \left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right)^{-1}\boldsymbol{x}, \boldsymbol{y}_i \geq \left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right)\boldsymbol{y}\right)\right].
\end{aligned} \tag{A.21}$$

In addition, from the definition of W_n in (A.17) and the definition of $\widehat{H}(\cdot,\cdot)$,

$$W_n = \frac{\sqrt{n}}{G'(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}))}\frac{1}{n}\sum_{i=1}^{n}\left[(1-\alpha) - I\left(\boldsymbol{x}_i \leq \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})^{-1}\boldsymbol{x}, \boldsymbol{y}_i \geq \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})\boldsymbol{y}\right)\right]. \tag{A.22}$$

Then

$$\text{VAR}(Z_{t,n} - W_n) = \frac{n}{G'(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}))^2} \times \text{VAR}\left(\frac{1}{n}\sum_{i=1}^{n} A_i\right)$$

where

$$\begin{aligned}
A_i = G&\left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right) \\
&- I\left(\boldsymbol{x}_i \leq \left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right)^{-1}\boldsymbol{x}, \boldsymbol{y}_i \geq \left(\gamma_\alpha(\boldsymbol{x},\boldsymbol{y}) + \frac{t}{\sqrt{n}}\right)\boldsymbol{y}\right) \\
&- (1-\alpha) + I\left(\boldsymbol{x}_i \leq \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})^{-1}\boldsymbol{x}, \boldsymbol{y}_i \geq \gamma_\alpha(\boldsymbol{x},\boldsymbol{y})\boldsymbol{y}\right).
\end{aligned}$$

Clearly, $E(A_i) = 0$; hence $\text{VAR}(A_i) = E\left(A_i^2\right)$. Denote $\gamma_\alpha(\boldsymbol{x}, \boldsymbol{y})$ by γ_α. Then

$$\begin{aligned}
A_i^2 = G&\left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right)^2 \\
&-2G\left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right) I\left(\boldsymbol{x}_i \leq \left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right)^{-1} \boldsymbol{x}, \boldsymbol{y}_i \geq \left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right) \boldsymbol{y}\right) \\
&-2(1-\alpha)G\left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right) \\
&+2G\left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right) I\left(\boldsymbol{x}_i \leq \gamma_\alpha^{-1}\boldsymbol{x}, \boldsymbol{y}_i \geq \gamma_\alpha \boldsymbol{y}\right) \\
&+I\left(\boldsymbol{x}_i \leq \left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right)^{-1} \boldsymbol{x}, \boldsymbol{y}_i \geq \left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right) \boldsymbol{y}\right) \\
&-2(1-\alpha)G\left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right) \\
&+2(1-\alpha)I\left(\boldsymbol{x}_i \leq \left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right)^{-1} \boldsymbol{x}, \boldsymbol{y}_i \geq \left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right) \boldsymbol{y}\right) \\
&-2(1-\alpha)G\left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right) \\
&-2(1-\alpha)I\left(\boldsymbol{x}_i \leq \left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right)^{-1} \boldsymbol{x}, \boldsymbol{y}_i \geq \left(\gamma_\alpha + \frac{t}{\sqrt{n}}\right) \boldsymbol{y}\right) \\
&-2I\left(\boldsymbol{x}_i \leq (\gamma_\alpha + \zeta)^{-1} \boldsymbol{x}, \boldsymbol{y}_i \geq (\gamma_\alpha + \zeta)\, \boldsymbol{y}\right) \\
&+(1-\alpha)^2 \\
&-2(1-\alpha)I\left(\boldsymbol{x}_i \leq \gamma_\alpha^{-1}\boldsymbol{x}, \boldsymbol{y}_i \geq \gamma_\alpha \boldsymbol{y}\right) \\
&+I\left(\boldsymbol{x}_i \leq \gamma_\alpha^{-1}\boldsymbol{x}, \boldsymbol{y}_i \geq \gamma_\alpha \boldsymbol{y}\right)
\end{aligned}$$

where $\zeta = \max\left(0, \frac{t}{\sqrt{n}}\right)$.

Taking expectations,

$$\begin{aligned}
E\left(A_i^2\right) = G&\left(\gamma_\alpha(\boldsymbol{x}, \boldsymbol{y}) + \frac{t}{\sqrt{n}}\right)^2 \\
&+\left[1 + 2(1-\alpha)\right] G\left(\gamma_\alpha(\boldsymbol{x}, \boldsymbol{y}) + \frac{t}{\sqrt{n}}\right) \\
&-2G\left(\gamma_\alpha(\boldsymbol{x}, \boldsymbol{y}) + \zeta\right) \\
&-(1-\alpha)^2 + (1-\alpha).
\end{aligned}$$

Since $\lim_{n\to\infty} E\left(A_i^2\right) = 0$, $E\left(Z_{t,n} - W_n\right)^2 \to 0$ as $n \to \infty$ and hence

$$(Z_{t,n} - W_n) \xrightarrow{p} 0 \text{ as } n \to \infty. \tag{A.23}$$

Ghosh (1971, lemma 1) gives two conditions that are sufficient for $R_n \xrightarrow{p} 0$:

(i) $\forall\, \delta > 0,\ \exists$ a λ such that $\Pr(|W_n| > \lambda) < \delta$; and

(ii) $\forall\, k$ and $\forall\, \epsilon > 0$,

$$\lim_{n\to\infty} \Pr\left(V_n \geq k + \epsilon,\ W_n \leq k\right) = 0, \tag{A.24}$$

$$\lim_{n\to\infty} \Pr\left(V_n < k,\ W_n \geq k + \epsilon\right) = 0. \tag{A.25}$$

Condition (i) follows trivially from the Bienaymé-Chebyshev inequality (*e.g.,* see Stuart and Ord, 1994, p. 113) whenever λ is chosen to satisfy $\lambda^2 > \sigma_\alpha^2(\boldsymbol{x}, \boldsymbol{y})\delta^{-1}$. In addition, for any k and any $\epsilon > 0$, set $t = k + \epsilon$; then by (A.18),

$$\begin{aligned}
\Pr\left(V_n \geq k + \epsilon,\ W_n \leq k\right) &= \Pr\left(Z_{t,n} > T_{t,n},\ W_n \leq t - \epsilon\right) \\
&= \Pr\left(\left|(Z_{t,n} - W_n) - (T_{t,n} - t)\right| \geq \epsilon\right).
\end{aligned}$$

Then (A.24) follows from (A.20) and (A.23).

Setting $t = k$ and applying similar reasoning leads to the result in (A.25). Hence $R_n \xrightarrow{p} 0$ by lemma 2 of Ghosh (1971). Therefore V_n has the same limiting distribution as W_n. □

Proof of Lemma 4.1. From (3.13),

$$\begin{aligned}
\widehat{\gamma}_n(\boldsymbol{x}, \boldsymbol{y}) - \widehat{\gamma}_{\alpha(n),n}(\boldsymbol{x}, \boldsymbol{y}) &= (\omega_{(n)} - \omega_{(\alpha(n)n)})I(\alpha(n)n \in \mathbb{N}) \\
&\quad + (\omega_{(n)} - \omega_{([\alpha(n)n]+1)})I(\alpha(n)n \notin \mathbb{N}).
\end{aligned}$$

Let $C_{\boldsymbol{x},\boldsymbol{y},k}(n) = \frac{\omega_{(n)} - \omega_{(k)}}{1 - n^{-1}k}\ \forall\, k \in \{1,\ \ldots,\ n-1\}$ and $C_{\boldsymbol{x},\boldsymbol{y}}(n) = \max_{k\in\{1,\ \ldots,\ n-1\}} C_{\boldsymbol{x},\boldsymbol{y},k}(n)$. Then

$$\begin{aligned}
&(\omega_{(n)} - \omega_{(\alpha(n)n)})I(\alpha(n)n \in \mathbb{N}) \\
&\quad \leq C_{\boldsymbol{x},\boldsymbol{y}}(n)(1 - \alpha(n))I(\alpha(n)n \in \mathbb{N}), (\omega_{(n)} - \omega_{([\alpha(n)n]+1)})I(\alpha(n)n \notin \mathbb{N}) \\
&\quad \leq C_{\boldsymbol{x},\boldsymbol{y}}(n)(1 - \alpha(n))I(\alpha(n)n \notin \mathbb{N}).
\end{aligned}$$

Hence $n^{1/(p+q)}(\widehat{\gamma}_n(\boldsymbol{x}, \boldsymbol{y}) - \widehat{\gamma}_{\alpha(n),n}(\boldsymbol{x}, \boldsymbol{y})) \leq n^{1/(p+q)} C_{bx,\boldsymbol{y}}(n)(1 - \alpha(n))$. Since by Assumptions 3.2 and 3.3 the support of $(\boldsymbol{x}, \boldsymbol{y})$ is bounded, there exists a constant $M_{xy} > 0$ such that

$\omega_i \leq M_{xy}$ almost surely, for any $i = 1, \ldots, n$. Therefore $\omega_{(n)} - \omega_{(k)} \leq \omega_{n)} \leq M_{xy}$ almost surely, for any $k = 1, \ldots, n-1$. Since $\frac{1}{1-n^{-1}k} \leq n$ (provided $n \geq 2$), $C_{\boldsymbol{x},\boldsymbol{y}}(n) \leq M_{xy}n$ almost surely. Hence

$$\begin{aligned} n^{1/(p+q)}(\widehat{\gamma}_n(\boldsymbol{x},\boldsymbol{y}) - \widehat{\gamma}_{\alpha(n),n}(\boldsymbol{x},\boldsymbol{y})) &\leq n^{1/(p+q)} M_{xy} n(1-\alpha(n)) \\ &= M_{xy} n^{(p+q+1)/(p+q)}(1-\alpha(n)) \end{aligned}$$

almost surely. The lemma follows after applying the strong law of large numbers. □

References

Aigner, D., Lovell, C. A. K., and Schmidt, P. (1977), "Formulation and estimation of stochastic frontier production function models," *Journal of Econometrics,* 6, 21–37.

Andersen, P., and Petersen, N. C. (1993), "A procedure for ranking efficient units in data envelopment analysis," *Management Science,* 39, 1261–1264.

Aragon, Y., Daouia, A., and Thomas-Agnan, C. (2005), "Nonparametric frontier estimation: A conditional quantile-based approach," *Econometric Theory,* 21, 358–389.

Banks, J., Blundell, R., and Lewbel, A. (1997), "Quadratic engel curves and consumer demand," *Review of Economics and Statistics,* 79, 527–539.

Barnett, W. A., Geweke, J., and Wolfe, M. (1991), "Seminonparametric Bayesian estimation of the asymptotically ideal production model," *Journal of Econometrics,* 49, 5–50.

Bauer, P. W., and Ferrier, G. D. (1996), "Scale economies, cost efficiencies, and technological change in federal reserve payments processing," *Journal of Money, Credit, and Banking,* 28, 1004–1039.

Bauer, P. W., and Hancock, D. (1993), "The efficiency of the federal reserve in providing check processing services," *Journal of Banking and Finance,* 17, 287–311.

Berger, A. N., and Humphrey, D. B. (1997), "Efficiency of financial institutions: International survey and directions for future research," *European Journal of Operational Research,* 98, 175–212.

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

Chalfant, J. A., and Gallant, A. R. (1985), "Estimating substitution elasticities with the Fourier cost function," *Journal of Econometrics,* 28, 205–222.

Cooper, R. J., and McLaren, K. R. (1996), "A system of demand equations satisfying effectively global regularity conditions," *Review of Economics and Statistics,* 78, 359–364.

Daouia, A. (2003), "Nonparametric analysis of frontier production functions and efficiency measurement using nonstandard conditional quantiles," PhD dissertation, Groupe de Recherche en Economie Mathématique et Quantititative, Université des Sciences Sociales, Toulouse I, et Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse III.

Daouia, A., and Simar, L. (2007), "Nonparametric efficiency analysis: A multivariate conditional quantile approach," *Journal of Econometrics,* 140, 375–400.

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

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

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

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

Gallant, A. R. (1981), "On the bias in flexible functional forms and an essentially unbiased form," *Journal of Econometrics,* 15, 211–245.

—— (1982), "Unbiased determination of production technologies," *Journal of Econometrics,* 20, 285–323.

Gerdes, G. R., and Walton, J. K. (2002), "The use of checks and other noncash payment instruments in the united states," *Federal Reserve Bulletin,* 88, 3600–3674.

Ghosh, J. K. (1971), "A new proof of the Bahadur representation of quantiles and an application," *The Annals of Mathematical Sciences,* 42, 1957–1961.

Gilbert, A., and Wilson, P. W. (1998), "Effects of deregulation on the productivity of korean banks," *Journal of Economics and Business,* 50, 133–155.

Gilbert, A., Wheelock, D. C., and Wilson, P. W. (2004), "New evidence on the fed's productivity in providing payments services," *Journal of Banking and Finance,* 28, 2175–2190.

Guilkey, D. K., Lovell, C. A. K., and Sickles, R. C. (1983), "A comparison of the performance of three flexible functional forms," *International Economic Review,* 24, 591–616.

Hoeffding, W. (1963), "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association,* 58, 13–30.

Hsu, P. L., and Robbins, H. (1947), "Complete convergence and the law of large numbers," *Proceedings of the National Academy of Sciences of the United States of America,* 33, 25–31.

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

Kneip, A., Simar, L., and Wilson, P. W. (2007), "Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models," *Econometric Theory.* Forthcoming.

Korostelev, A., Simar, L., and Tsybakov, A. B. (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.

Kuntz, L., and Scholtes, S. (2000), "Measuring the robustness of empirical efficiency valuations," *Management Science,* 46, 807–823.

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

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

Porembski, M., Breitenstein, K., and Alpar, P. (2005), "Visualizing efficiency and reference relations in data envelopment analysis with an application to the branches of a german bank," *Journal of Productivity Analysis,* 23(2), 203–221.

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

Simar, L. (2003), "Detecting outliers in frontier models: A simple approach," *Journal of Productivity Analysis,* 20, 391–424.

Simar, L., and Wilson, P. W. (1998), "Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models," *Management Science,* 44, 49–61.

—— (2000a), "A general methodology for bootstrapping in non-parametric frontier models," *Journal of Applied Statistics,* 27, 779–802.

—— (2000b), "Statistical inference in nonparametric frontier models: The state of the art," *Journal of Productivity Analysis,* 13, 49–78.

Stuart, A., and Ord, K. (1994), *Kendall's Advanced Theory of Statistics,* sixth ed., vol. 1, London: Edward Arnold.

Wheelock, D. C., and Wilson, P. W. (1995), "Explaining bank failures: Deposit insurance, regulation, and efficiency," *Review of Economics and Statistics,* 77, 689–700.

—— (2000), "Why do banks disappear? the determinants of US bank failures and acquisitions," *Review of Economics and Statistics,* 82, 127–138.

—— (2001), "New evidence on returns to scale and product mix among US commercial banks," *Journal of Monetary Economics,* 47, 115–132.

—— (2003), "Robust nonparametric estimation of efficiency and technical change in U. S. commercial banking," Unpublished working paper, The John E. Walker Department of Economics, 222 Sirrine Hall, Clemson University, Clemson, South Carolina 29634, USA.

—— (2004), "Trends in the efficiency of federal reserve check processing operations," *Federal Reserve Bank of St. Louis Review,* 86, 7–19.

Wilson, P. W. (1993), "Detecting outliers in deterministic nonparametric frontier models with multiple outputs," *Journal of Business and Economic Statistics,* 11, 319–323.

—— (1995), "Detecting influential observations in data envelopment analysis," *Journal of Productivity Analysis,* 6, 27–45.

—— (2007), "FEAR: A software package for frontier efficiency analysis with R," *Socio-Economic Planning Sciences.* Forthcoming.

Wilson, P. W., and Carey, K. (2004), "Nonparametric analysis of returns to scale and product mix among US hospitals," *Journal of Applied Econometrics,* 19, 505–524.

Table 1: Definitions and Measurement of Inputs

1. *Personnel*—number of employee work hours.

2. *Materials, Software, Equipment and Support*—expenditures are deflated by the following price measures:

   - *Materials:* Consumer Price Index (CPI) (n.s.a, 2000=100)
   - *Software:* private nonresidential fixed investment deflator for software (n.s.a., 2000=100)
   - *Equipment:* for 1979–1989, PPI for check-handling machines; for 1990–2003, PPI for the net output of select industries-office machines, n.e.c. (n.s.a., 2000=100)
   - *Other Support:* CPI (n.s.a., 2000=100)

3. *Transit*—expenditures for shipping, travel, communications, and data communications support deflated by the following price measures:

   - *Shipping and Travel:* private nonresidential fixed investment deflator for aircraft (n.s.a., 2000=100).
   - *Communications and Communications Support:* private nonresidential fixed investment deflator for communications equipment (n.s.a., 2000=100).

4. *Facilities*—expenditures on facilities support deflated by the following price index: "Historical Cost Index" from *Means Square Foot Costs Data 2000* (R.S. Means Company: Kingston, MA), pp. 436-442. Data are January values.

Sources: Federal Reserve Planning and Control System documents unless otherwise noted. Additional details are available from the authors.

Table 2: Summary Statistics for Inputs and Outputs

	1980:3	1992:3	2003:3
Inputs:			
labor	43438.5	33525.3	25994.5
	33503.7	22121.1	15698.1
materials	2007.9	1694.9	2373.5
	1321.8	1023.9	1252.2
facilities	644.9	714.6	919.6
	585.1	475.7	572.8
transit	1906.3	1281.4	492.3
	1287.9	701.7	301.8
Outputs:			
checks	84325.1	84896.5	97909.7
	52406.7	46057.8	49676.4
endpts	399.8	430.1	307.8
	296.3	303.6	164.9

NOTE: Each entry gives sample mean (top) and sample standard deviation (bottom).

Table 3: Hyperbolic Quantile, DEA, and FDH Efficiency Estimates, 2003:3, $\alpha = 0.9$

		95% CI		input		(inverse) output	
site	$\widehat{\gamma}_\alpha$	$\widehat{\gamma}_{\text{lo}}$	$\widehat{\gamma}_{\text{hi}}$	DEA	FDH	DEA	FDH
1	0.3427	0.3045	0.4340	1.0000	1.0000	1.0000	1.0000
2	0.4122	0.3773	0.5638	1.0000	1.0000	1.0000	1.0000
3	0.4886	0.4493	0.6369	1.0000	1.0000	1.0000	1.0000
4	0.5190	0.4809	0.6920	1.0000	1.0000	1.0000	1.0000
5	0.5395	0.5043	0.6847	1.0000	1.0000	1.0000	1.0000
6	0.5411	0.5044	0.6736	1.0000	1.0000	1.0000	1.0000
7	0.5514	0.5131	0.7146	1.0000	1.0000	1.0000	1.0000
8	0.5675	0.5287	0.7385	1.0000	1.0000	1.0000	1.0000
9	0.5733	0.5035	0.7315	1.0000	1.0000	1.0000	1.0000
10	0.5817	0.5291	0.7021	1.0000	1.0000	1.0000	1.0000
11	0.5908	0.5572	0.7251	1.0000	1.0000	1.0000	1.0000
12	0.5979	0.5691	0.7277	1.0000	1.0000	1.0000	1.0000
13	0.6396	0.5808	0.8226	1.0000	1.0000	1.0000	1.0000
14	0.6551	0.6019	0.7841	1.0000	1.0000	1.0000	1.0000
15	0.6675	0.5964	0.7506	1.0000	1.0000	1.0000	1.0000
16	0.6719	0.6253	0.8308	1.0000	1.0000	1.0000	1.0000
17	0.6845	0.6338	0.8499	1.0936	1.0000	1.1094	1.0000
18	0.6916	0.6297	0.8223	1.1498	1.0000	1.2191	1.0000
19	0.6922	0.6406	0.8506	1.0027	1.0000	1.0021	1.0000
20	0.7024	0.6416	0.8555	1.0000	1.0000	1.0000	1.0000
21	0.7206	0.6137	0.7685	1.0236	1.0000	1.0220	1.0000
22	0.7208	0.6486	0.8219	1.0000	1.0000	1.0000	1.0000
23	0.7271	0.6333	0.8303	1.0000	1.0000	1.0000	1.0000
24	0.7608	0.6864	0.8945	1.1173	1.0000	1.1038	1.0000
25	0.7867	0.7696	0.9923	1.0000	1.0000	1.0000	1.0000
26	0.7892	0.7397	0.9928	1.0000	1.0000	1.0000	1.0000
27	0.8000	0.7178	0.8910	1.1414	1.0000	1.1186	1.0000
28	0.8069	0.7527	1.0136	1.1943	1.0133	1.1488	1.0445
29	0.8101	0.6822	0.8416	1.0541	1.0000	1.0500	1.0000
30	0.8108	0.7256	0.9825	1.0164	1.0000	1.0148	1.0000
31	0.8164	0.7360	0.9373	1.0431	1.0000	1.0337	1.0000
32	0.8204	0.7283	0.9177	1.0061	1.0000	1.0057	1.0000
33	0.8228	0.6964	0.8981	1.0336	1.0000	1.0302	1.0000
34	0.8424	0.7745	0.9901	1.2330	1.0000	1.2137	1.0000
35	0.8477	0.7733	1.0261	1.2852	1.2520	1.2045	1.0212
36	0.8523	0.7652	0.9488	1.2074	1.0473	1.1783	1.0767
37	0.8844	0.7571	0.9336	1.3973	1.0000	1.2898	1.0000
38	0.8872	0.7182	0.8996	1.2957	1.0000	1.2726	1.0000
39	0.9163	0.7922	0.9870	1.4814	1.0000	1.3214	1.0000
40	0.9182	0.8345	1.0233	1.4881	1.0927	1.3659	1.0942
41	0.9226	0.8334	1.0309	1.3026	1.0173	1.2694	1.0753
42	0.9295	0.8437	1.0911	1.4955	1.1647	1.3822	1.1880
43	1.0253	0.9183	1.1456	1.8164	1.3685	1.5788	1.3514
44	1.0584	0.9034	1.1580	2.2161	1.2831	2.0276	1.5150

NOTE: Output-oriented DEA and FDH efficiency estimates are reported as reciprocals of Shephard output distance function estimators to facilitate comparisons; i.e., increasing values correspond to increasing inefficiency. The column labeled $\widehat{\gamma}_\alpha$ gives estimates of the hyperbolic α-quantile distance function defined in (3.4, while the columns labeled $\widehat{\gamma}_{\text{lo}}$ and $\widehat{\gamma}_{\text{hi}}$ give bounds of confidence intervals estimated using the bootstrap procedure described in footnote 17.

Table 4: Hyperbolic Quantile, Input- and (Inverse) Output-Conditional Quantile Efficiency Estimates, 2003:3, $\alpha = 0.9,\ 0.95$

	$\alpha = 0.9$			$\alpha = 0.95$		
site	$\widehat{\gamma}_\alpha$	$\widehat{\theta}_\alpha$	$1/\widehat{\lambda}_\alpha$	$\widehat{\gamma}_\alpha$	$\widehat{\theta}_\alpha$	$1/\widehat{\lambda}_\alpha$
1	0.3427	0.3359	1.0000	0.4409	0.3377	1.0000
2	0.4122	0.4039	1.0000	0.4464	1.0000	1.0000
3	0.4886	0.4336	1.0000	0.6575	0.6575	1.0000
4	0.5190	1.0000	0.5031	0.5469	1.0000	0.6109
5	0.5395	0.5395	1.0000	0.7606	0.8000	1.0000
6	0.5411	0.6283	1.0000	0.5575	1.0000	1.0000
7	0.5514	1.0000	0.4759	0.5733	1.0000	0.5733
8	0.5675	1.0000	0.5258	0.6117	1.0000	0.5675
9	0.5733	0.5733	1.0000	0.6894	0.6894	1.0000
10	0.5817	0.5063	1.0000	0.6021	0.6021	1.0000
11	0.5908	1.0000	1.0000	0.5950	1.0000	1.0000
12	0.5979	1.0000	1.0000	0.6849	1.0000	1.0000
13	0.6396	1.0000	0.6014	0.6744	1.0000	1.0000
14	0.6551	0.4625	1.0000	0.6851	1.0000	1.0000
15	0.6675	0.5756	1.0000	0.6871	1.0000	1.0000
16	0.6719	1.0000	0.6567	0.7494	1.0000	0.7494
17	0.6845	0.9068	1.0000	0.7243	1.0000	1.0000
18	0.6916	0.6916	1.0000	0.9070	0.9070	1.0000
19	0.6922	0.6922	1.0000	0.7590	1.0000	1.0000
20	0.7024	0.6916	1.0000	0.7986	0.7024	1.0000
21	0.7206	0.8357	1.0000	0.7672	1.0000	1.0000
22	0.7208	1.0000	0.6819	0.7971	1.0000	0.8301
23	0.7271	1.0000	0.6589	0.7388	1.0000	1.0000
24	0.7608	0.6666	1.0000	0.7976	0.7976	1.0000
25	0.7867	0.7547	0.8552	0.8187	0.7560	1.0000
26	0.7892	1.0000	0.7466	0.8471	1.0000	0.7892
27	0.8000	1.0000	0.7703	0.8344	1.0000	1.0000
28	0.8069	1.0133	0.8670	0.9358	1.0133	1.0000
29	0.8101	0.8683	1.0000	0.8619	1.0000	1.0000
30	0.8108	0.7736	1.0000	0.8737	0.7753	1.0000
31	0.8164	1.0000	0.8164	0.9249	1.0000	0.8287
32	0.8204	0.6560	1.0000	0.9099	0.7117	1.0000
33	0.8228	0.7868	1.0000	0.8271	0.8230	1.0000
34	0.8424	0.8424	0.8436	0.8436	1.0000	1.0000
35	0.8477	1.2520	0.9150	0.9822	1.2520	1.0000
36	0.8523	1.0473	0.8523	0.8823	1.0473	1.0000
37	0.8844	1.0000	0.9372	0.9372	1.0000	0.9523
38	0.8872	0.8666	1.0000	0.9085	0.8929	1.0000
39	0.9163	0.9443	0.9163	0.9443	0.9904	1.0000
40	0.9182	1.0012	1.0000	1.0000	1.0927	1.0345
41	0.9226	1.0000	0.7978	0.9886	1.0173	1.0000
42	0.9295	1.0000	0.9759	0.9759	1.1647	1.0000
43	1.0253	1.3088	1.0757	1.0757	1.3685	1.3095
44	1.0584	1.1063	1.4815	1.1188	1.2199	1.5150

NOTE: Hyperbolic α-quantile distance function estimates are shown in the columns labeled $\widehat{\gamma}_\alpha$ for each office. Offices are sorted by the value of these estimates for $\alpha = 0.9$ as in Table 3. Columns labeled $\widehat{\theta}_\alpha$ and $1/\widehat{\lambda}_\alpha$ give estimates for the conditional input- and output-α-quantile distance functions; reciprocals of the output-oriented estimates are given so that in all cases, increasing values correspond to increasing inefficiency.

Table 5: Weak Dominance Relationships, 2003:3, $\alpha = 0.9$

site	relevant peers					# times relevant
1	3	5	10	18		0
2	3	6	10	17		10
3	5	6	15	18		7
4	7	8	16	23		2
5	2	3	18	21		4
6	9	15	19	21		9
7	4	16	22	35		2
8	4	7	12	16		1
9	3	6	15	19		5
10	5	14	18	24		3
11	9	15	17	19		16
12	11	17	19	22		6
13	11	22	23	27		4
14	2	21	29	33		4
15	2	6	19	21		14
16	12	23	31	37		4
17	9	11	15	39		9
18	2	3	5	21		4
19	9	11	15	29		11
20	3	14	24	33		1
21	2	6	15	38		11
22	11	12	23	36		10
23	11	13	22	27		11
24	10	14	21	30		3
25	15	29	32	34		0
26	11	22	23	31		0
27	11	17	19	23		4
28	11	12	22	37		0
29	2	6	15	21		5
30	2	3	21	33		1
31	11	13	23	27		2
32	6	15	29	38		2
33	2	15	20	21		4
34	9	15	19	32		1
35	11	12	16	22		1
36	11	13	17	19	22	2
37	11	12	22	36		2
38	2	6	15	21		3
39	15	29	33	38		1
40	13	17	19	23		0
41	11	17	23	27		0
42	11	17	19	23		0
43	11	17	19	22	23	0
44	2	6	14	21	24	0

Figure 1: Conditional Input, Conditional Output, and Hyperbolic Quantile Frontiers

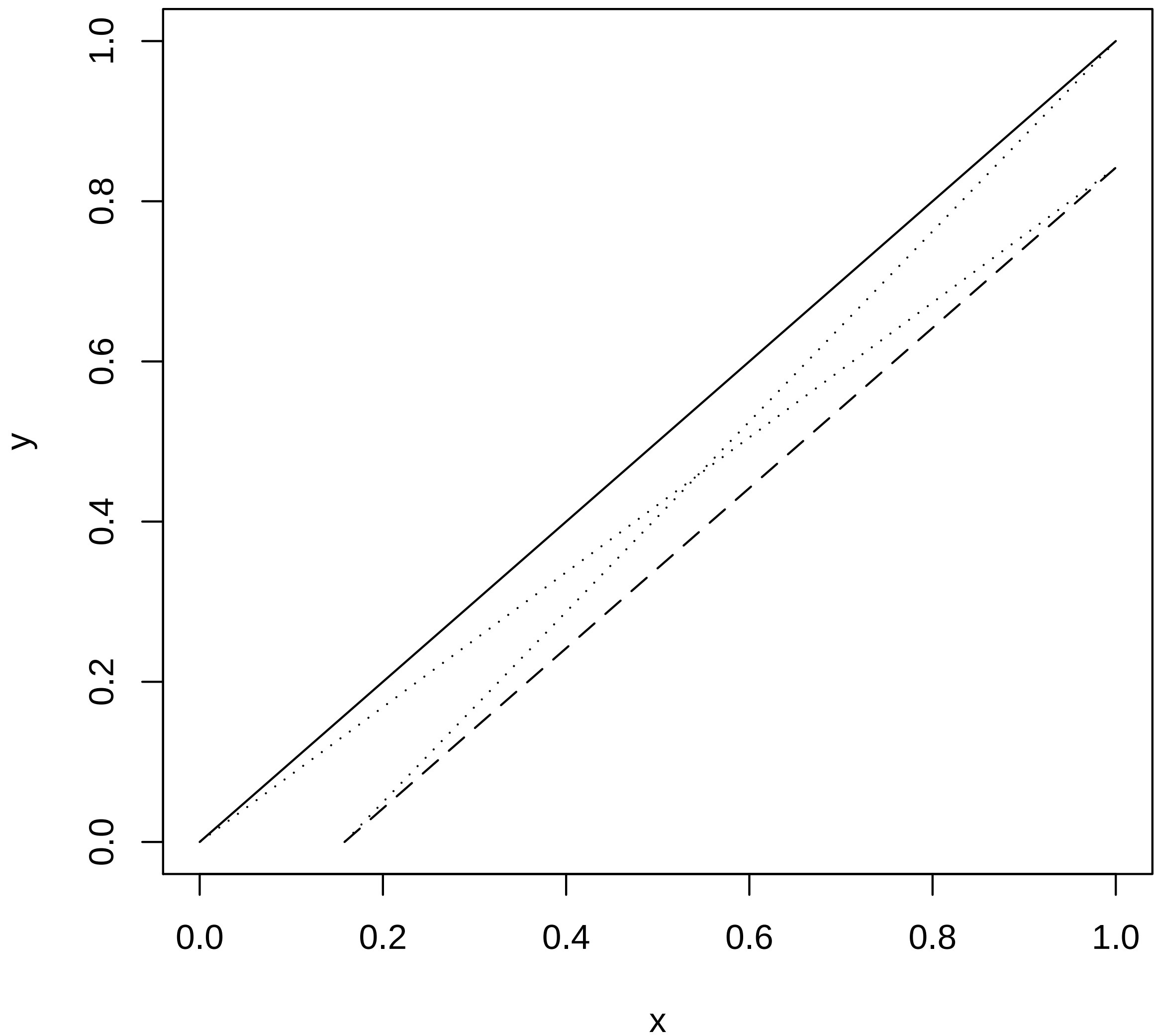

Figure 2: Example Illustrating DEA and Hyperbolic α-Quantile Estimators

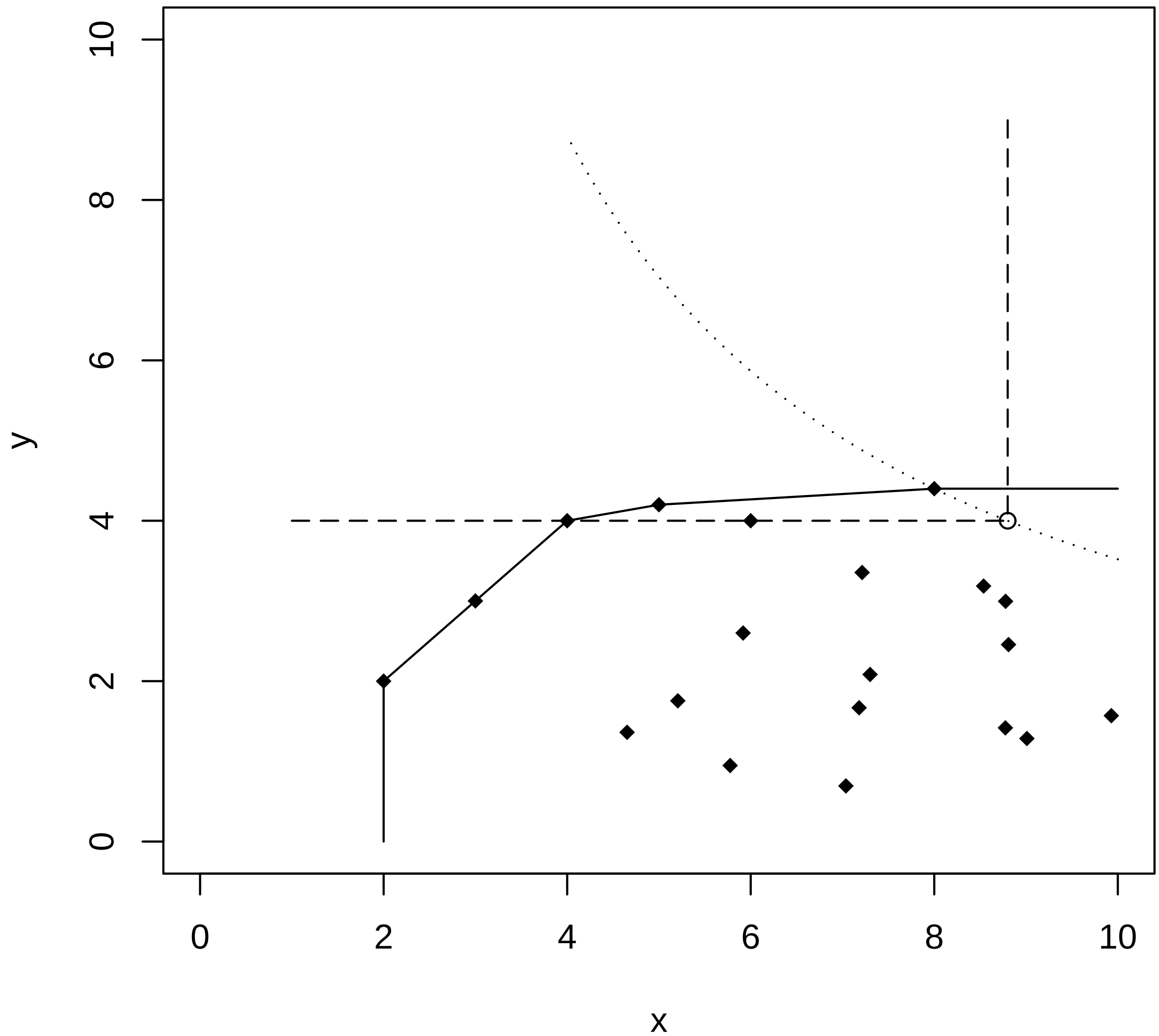

Figure 3: Hyperbolic Graph Efficiency Estimates, 2003:3

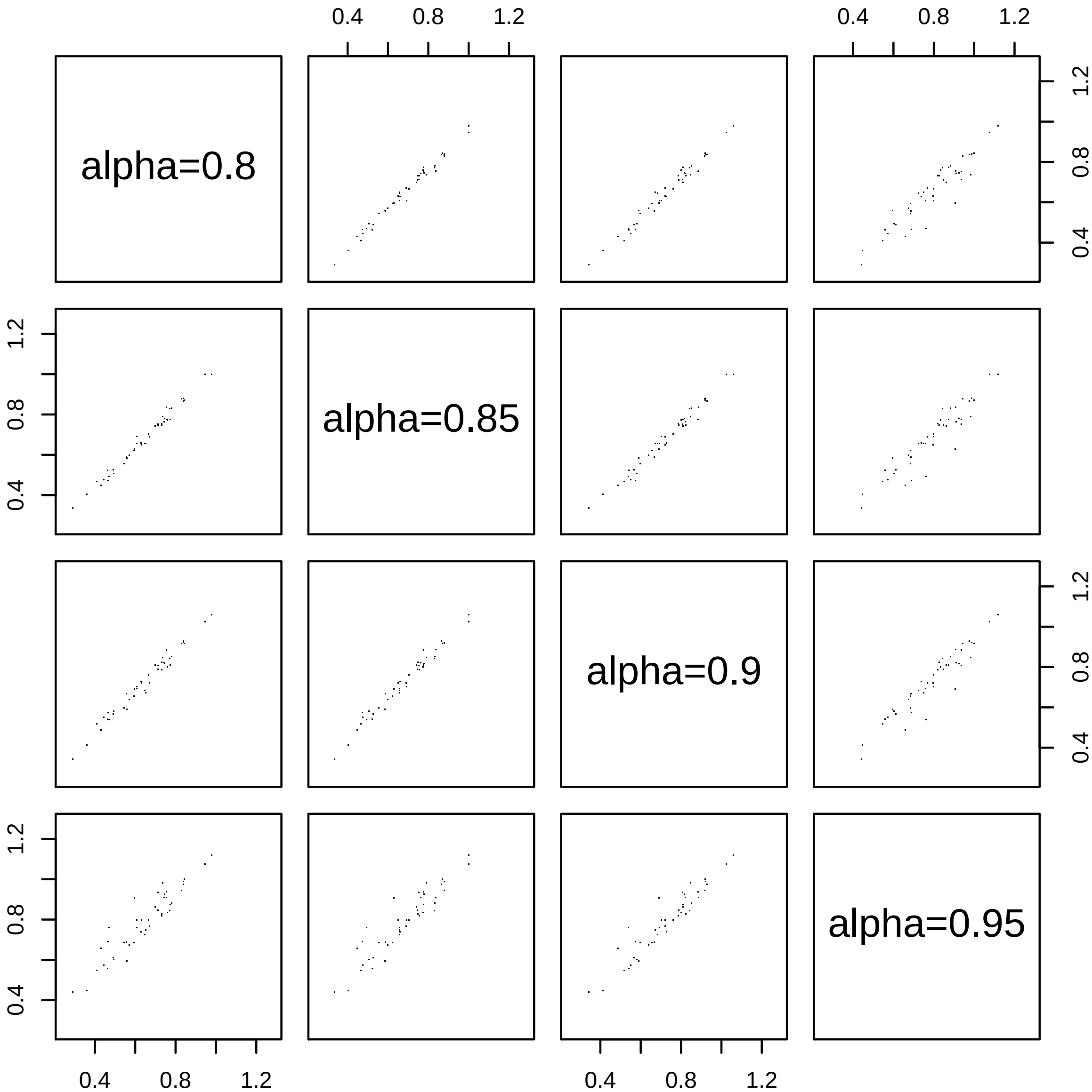

Figure 4: Mean Hyperbolic Graph Efficiency Estimates

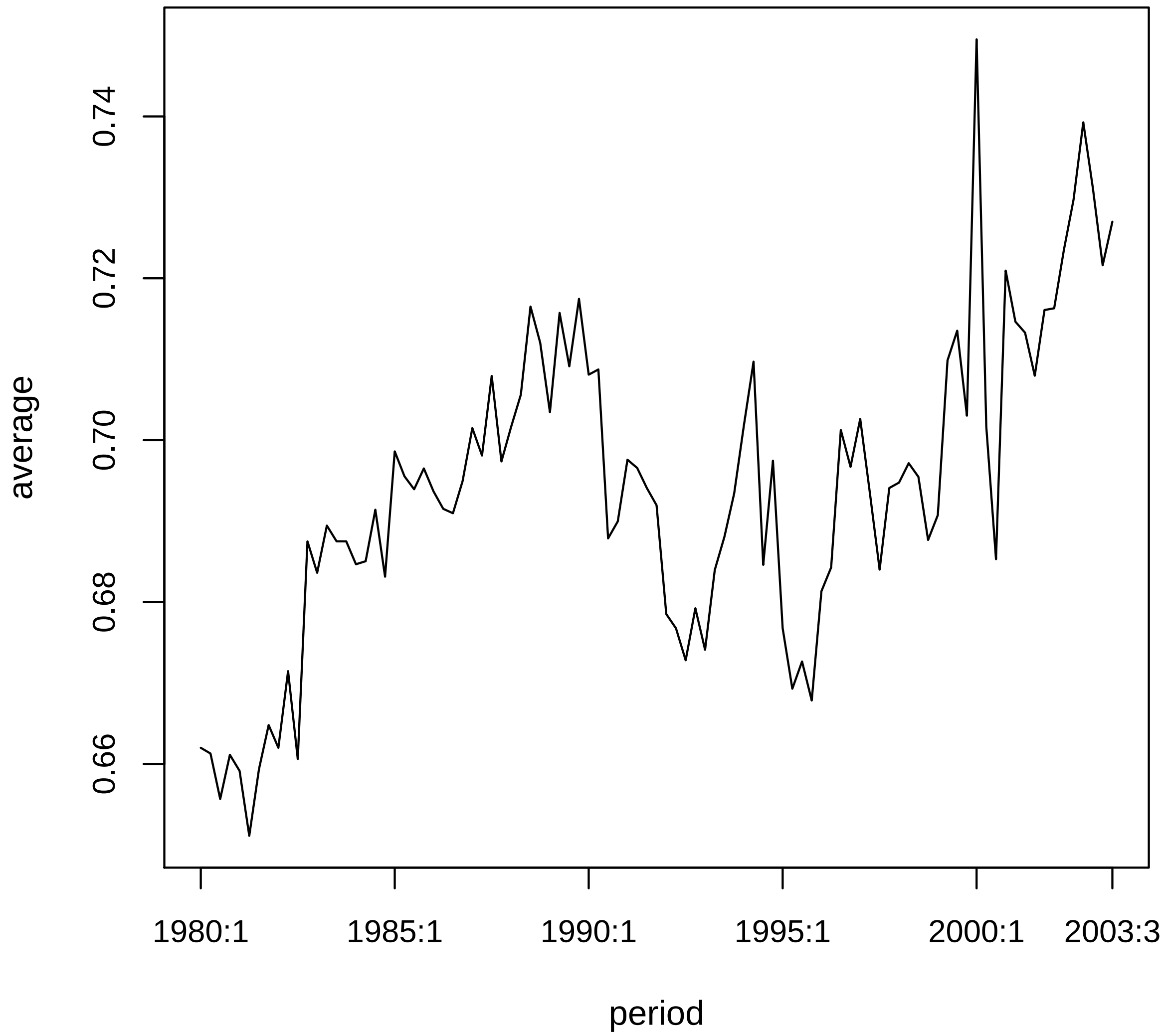

Figure 5: Median Hyperbolic Graph Efficiency Estimates

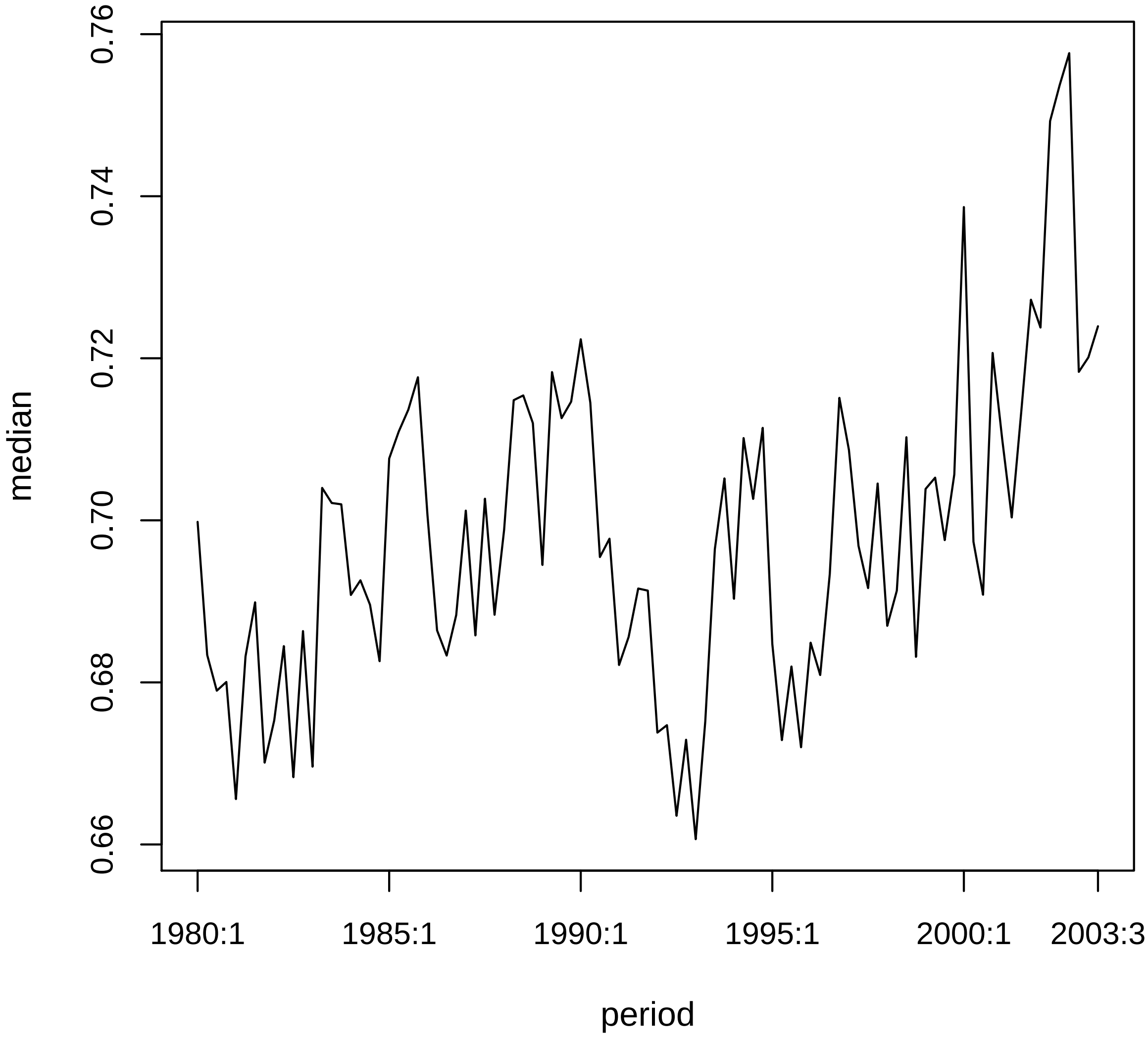

Figure 6: Variance of Hyperbolic Graph Efficiency Estimates