New Evidence on the Fed's Productivity in Providing Payments Services

Authors	R. Alton Gilbert, David C. Wheelock, and Paul W. Wilson
Working Paper Number	2002-020A
Citable Link	https://doi.org/10.20955/wp.2002.020
Suggested Citation	Gilbert, R.A., Wheelock, D.C., Wilson, P.W.; New Evidence on the Fed's Productivity in Providing Payments Services, Federal Reserve Bank of St. Louis Working Paper 2002-020. URL https://doi.org/10.20955/wp.2002.020

Published In	Journal of Banking & Finance
Publisher Link	https://doi.org/10.1016/j.jbankfin.2003.08.006

New Evidence on the Fed's Productivity in Providing Payments Services

R. Alton Gilbert

David C. Wheelock

and

Paul W. Wilson*

September 5, 2002

ABSTRACT

As the dominant provider of payments services, the efficiency with which the Federal Reserve provides such services in an important public policy issue. This paper examines the productivity of Federal Reserve check-processing offices during 1980-1999 using non-parametric estimation methods and newly developed methods for non-parametric inference and hypothesis testing. The results support prior studies that found little initial improvement in the Fed's efficiency with the imposition of pricing for Federal Reserve services in 1982. However, we find that median productivity improved substantially during the 1990s, and the dispersion across Fed offices declined.

JEL classification numbers: C1, G2, L2

---

*Gilbert and Wheelock: Research Department, Federal Reserve Bank of St. Louis, P. O. Box 442, St. Louis, MO 63166-0442. Wilson: Department of Economics, University of Texas, Austin, Texas 78712.

This research was conducted while Wilson was a visiting scholar in the Research Department of the Federal Reserve Bank of St. Louis. We thank Paul Bauer for discussions at an early stage of this project and Heidi Beyer for research assistance.

1. INTRODUCTION

The Federal Reserve is the largest provider of payments services in the world, offering electronic funds transfer, automatic clearinghouse, check clearing, and cash services. Despite inroads by electronic funds transfer and ACH, the processing of paper checks accounts for some 80 percent of all payments revenue and cost for Fed payments operations. The Federal Reserve processes approximately one-half of all checks deposited with U.S. banks other than those on which the checks are drawn (Federal Reserve System, 2002). In 2001, the Fed processed 16,905 million commercial checks, on which it collected $765 million in revenue and entailed operating expenses of $684 million (Board of Governors of the Federal Reserve System, 2001).

The Monetary Control Act of 1980 requires the Federal Reserve to recover its costs of providing payments services plus a "private sector adjustment factor" that reflects estimates of the taxes that a private firm would pay, and a return on investment for its shareholders. By requiring the Fed to price its services in this manner, this provision of the Monetary Control Act sought to use market discipline to improve the efficiency with which Fed offices provide payments services.

Recently, the Federal Reserve has reconsidered and reaffirmed its role as a provider of retail payments services (Rivlin *et al.,* 1998). One objective the Fed set for continuing to provide payments services was to encourage greater efficiency of the payments system. Evidence that Fed check facilities waste resources in providing payments services would indicate that the legal framework for pricing services imposed by the Monetary Control Act has not had its intended effect on the efficiency of Fed payments operations. Further it would indicate that the Fed could pursue greater efficiency in the operation of the payments system by focusing on the efficiency of its own payments facilities.

Previous studies of the efficiency with which the Fed provides check clearing services found little or no evidence of efficiency gains with the advent of pricing in the early 1980s. The increased availability of data with the passage of time, however, as well as recent advances in econometric methodology, provide an opportunity to gain insights that were not possible in earlier studies of Fed payments services.

This paper examines the productivity and technical efficiency of Fed check-processing offices using a non-parametric distance function estimator and newly-developed methods for non-parametric inference and hypothesis testing. We perform tests of several model restrictions, including independence of the inefficiency process, constant returns to scale, and the appropriate number of outputs. We find clear evidence that Federal Reserve check processing facilities have become more productive over time, consistent with the goals of the Monetary Control Act of 1980, but that substantial improvement did not occur until the late 1980s. Unlike previous studies, we present evidence of the statistical precision of our productivity estimates for individual offices. Finally, we determine that the technology of check processing is characterized by variable returns to scale, though we fail to reject operation at constant returns for any individual Fed offices.

The next section presents a conceptual framework for analyzing the effects of pricing on the incentives of Fed check office managers to operate efficiently. Section 3 briefly discusses the findings of previous studies of the efficiency of Fed check-processing since the implementation of pricing. Section 4 presents our statistical model, Section 5 describes our data, and Section 6 presents results. Section 7 concludes.

2. CONCEPTUAL FRAMEWORK

The conceptual framework for analyzing the effects of pricing on the incentives for the managers of Fed check offices to operate efficiently includes a production function and a cost function. Figure 1 illustrates a production function with one type of input and one type of output. The curve labeled “frontier” is the maximum output possible for each level of the input. None of the offices could operate with combinations of the input and output to the left of the curve. Any office with a combination off the frontier, such as Office A, is inefficient. Office A could produce more output with the same level of input, or produce the same output with less of the input.

For a given input price, it is possible to derive a frontier cost curve from the production function in Figure 1. To calculate the average cost associated with each level of output on the frontier cost curve in Figure 2, multiply the price of a unit of the input by the number of units of the input on the production function (Figure 1), and divide by the level of output. The frontier cost curve indicates the minimum average cost possible for each level of output.

Figure 3 illustrates the effects of the Monetary Control Act pricing provision on the price for check clearing services set by a Fed check office and the quantity of check services demanded by banks. The curve labeled "frontier cost curve" is the minimum possible average cost for each level of output, based on existing technology for clearing checks. The cost curves in Figure 3 incorporate the private sector adjustment factor that Federal Reserve offices are required to include in their pricing. The average cost of clearing checks at Office A is above the minimum possible average cost at each level of output.

Figure 3 includes two demand curves for the check clearing services of Office A. The curve labled D_M is the demand by Federal Reserve member banks for the check clearing services of Office A. Prior to the Monetary Control Act, only member banks had direct access to Fed check clearing services, and they were charged an explicit price of zero for those services. Thus, in the environment that existed before the Act, Office A would provide check clearing services of Q_3. The cost structure of Office A would not affect the quantity of services it provided to member banks.

The second demand curve in Figure 3, labeled D_T, is the demand by all banks for the check clearing services of Office A. There is reason to believe that permitting all banks to have direct access to Reserve Bank services had only a small effect on the demand for check clearing services (shifting the demand curve to the right). Prior to the Monetary Control Act, nonmember banks had indirect access to Fed services through correspondents that were Fed members.

Under the pricing regime imposed by the Monetary Control Act, the cost structure of a Reserve Bank office affects the quantity of output it provides. Office A charges the price P_1 and provides Q_1 of check collection services. Office A could increase the quantity of its check collection services to Q_2 by lowering its cost structure to that of the frontier cost curve. We assume that the managers of Fed check offices prefer higher check volume, within the pricing guidelines imposed by the Monetary Control Act. Figure 3 illustrates how the Act's pricing requirement gives an incentive to Fed check office managers to eliminate waste in their operations. Because the managers of offices with average cost highest above the frontier cost curve have the most to gain from increased efficiency, we would expect the dispersion of efficiency across Fed offices to decline after the imposition of pricing.

3. PREVIOUS STUDIES

The first studies of the efficiency of Federal Reserve check processing after the implementation of pricing mandated by the Monetary Control Act concluded that the pricing regime had improved resource allocation in the processing of checks. For example, whereas Humphrey (1981) found evidence of scale diseconomies at large Fed check offices during the 1970s, Humphrey (1985) found that by 1983 no Fed office experienced diseconomies, and concluded that "the pricing of the Federal Reserve's check service has clearly improved resource allocation for society as a whole" (p. 49).

More recent studies of the Fed's efficiency have tended to support Humphrey's (1985) findings about scale efficiency, but nevertheless 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) find evidence of considerable cost inefficiency at Fed check offices both before and after the implementation of pricing, with no significant difference in average inefficiency between the two periods. Further, they conclude that during 1983-90, Fed check facilities experienced a slight, though statistically insignificant, decline in average productivity.

Bauer and Ferrier (1996) use quarterly data for 1990-94 to estimate cost functions for Fed check processing, wire transfer, and ACH transfer services. For check processing, Bauer and Ferrier (1996) find that both average cost inefficiency and the dispersion of inefficiency across Fed offices were high during this period. Further, Bauer and Ferrier (1996) detect evidence of technological regress in check processing during the early 1990s, which they associate with declining processing volume at some sites, migration of "high-quality" check business (*e.g.,* Social Security and payroll checks) to ACH, and improved quality of Fed check services (*e.g.,* application of magnetic ink character recognition).

The evidence presented by Bauer and Hancock (1993) and by Bauer and Ferrier (1996) suggests that the efficiency with which the Fed provides check processing services did not improve with the implementation of the pricing regime. The Fed has retained significant market share in the processing of checks, however, and its volumes continued to rise through 1999. Further, although both studies employ fairly flexible methods to estimate cost efficiency, and report results that are robust to different methods, alternative estimation methods exist that offer even more flexibility. Hence, the Fed's continued presence in check clearing, the benefit of additional years of data since the advent of pricing to study the efficiency with which the Fed provides payments services, and the availability of flexible estimation methods (and newly developed means of testing hypotheses based on those methods) justify and enable a new look at the productivity of the Fed's provision of check clearing services.

4. THE STATISTICAL MODEL

We use data envelopment analysis (DEA), which is a non-parametric distance function estimator, to estimate the productivity of Federal Reserve offices in providing check clearing services. DEA has been used widely to study efficiency, but almost never with any attempts at statistical inference. Indeed, DEA and similar estimators are often said to be *deterministic* or *non-stochastic.* They are, however, actually *estimators* of *unknown* distance functions and, consequently, statistical inference is necessary to learn what an *estimate* might reveal about a true distance. Recently, methods of statistical inference have been developed for DEA and similar estimators.[1] In the present context, these methods permit us to discriminate among alternative models of Federal Reserve check production, to test for economies of scale, and to test for differences in productivity across different Fed offices. This section lays out our statistical model, focusing on the assumptions required to estimate the model and test hypotheses. The specific application to Fed check offices and our data are described in Section 5.

We begin by defining the production set

$$\mathcal{P} = \{(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \text{ can produce } \boldsymbol{y}\} \subset \mathbb{R}_{+}^{p+q}, \tag{4.1}$$

where $\boldsymbol{x} \in \mathbb{R}_{+}^{p}$ denotes a vector of p inputs and $\boldsymbol{y} \in \mathbb{R}_{+}^{q}$ denotes a vector of q outputs. The boundary of $\mathcal{P}$, denoted $\mathcal{P}^{\partial}$, is frequently referred to as the ***technology*** or the ***production frontier,*** and is given by the intersection of $\mathcal{P}$ and the closure of its complement.

Given a point $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathbb{R}^{p+q}$, we measure distance from $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ to $\mathcal{P}^{\partial}$ by the Shephard (1970) input distance function

$$\delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{P}) \equiv \sup\{\theta > 0 | (\theta^{-1}\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{P}\}, \tag{4.2}$$

which is merely a normalized Euclidean distance measure and provides a meaningful measure of input technical efficiency. For $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{P}$, this measure provides an indication of whether, and by how much, the input vector $\boldsymbol{x}_0$ could feasibly be scaled back without reducing the quantities of outputs $\boldsymbol{y}_0$. Clearly, $\delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{P}) \geq 1$ for all $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{P}$. If $\delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{P}) = 1$, then the production unit (a Fed check office in our context) is technically efficient; given the production possibilities defined by $\mathcal{P}$ in (4.1), the unit cannot reduce its input quantities without simultaneously reducing output quantities. In this

[1] See Simar and Wilson (2000b) for a survey of these methods and additional discussion.

case, $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{P}^\partial$. On the other hand, if $\delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{P}) > 1$, then the unit is technically inefficient; given $\mathcal{P}$, the unit could reduce input quantities without reducing output quantities.

We make several standard assumptions about the production set $\mathcal{P}$ to enable estimation of inefficiency: (i) $\mathcal{P}$ is closed and convex, the corresponding output feasibility sets are closed, convex, and bounded for all $\boldsymbol{x} \in \mathbb{R}^p_+$, and the corresponding input requirement sets are closed and convex for all $\boldsymbol{y} \in \mathbb{R}^q_+$; (ii) all production requires use of some inputs; and (iii) both inputs and outputs are strongly disposable. In addition, we assume (iv) that the set of n sample observations on inputs and outputs, $\mathcal{S}_n = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^n$, results from n independent draws from a probability density function $f(\boldsymbol{x}, \boldsymbol{y})$ with bounded support over $\mathcal{P}$. We assume (v) $f(\boldsymbol{x}, \boldsymbol{y})$ is strictly positive for all $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^\partial$, and starting from any point in $\mathcal{P}^\partial$, $f(\boldsymbol{x}, \boldsymbol{y})$ is continuous in any direction toward the interior of $\mathcal{P}$. Finally, we assume (vi) the distance function $\delta(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ is Lipschitz continuous in both arguments for $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$.

Together, assumptions (i)–(vi) define the data-generating process that produces the sample observations in $\mathcal{S}_n$, and permit statistical estimation and inference about the unobserved technology $\mathcal{P}^\partial$ and distance function $\delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{P})$. Assumptions (i)–(iii) are standard assumptions that relate to the economic aspects of the problem (see Shephard (1970) and Färe (1988)). Assumptions (iv)–(vi) are statistical in nature and are necessary to establish statistical properties of the estimators we use.[2]

We use the convex hull of the free disposal hull of $\mathcal{S}_n$, denoted $\widehat{\mathcal{P}}$, to estimate $\mathcal{P}$ and to construct an estimator

$$\delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \widehat{\mathcal{P}}) = 1/\left\{\min\left[\theta > 0 \mid \boldsymbol{Y}\boldsymbol{q} \geq \boldsymbol{y}_0,\ \boldsymbol{X}\boldsymbol{z} \leq \theta\boldsymbol{x}_0,\ \boldsymbol{i}'\boldsymbol{q} = 1,\ \boldsymbol{q} \in \mathbb{R}^n_+\right]\right\} \tag{4.3}$$

of the Shephard input distance function defined in (4.2), where $\boldsymbol{Y} = [\,\boldsymbol{y}_1 \quad \ldots \quad \boldsymbol{y}_n\,]$, $\boldsymbol{X} = [\,\boldsymbol{x}_1 \quad \ldots \quad \boldsymbol{x}_n\,]$, $\boldsymbol{i}$ denotes an $(n \times 1)$ vector of ones, and $\boldsymbol{q}$ is an $(n \times 1)$ vector

[2]See Simar and Wilson (2000a) for additional discussion of the assumptions used here, and Simar and Wilson (2000b) for a survey of estimation and inference methods for distance function estimation.

of intensity variables whose values are determined by solution of the linear programs in each case. The estimator $\delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \widehat{\mathcal{P}})$ measures normalized distance from a point $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ to the boundary of $\widehat{\mathcal{P}}$, and hence gives an estimate of the distances from $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ to $\mathcal{P}^\partial$. Korostelev *et al.* (1995) prove that $\widehat{\mathcal{P}}$ is a consistent estimator of $\mathcal{P}$ under conditions met by assumptions (i)–(vi) listed above, and Kneip *et al.* (1998) prove consistency of the distance function estimator under assumptions (i)–(vi) and establish its convergence rate of $O_p\left(n^{-\frac{2}{p+q+1}}\right)$.[3]

Unfortunately, few results exist on the sampling distribution of the distance function estimator in (4.3), which makes statistical inference using that estimator difficult. Bootstrap methods described in Simar and Wilson (1998, 2000a) allow one to approximate the asymptotic distribution of distance function estimators in multivariate settings, however, and hence to make inferences about the true distance function in (4.2).[4]

A second complication arises because the estimator $\delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \widehat{\mathcal{P}})$ is biased downward, *i.e.*, $\delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \widehat{\mathcal{P}}) \leq \delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{P})$ because $\widehat{\mathcal{P}} \subseteq \mathcal{P}$. Intuitively, the bias arises because the estimate of the production frontier is based on actual observations, so in finite samples the location of the estimated frontier will lie on or below the true frontier (assuming no measurement error). Hence, estimates of the productivity of individual offices will be biased upward. Fortunately, this bias can be corrected using the heterogeneous bootstrap method of Simar and Wilson (2000a). We report both original and bias-corrected estimates for comparison.

Comparison with other Estimation Methods

[3]The rate of convergence is slow, as is typical in non-parametric estimation; the rate becomes even slower as $p+q$ is increased—this is the well-known curse of dimensionality that commonly plagues non-parametric estimators. The free disposal hull estimator used by Bauer and Hancock (1993) relaxes the convexity assumption, but otherwise is similar to the estimator in (4.3). Imposing the convexity assumption yields a slightly faster rate of convergence; see Park *et al.* (2000) for details.

[4]Gijbels *et al.* (1999) derived the asymptotic distribution of the output distance function estimator corresponding to (4.3) for the special case of one input and one output ($p = q = 1$), along with an analytic expression for its large sample bias and variance; these results easily extend to the input distance function estimator in (4.3). Unfortunately, however, derivation of similar results for the more general multivariate setting where $p + q > 2$ is complicated by the radial nature of the distance functions and the complexity of the estimated frontier.

With the ability to make statistical inferences, non-parametric distance function estimators such as (4.3) offer several advantages over the parametric and semi-non-parametric estimators that have been used widely in studies of efficiency. Being fully non-parametric, the estimator in (4.3) can be used without specifying functional forms for the technology or the distribution of the inefficiency process. By contrast, parametric estimation often has involved the specification of a translog cost function containing a two-sided random noise term and a one-sided random inefficiency term. Estimation of the parameters of the model can be done by the method of maximum likelihood, and the inefficiency term can then be estimated using the methods of Aigner *et al.* (1977), Meeusen and van den Broeck (1977), and Jondrow *et al.* (1982). Meaningful estimates of efficiency are possible only if the translog function accurately represents the underlying cost relationship, however, and in contexts where the sizes of production units varies widely, as is the case with Fed check offices, researchers have found the translog function to mis-specify cost relationships.[5]

To illustrate the relevance of assumptions about the shape of the cost curve for estimating the efficiency of check offices, consider the efficiency of Office A in Figure 3. The difference between the average cost of Office A and the average cost of an office on the frontier cost curve with the same level of output is a measure of the inefficiency of Office A. Obviously, if the cost curve has a different shape, the estimate of inefficiency of Office A is likely to be different.

To overcome the problem of mis-specification inherent in parametric models, the translog function is often augmented with sine and cosine terms to improve model fit; the resulting models are sometimes labeled *flexible Fourier*. The flexible Fourier approach introduces several problems, however, that are often ignored by practitioners. One difficulty is in choosing the number of sine and cosine terms to include in the model. If too

[5]During the fourth quarter of 1999, for example, the number of forward commercial and government check items processed at different Fed offices ranged from 22,141,000 items to 280,006,000 items. Specification tests reported in McAllister and McManus (1993) and Wheelock and Wilson (2001) reject the translog cost function for commercial banks, whose sizes also vary over a wide range. Wilson and Carey (2002) report similar results for hospitals.

many terms are included, over-fitting will necessarily result and the model will explain noise as well as the functional relationship. If too few terms are included, however, the problem of mis-specification will remain.[6]

Additional problems with the flexible Fourier approach arise in the estimation of inefficiency. The typical method of modeling inefficiency as a one-sided component of a composite error term necessitates maximum likelihood estimation. Maximum likelihood, however, is often not tractable when sample size, and hence the optimal number of trigonometric terms included in the model, is large. Some authors have used panel data methods to estimate fixed effects, and then used these to construct estimates of inefficiency along the lines of Schmidt and Sickles (1984). Doing so, however, imposes the assumption that inefficiency is constant over time, which assumes no learning on the part of managers. Furthermore, efficiency estimates from this approach are consistent only as the number of time periods tends to infinity. Over long time horizons, however, the assumption that inefficiency does not change is increasingly suspect. Hence, at least in our application, the flexible Fourier estimator is not as advantageous as our fully non-parametric estimator.[7]

5. FEDERAL RESERVE CHECK PROCESSING

The clearing of checks involves receiving checks from depositing banks (defined broadly to include all depository institutions), sorting them, crediting the accounts of the depositing banks, and delivering the checks to the banks upon which they are drawn. Such

[6] Results reported by Gallant (1981; 1982) suggest that the optimal number of terms to include is of order $n^{2/3}$, where n represents sample size. The choice of the number of Fourier terms involves a tradeoff between bias and variance – adding terms reduces the bias of the predicted values but increases variance, whereas reducing the number of terms reduces variance but at the expense of increased bias.

[7] Whereas our non-parametric estimator (4.3) avoids the necessity of imposing a particular cost relationship and permits efficiency estimates to vary over time, some might argue that our approach compares unfavorably to a regression framework because it does not permit the inclusion of a noise term to capture potential measurement error in the left-hand side variable. However, besides incurring one or more of the specification problems noted above in using a regression framework to estimate inefficiency, two features of our study work to mitigate the effect of omitting a noise term: first, we use highly aggregated data, and so by standard central limit theorem arguments, the effect of any measurement error is smoothed away; second, we employ techniques to scan our data for outliers that might result from measurement errors or other problems. In addition, we employ a statistical framework to allow for uncertainty due to sampling error that does not require the inclusion of an explicit error term.

"forward item" processing is the main source of revenue and total cost for Fed check operations. 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. Finally, Fed check offices entail costs associated with making adjustments necessitated by processing and other errors. Following the convention of other studies, we focus here on the forward processing of commercial and Federal Government check items.

The methods we use permit the estimation of productivity of check offices with multiple outputs. In addition to treating the number of forward items processed as an output, we consider whether the number of endpoints served by an office should be treated as a second output. An endpoint is an office of a depository institution to which a Fed office delivers check items. Other studies have suggested that differences in the number or location of endpoints may help explain why some Fed check offices appear less efficient than others. The number of endpoints (or a measure of the location of endpoints) could be treated as an environmental characteristic affecting the efficiency of check processing. Alternatively, the number of endpoints might be thought of as a measure of the level of service provided by a check office – an office serving many endpoints, all else equal, is providing a higher level of service than an office serving fewer endpoints. We estimate the productivity of Fed check offices both for a single-output model (number of forward commercial and Federal Government check items processed), as well as a two-output model that includes the number of endpoints as a second output. We perform a statistical test to determine whether the data support the treatment of the number of endpoints served as a distinct output. Our data consist of quarterly observations for each Federal Reserve Bank main office, branch office, and dedicated check processing center from 1980:Q1 through 1999:Q4, totaling 3761 office-quarters.[8]

[8]Quarterly data on the number of items processed and number of endpoints served by each office are

Federal Reserve check facilities use a variety of inputs to process checks and deliver them to paying banks. Estimation of productivity using statistical methods requires the specification of a model of the production process with a limited number of inputs. We follow the convention of other studies of check office productivity (Bauer and Hancock, 1993, and Bauer and Ferrier, 1996) 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. Our model of productivity 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. Table 2 gives summary statistics for both inputs and outputs.

6. EMPIRICAL RESULTS

We estimated input distance functions for both the one- and two-output models described above. Because our estimator of the production frontier is based on the convex hull of the free-disposal hull of the sample observations, the efficiency estimates are not independent of one another, and a single outlier has the potential to severely distort efficiency estimates for possibly many observations. Using the outlier-detection technique described in Simar (2002), however, we found no evidence of outliers in our data that might serve to distort estimates of the production frontier.

Tests of Specification and Returns to Scale

Next, we examine our conjecture that the number of endpoints served by a check office is a distinct output of check production in addition to the number of items processed. If the number of endpoints is in fact irrelevant, including it would have no influence on the shape of the estimated production frontier. Figure 4 plots the bias-corrected distance function estimates for 1999:Q4 from the two-output model (model #2) as a function of the corresponding estimates from the single-output model (model #1). The same scale is used

from an internal Federal Reserve database containing Federal Reserve expense reports.

on both axes to facilitate comparison.[9] If the estimates from each model were identical, the points in Figure 4 would fall on a 45-degree line running from the lower-left to the upper-right corner of the figure. Several points lie below the 45-degree diagonal, however, indicating that treating the number of endpoints as an output increases the estimated efficiency of some sites.[10] Treating endpoints as an output also changes the shape of the estimated frontier, which would not be expected if endpoints were irrelevant to the production process.

A formal statistical test of the null hypothesis of one output (forward items processing) against the alternative hypothesis that Fed check office production is more appropriately modeled as involving two outputs (with number of endpoints as the second output) further indicates that the number of endpoints should be treated as a distinct output. The test, which is described in detail in Simar and Wilson (2001) and uses distance function estimates for all quarters in the sample, is based on the idea that under the null hypothesis, the irrelevant output will be unrelated to the true production frontier, $\mathcal{P}^{\partial}$. Statistics for the test are based on ratios and differences of distance function estimates from the two models; under the null, the statistics are expected to be small in value, *i.e.*, the distance functions will be similar. To carry out the test, which involves drawing inferences about the true distance function estimates for each observation, we use the heterogeneous bootstrap described in Simar and Wilson (2000a) to approximate the distribution of the true

[9] We used the bootstrap methods of Simar and Wilson (2000a) to produce estimates of the bias for each observation, which we subtracted from the original distance function estimates to obtain bias-corrected distance function estimates. Because the bias-corrected estimates are obtained by subtracting a potentially noisy estimate of bias from the original distance function estimates, the bias-corrected estimates might have higher mean-square error than the original estimates. To check this, we computed the value 1/3 times the square of the bootstrap bias estimate divided by the sample variance of the bootstrap estimates, which serves as an indicator of whether mean-square error is worsened when the bootstrap bias estimate is subtracted from the original estimate to obtain the bias-corrected estimate. As discussed in Simar and Wilson (2000a), this ratio should exceed unity if the bias-corrected estimator is to be used; otherwise, the bias-corrected estimator will likely have greater mean-square error than the original, uncorrected estimator. In every case, the ratio is well above unity, and so we rely on the bias-corrected estimator of the distance function.

[10] By construction, the uncorrected distance function estimates from model #2 are less than or equal to the corresponding uncorrected distance function estimates from model #1 due to the increased dimensionality in model #2. This is not true for the bias-corrected estimates, however.

distance function estimator and thereby derive p-values for the test.[11] Using 2000 bootstrap replications to obtain p-values, our test rejects the one-output model in favor of the two-output model with a p-value less than 0.0005 in each case.[12]

Next, we investigate returns to scale in Fed check processing. Using a bootstrap test described in Simar and Wilson (2002) we test the null hypothesis of globally constant returns to scale in the technology $\mathcal{P}^{\partial}$ of the two-output model versus the alternative hypothesis of variable returns to scale. Using the six statistics described by Simar and Wilson (2002) and 2,000 bootstrap replications, we reject the null hypothesis of constant returns with p-values of less than 0.001 for each statistic. We are unable to reject the hypothesis of operation at constant returns for any individual Fed office, however, and hence our results conform with conclusions about scale economies in Humphrey (1985) and other studies.[13]

Productivity Change

Because we use cross-sectional, time-series data, distance function estimates for a particular observation measure distance to the estimated boundary of the production set at a single point in time – the fourth quarter of 1999. Consequently, distance function estimates reflect *efficiency*, which relates a unit's performance to the current technology, only for 1999:Q4. However, changes in the distance function estimates for an office over time reflect changes in the *productivity* of that office. An indication of how productivity changed for the System as a whole can be obtained by aggregating across all offices.

Figure 5 plots the median and variance of the bias-corrected distance function estimates

[11] We use the heterogeneous bootstrap, rather than the homogeneous bootstrap of Simar and Wilson (1998), because the latter requires the true inefficiency estimates measured by (4.2) to be statistically independent of output levels and input mix. Using two bootstrap versions of the Kolmogorov- Smirnov test described in Wilson (2002), we reject the null hypothesis of independence for both the one- and two-output models with p-values of less than 0.0001.

[12] For each statistic, we find no bootstrap values among the 2000 bootstrap replications that are smaller than the original value computed from the sample; hence the estimated p value is less than $1/2000 = 0.0005$.

[13] As with the test for an irrelevant output, our test of returns to scale is based on the idea that under the null hypothesis, distance function estimates obtained while imposing constant returns should not differ greatly from corresponding estimates obtained without imposing constant returns. Although the Monte Carlo experiments in Simar and Wilson (2002) were based on the homogeneous bootstrap to reduce computational burden, our tests here are based on the heterogeneous bootstrap, having rejected independence between the input distance function values and the set of outputs and input angles.

from the two-output model for each quarter of the sample. The median is measured on the left vertical axis, while variance is measured on the right vertical axis. The median varies considerably over the sample period, but after increasing in the mid-1980s, it tends to decline over the remaining years through the 1990s. Because the median distance function estimates are smaller during much of the 1990s than before 1982, our results suggest that the median productivity of Federal Reserve check offices was higher by the end of our sample than before the implementation of pricing in 1982. We find that median productivity worsened initially after the implementation of pricing, however, consistent with the findings of Bauer and Hancock (1993). Finally, we find that the dispersion in productivity across offices was considerably smaller by the end of the sample period than it had been during the 1980s. It appears that pricing, and perhaps other factors, have narrowed productivity differences across Fed offices.[14]

Table 3 reports input distance function estimates from the two-output model for each Federal Reserve check-processing office in 1999:Q4, the final period of our sample. The column labeled $\widehat{\delta}$ gives the original distance function estimates, while the column labeled $\widehat{\widehat{\delta}}$ gives the bias-corrected distance function estimates obtained as described in footnote 9. The remaining columns of Table 3 contain estimated upper and lower bounds (a^*_α, b^*_α) for confidence intervals at $\alpha = .1$ and $\alpha = .05$ significance levels, respectively, which were obtained using the methods described in Simar and Wilson (2000a).[15]

The bias-corrected distance function estimates shown in Table 3 range from 1.0556 to 2.0162, with a mean of 1.4528. Hence, our estimates indicate that in 1999:Q4 the average Federal Reserve check processing site could have feasibly reduced its inputs by a factor

[14] Even though we reject the one-output model in favor of the two-output model, median distance function estimates based on the one-output model (plotted in Figure 6) show the same trend as estimates for the two-output model. These results further support our conclusion that the System's median productivity improved beginning in the mid-1980s through the 1990s.

[15] Note that the original estimate of the input distance function, $\widehat{\delta}$, lies outside the corresponding estimated confidence interval in each case. As discussed in Simar and Wilson (2000a), the confidence interval estimates incorporate an implicit bias correction that does not depend on an explicit estimate of the bias. The original distance function estimates always lie to the left of the corresponding confidence interval estimates, reflecting the downward bias of the input distance function estimator.

of $1/1.4528 = 0.6883$. As is typical in efficiency estimation, however, the distribution of efficiency estimates is skewed. Figure 7 shows a non-parametric kernel estimate of the density of the bias-corrected distance function estimates for 1999:Q4.[16] In addition to skewness, the estimated density first increases as one moves to the right from 1.0, before eventually decreasing on the right. This suggests that in terms of technical efficiency, the sites are not clustered along the frontier; rather, only a few sites define the frontier estimate, with most lying in the interior of the estimated production set. Our results indicate, therefore, that despite improvement over time, at the end of the sample period many Fed check offices remained considerably less efficient than they could be.

7. SUMMARY AND CONCLUSIONS

The Monetary Control Act of 1980, among other things, sought to improve the efficiency with which the Federal Reserve provides payments services by requiring Fed offices to recover their costs of providing services plus a private sector adjustment factor. Although prior studies found that Fed offices operated at efficient scale in processing checks after the introduction of the pricing requirement, they also concluded that the introduction of pricing produced no improvement in overall operating efficiency. When the Fed reconsidered its role in the payments system in 1998, one objective it stated for continuing to provide retail payments services was to improve the efficiency of the payments system. Hence, evidence that Fed offices waste resources in the processing of checks would indicate that the Fed could contribute to the efficiency of the payments system by improving the efficiency of its own operations.

The present study presents new evidence on the productivity of Federal Reserve check offices using non-parametric estimation and recently developed methods of statistical inference for non-parametric estimators. Like prior studies, we treat forward check items

[16]The bandwidth for the density estimate was selected using a modified version of the normal reference rule suggested by Hjort and Jones (1996), which incorporates information from the third and fourth sample moments of the data; the bandwidth used was 0.2078. The density estimate was constructed using the reflection method described by Silverman (1986) to overcome the problem of bias near the left boundary.

processing as an output of Fed offices. However, we also treat the number of endpoints served by an office as a distinct, second output. Our specification tests indicate that treating the number of endpoints served as an output is appropriate.

We find that median productivity of Fed check offices has improved markedly since the implementation of pricing in the early 1980s, though most of the improvement has come since the late 1980s, after some regress in the middle part of that decade. Further, the variance in productivity across offices has also declined substantially.

Unlike previous studies, we report robust confidence intervals around inefficiency estimates for individual Fed offices. Hence, we are able to test hypotheses about differences in inefficiency across offices. We find that although the variance across offices has declined over time while median productivity has improved, significant differences across offices remain. We find that few offices operate close to the efficient frontier, suggesting that further improvements in efficiency are possible at many offices.

REFERENCES

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

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

Board of Governors of the Federal Reserve System (2001), *Annual Report.*

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

Federal Reserve System (2002), *Retail Payments Research Project.*

Gallant, A. R. (1981), On the bias in flexible functional forms and an essentially unbiased form, *Journal of Econometrics* 15, 211–45.

Gallant, A. R. (1982), Unbiased determinantion of production technologies, *Journal of Econometrics* 20, 285–323.

Gijbels, I., E. Mammen, B.U. Park, and L. Simar, (1999), On estimation of monotone and concave frontier functions, *Journal of the American Statistical Association* 94, 220–228.

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

Humphrey, D.B. (1981), Economies to scale in Federal Reserve check processing operations, *Journal of Econometrics* 15, 155–173.

Humphrey, D.B. (1985), Resource use in Federal Reserve check and ACH operations after pricing, *Journal of Bank Research* 16, 45–53.

Jondrow, J., C.A.K. Lovell, I. Materov, and P. Schmidt (1982), On the estimation of technical inefficiency in the stochastic frontier production function model, *Journal of Econometrics* 19, 233–238.

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

Korostelev, A., Simar, L. and A. Tsybakov (1995), On estimation of monotone and convex boundaries, *Publications de l'Institut de Statistique de l'Université de Paris XXXIX* 1, 3-18.

McAllister, P.H. and D.A. McManus (1993), Resolving the scale efficiency puzzle in banking, *Journal of Banking and Finance* 17, 389–405.

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

444.

Park, B., L. Simar, and C. Weiner (2000), The FDH estimator for productivity efficiency scores: Asymptotic properties, *Econometric Theory* 16, 855–877.

Rivlin, Alice M., Edward W. Kelley, Jr., Willliam J. McDonough, and Thomas C. Melzer (1998). *The Federal Reserve in the Payments Mechanism.* Washington, D.C.: Committee on the Federal Reserve in the Payments Mechanism, Federal Reserve System.

Schmidt, P. and R. Sickles (1984), Production frontiers and panel data, *Journal of Business and Economic Statistics* 2, 367–374.

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

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

Simar, L. (2002), "Detecting Outliers in Frontier Models: A Simple Approach," *Journal of Productivity Analysis,* forthcoming.

Simar, L. and P.W. Wilson (1998), Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models, *Management Science* 44(1), 49–61.

Simar, L. and P.W. Wilson (2000a), A general methodology for bootstrapping in nonparametric frontier models, *Journal of Applied Statistics* 27, 779–802.

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

Simar, L. and P.W. Wilson (2001), Testing restrictions in non-parametric efficiency models, *Communications in Statistics* 30, 159–184.

Simar, L. and P.W. Wilson (2002), Nonparametric tests of returns to scale, *European Journal of Operational Research* 139, 115–132.

Wheelock, D.C. and P.W. Wilson (2001), New evidence on returns to scale and product mix among U.S. commercial banks, *Journal of Monetary Economics* 47, 653–674.

Wilson, P.W. (2002), "Testing Independence in Models of Productive Efficiency," unpublished working paper, Department of Economics, University of Texas, Austin, Texas 78712.

Wilson, P.W. and K. Carey (2002), Nonparametric analysis of returns to scale in the U.S. hospital industry, unpublished working paper, Department of Economics, University of Texas, Austin, Texas.

TABLE 1

Definitions and Measurement of Inputs

1. *Personnel*—expenditures on personnel divided by the number of employee hours.

2. *Materials, Software, Equipment and Support*—expenditures are deflated by the following price measures:
   - *Materials:* GDP implicit price deflator (sa, 1996=100).
   - *Software:* private nonresidential fixed investment deflator for software (sa, 1996=100).
   - *Equipment:* for 1979–1989, PPI for check-handling machines (June 1985=100); for 1990-1999, PPI for the net output of select industries-office machines, n.e.c. (nsa, June 1985=100).
   - *Other Support:* GDP implicit price deflator (sa, 1996=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 (sa, 1996=100).
   - *Communications and Communications Support:* private nonresidential fixed investment deflator for communications equipment (sa, 1996=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, Outputs
(3761 observations)

Variable	Mean	Std. Dev.	Minimum	Maximum
Outputs:				
# checks	83376.11	774.99	10413.00	280006.00
# endpoints	412.66	4.67	32.00	1686.00
Inputs:				
Personnel	34553.67	396.25	4905.04	201529.15
Materials, etc.	1914.33	19.32	154.83	10630.28
Transit	720.73	9.58	62.76	4438.23
Facilities	1262.45	14.93	20.34	6677.62

TABLE 3
Input Distance Function Estimates for Model #2, 1999:Q4

Site	$\widehat{\delta}$	$\widehat{\widehat{\delta}}$	$a^*_{.1}$	$b^*_{.1}$	$a^*_{.05}$	$b^*_{.05}$
1	1.0000	—	—	—	—	—
2	1.0192	1.0556	1.0412	1.0711	1.0377	1.0727
3	1.0587	1.1026	1.0854	1.1216	1.0817	1.1257
4	1.3312	1.3784	1.3586	1.3985	1.3552	1.4026
5	1.5137	1.5600	1.5402	1.5844	1.5361	1.5878
6	1.1668	1.2318	1.2071	1.2556	1.2034	1.2597
7	1.1585	1.2166	1.1925	1.2402	1.1895	1.2464
8	1.0000	1.1168	1.0887	1.1429	1.0826	1.1474
9	1.1113	1.1779	1.1498	1.2098	1.1440	1.2168
10	1.0631	1.1470	1.1147	1.1802	1.1089	1.1844
11	1.0000	1.1176	1.0852	1.1504	1.0770	1.1554
12	1.7265	1.8105	1.7775	1.8429	1.7721	1.8497
13	1.3568	1.4482	1.4131	1.4814	1.4073	1.4875
14	1.3552	1.4463	1.4131	1.4816	1.4076	1.4881
15	1.6024	1.7010	1.6666	1.7357	1.6608	1.7413
16	1.1458	1.2173	1.1834	1.2523	1.1787	1.2595
17	1.0591	1.1565	1.1204	1.1903	1.1131	1.1955
18	1.8110	1.9319	1.8940	1.9700	1.8863	1.9754
19	1.2240	1.2973	1.2622	1.3381	1.2558	1.3464
20	1.3158	1.4041	1.3633	1.4425	1.3544	1.4511
21	1.2036	1.2837	1.2472	1.3266	1.2366	1.3310
22	1.3113	1.4303	1.3898	1.4720	1.3772	1.4768
23	1.0455	1.1593	1.1169	1.2004	1.1085	1.2098
24	1.5487	1.6610	1.6190	1.7031	1.6094	1.7096
25	1.4242	1.5265	1.4833	1.5714	1.4767	1.5785
26	1.3553	1.4558	1.4118	1.5001	1.4067	1.5076
27	1.5254	1.6499	1.6028	1.6917	1.5905	1.7029
28	1.0913	1.1779	1.1345	1.2244	1.1277	1.2322
29	1.5590	1.6784	1.6322	1.7225	1.6230	1.7302
30	1.0477	1.1542	1.1097	1.2034	1.1020	1.2157
31	1.3593	1.4573	1.4069	1.5085	1.3993	1.5173
32	1.2679	1.4140	1.3619	1.4626	1.3522	1.4731
33	1.8130	1.9411	1.8890	1.9914	1.8778	2.0005
34	1.1202	1.2439	1.1919	1.2972	1.1844	1.3079
35	1.2794	1.4236	1.3673	1.4826	1.3578	1.4942

TABLE 3 (continued)

Site	$\widehat{\delta}$	$\widehat{\widehat{\delta}}$	$a^*_{.1}$	$b^*_{.1}$	$a^*_{.05}$	$b^*_{.05}$
36	1.3489	1.5632	1.4907	1.6260	1.4710	1.6367
37	1.3232	1.5469	1.4724	1.6126	1.4604	1.6237
38	1.0000	1.8251	1.7375	1.8780	1.7099	1.8819
39	1.3494	1.4760	1.4023	1.5546	1.3912	1.5662
40	1.6042	1.8335	1.7466	1.9168	1.7342	1.9334
41	1.2378	1.4633	1.3776	1.5469	1.3595	1.5584
42	1.0000	1.6274	1.5299	1.7056	1.5083	1.7160
43	1.6600	2.0162	1.8796	2.1477	1.8508	2.1678
44	1.4103	1.7706	1.6127	1.9443	1.5812	1.9844
45	1.2245	1.6272	1.4316	1.8607	1.4050	1.9212

FIGURE 1

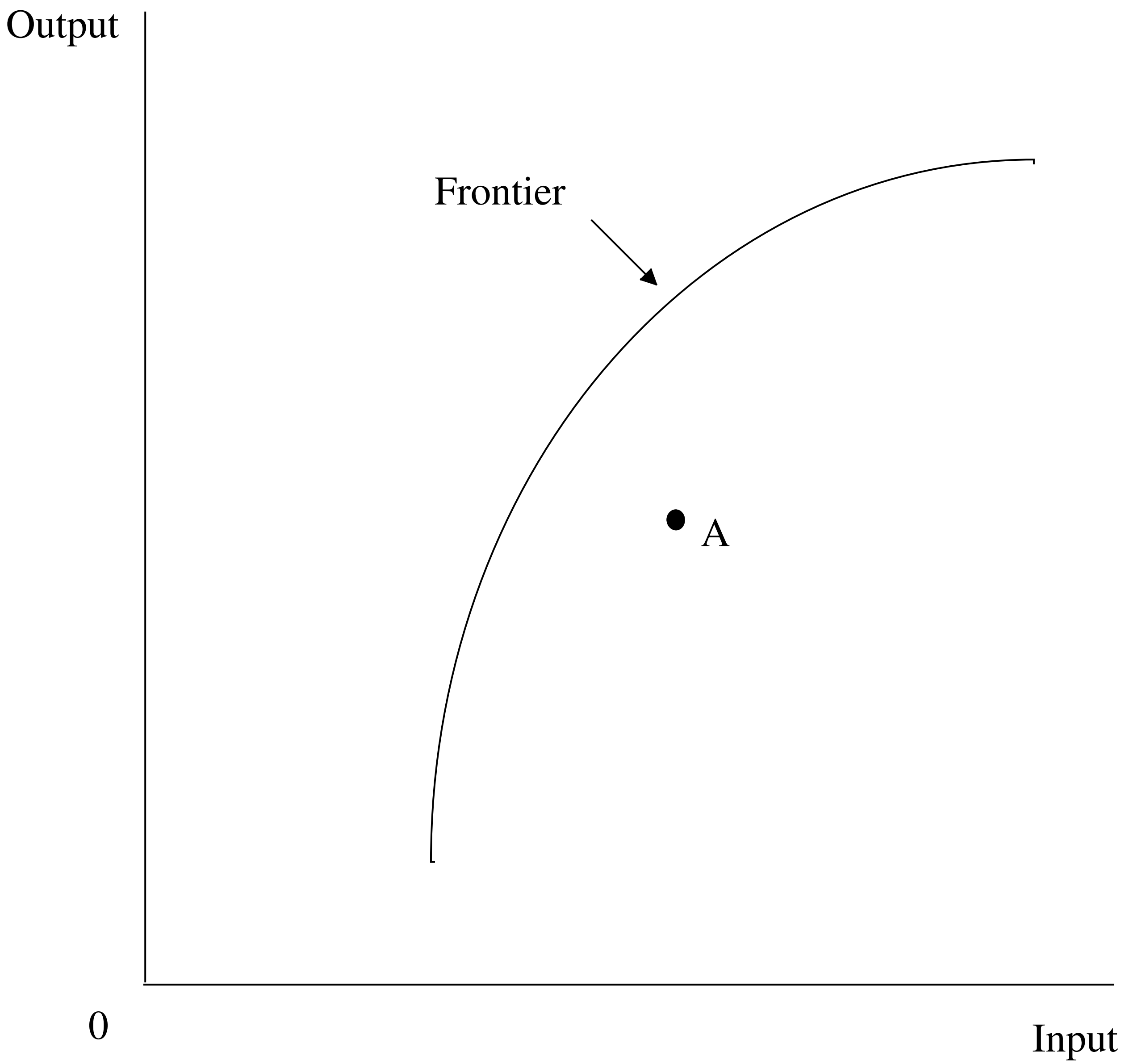


FIGURE 2

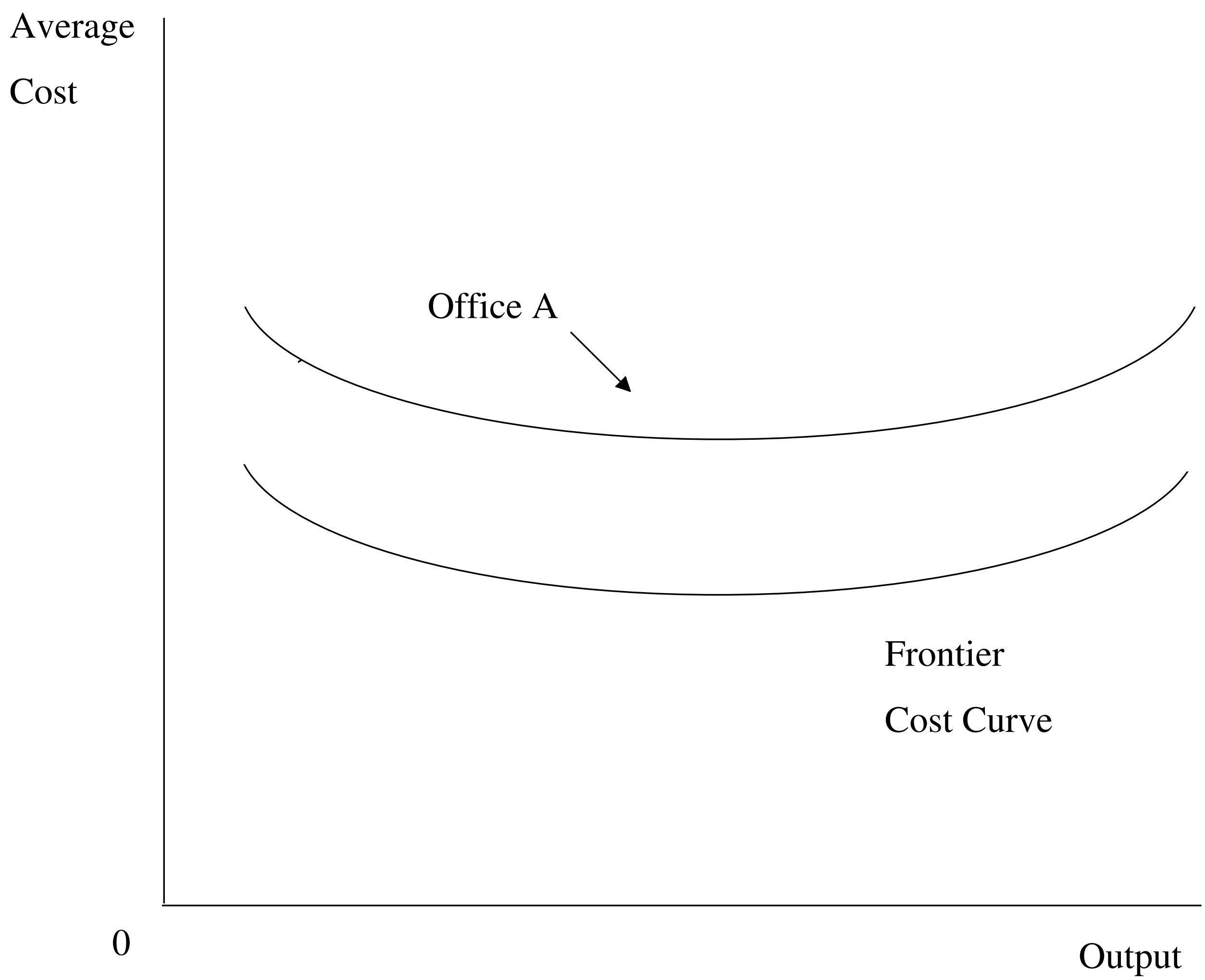


FIGURE 3

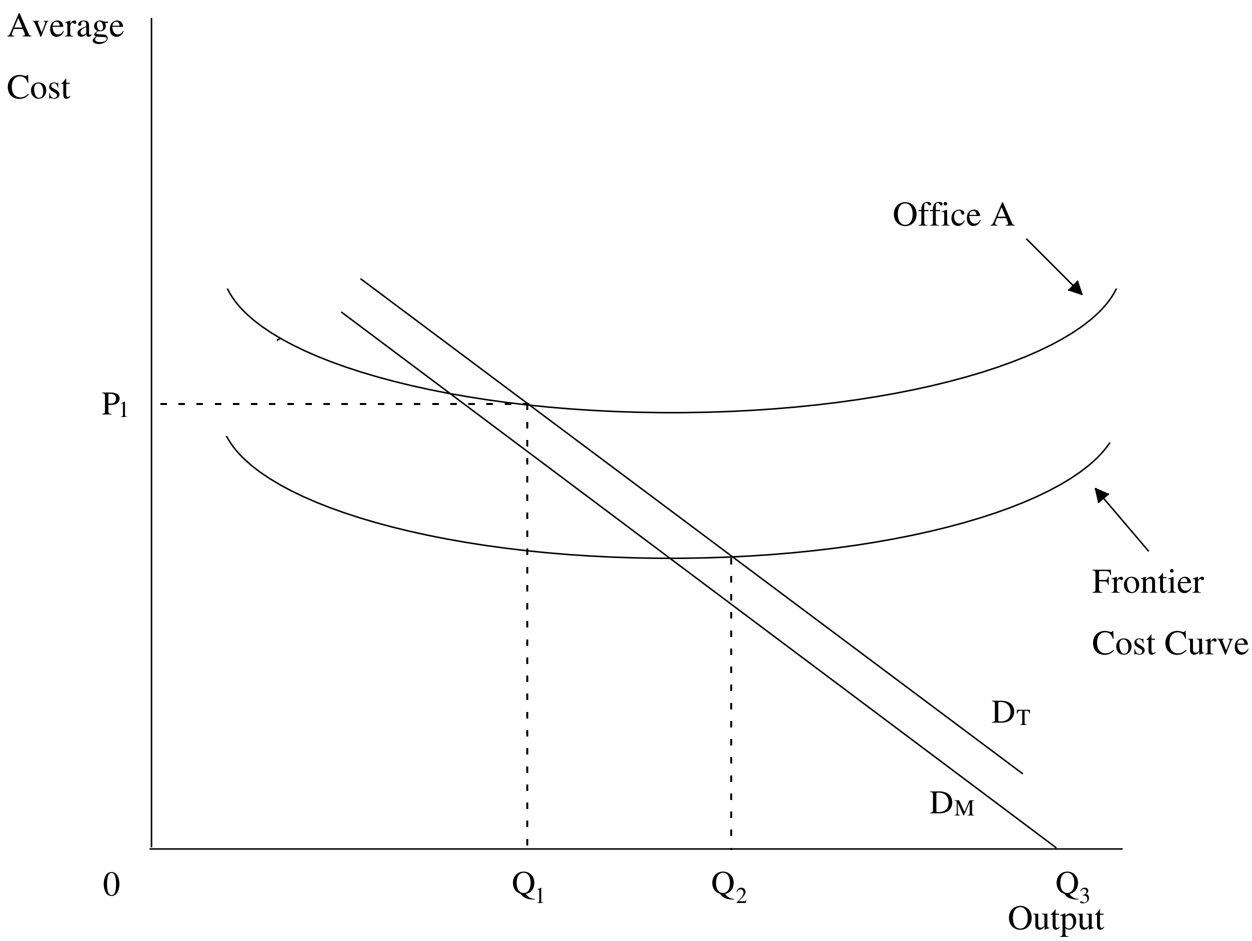

Average
Cost
Office A
P_1
Frontier
Cost Curve
D_T
D_M
0
Q_1
Q_2
Q_3
Output

FIGURE 4

Model #2 versus Model #1 Distance Function Estimates, 1999:Q4

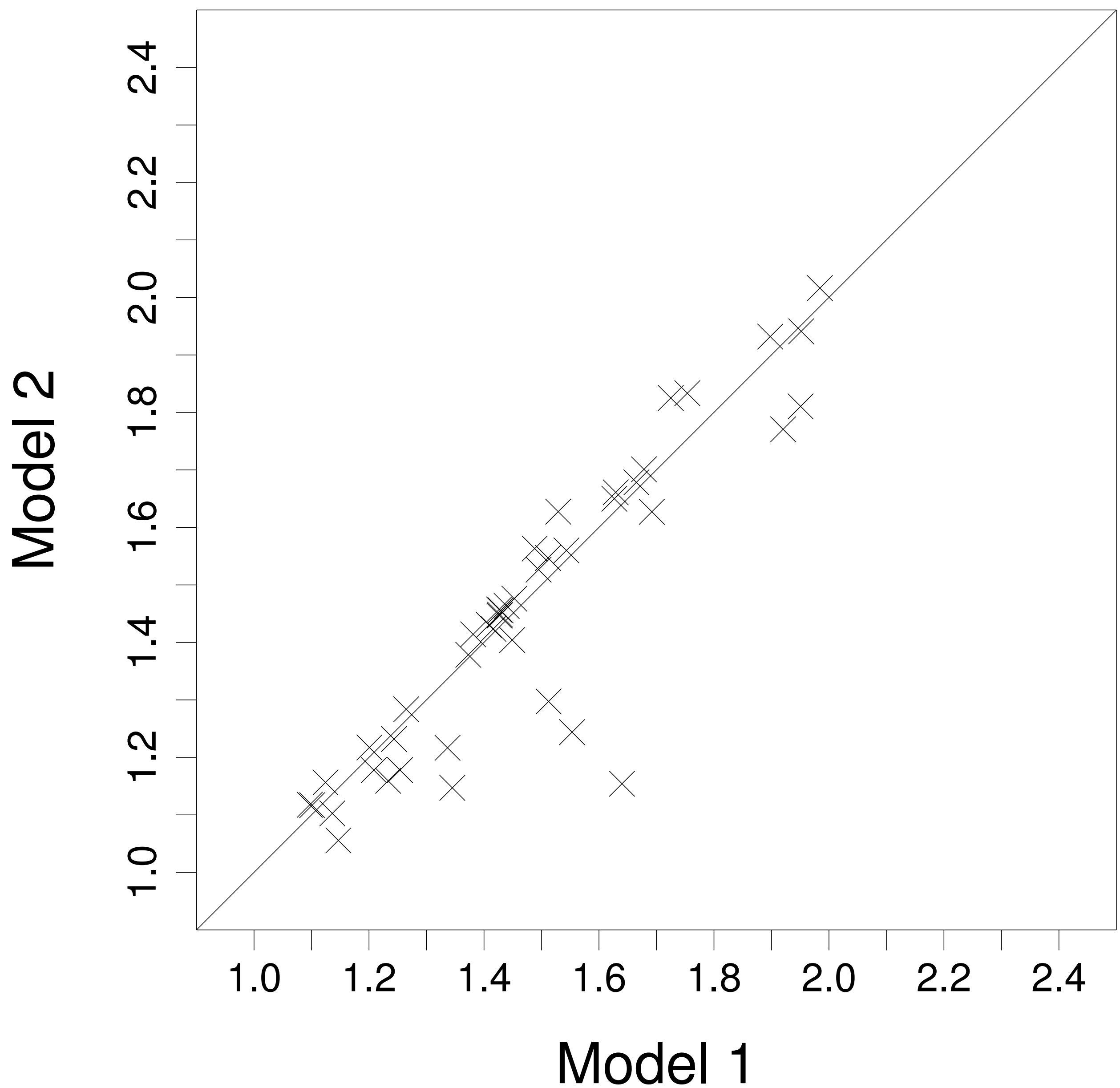
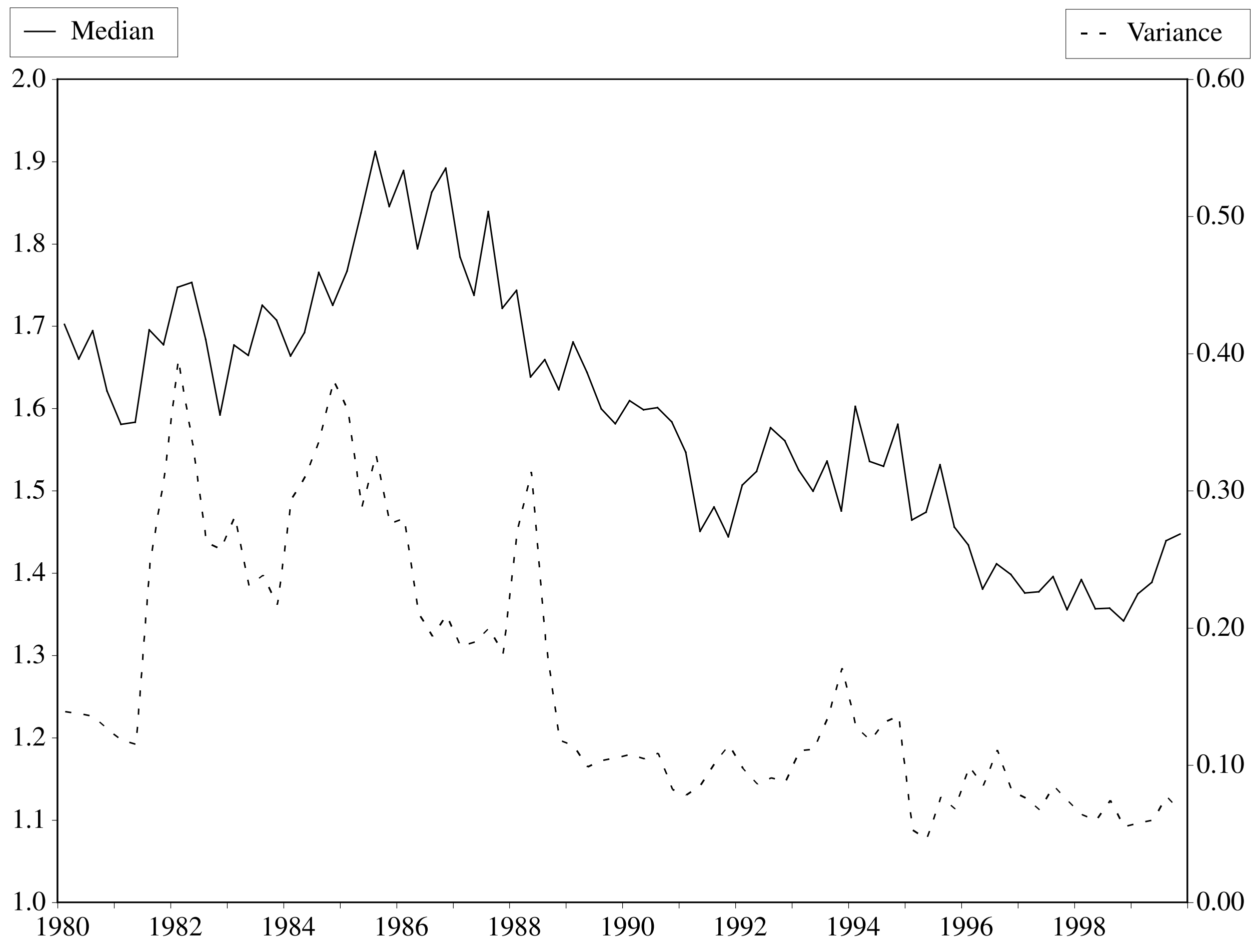


FIGURE 5

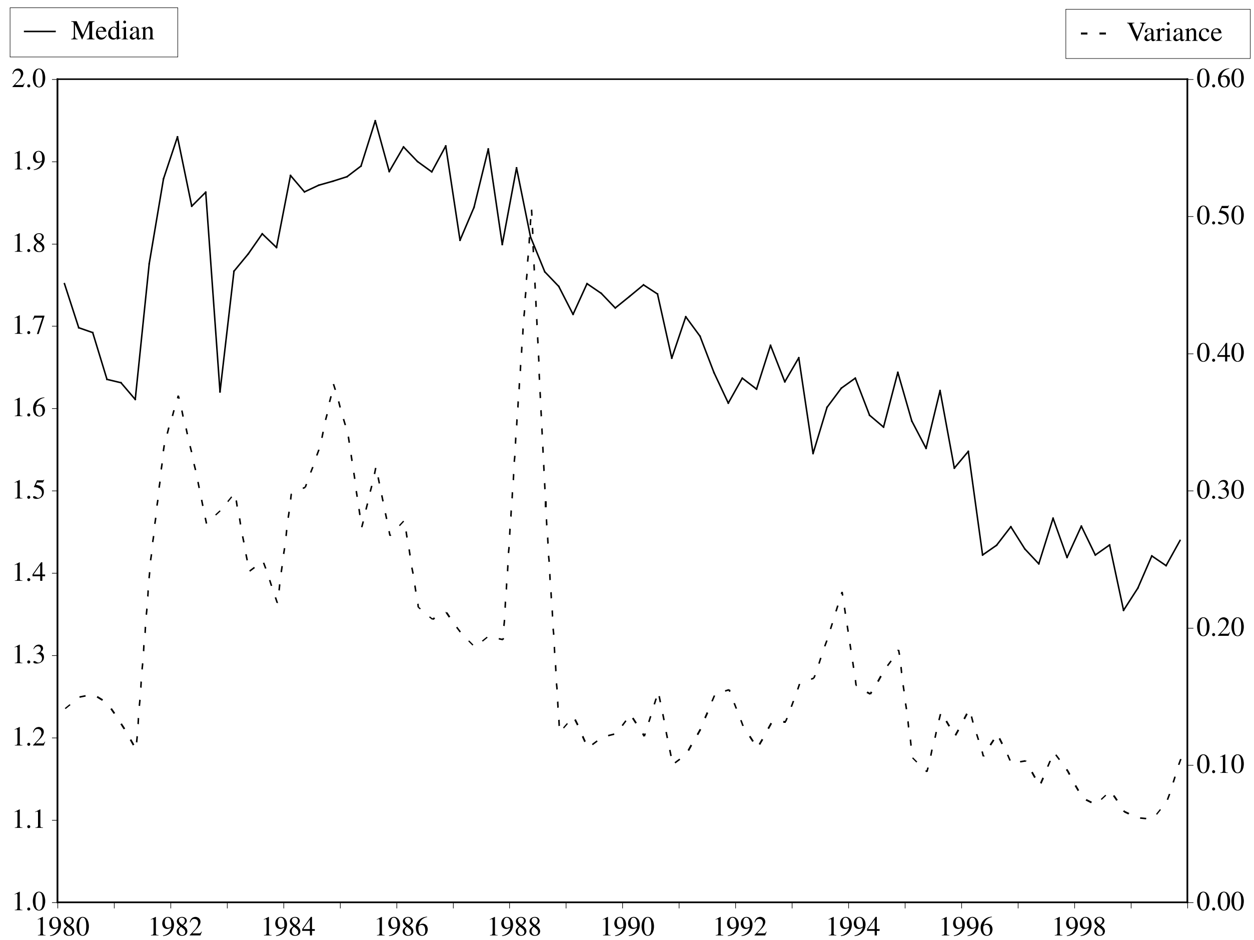
Median and Variance of Bias-Corrected Productivity Estimates
Across Fed Check-Processing Sites, Model 2 (two outputs),
1980:Q1–1999:Q4



Note: Solid line shows median productivity, measured on the left vertical axis; dashed line shows variance of estimated productivity, measured on the right vertical axis.

FIGURE 6

Median and Variance of Bias-Corrected Productivity Estimates
Across Fed Check-Processing Sites, Model 1 (one output),
1980:Q1–1999:Q4



Note: Solid line shows median productivity, measured on the left vertical axis; dashed line shows variance of estimated productivity, measured on the right vertical axis.

FIGURE 7

Kernel Estimate of Density of $\hat{\hat{\delta}}$
1999:Q4