Efficiency and Technical Progress in Check Processing

by Paul W. Bauer

Paul W. Bauer is an economist at the Federal Reserve Bank of Cleveland. For their comments, suggestions, and encouragement, the author would like to thank Randall Eberts and Mark Sniderman.

Introduction

By lowering the transaction costs associated with barter, a payments system greatly facilitates the exchange of goods and services.[1] Although vastly improved over the years, the process of transferring funds remains costly, and the evolution of the payments system has been at least partially determined by efforts to trim these costs further.[2] Increasing the productivity of the payments system improves economic welfare both by releasing resources to other sectors of the economy and by lowering the effective purchase price of goods and services.

In addition to its roles as the nation's central bank and as the primary federal regulator of state member banks and bank holding companies, the Federal Reserve System is also a major provider of payment services. Ordered by the Federal Reserve Act of 1913 to ensure the efficiency of the payments system, the central bank has directly participated in the market since its inception. Initially, it provided a national mechanism for clearing and settling checks — two major components of payment services — and instituted regulations that eliminated the incentive for the circuitous routing of checks.[3]

Prior to passage of the Depository Institutions Deregulation and Monetary Control Act (MCA) of 1980, the Federal Reserve did not charge fees for its payment services and provided them only to member banks. Consequently, it faced little competition from private providers serving nonmember financial institutions. Starting in 1981, the MCA required the Federal Reserve to make its services available to all depository institutions and to charge fees that would recover its costs. The goal was to foster a more efficient payments system by giving private providers of payment services the opportunity to compete.

Given this new competitive environment, it became even more important for the Federal Reserve to be able to track the performance of its various offices. Over the years, an extensive accounting system has been developed to identify costs associated with each of its services.

■ 1 The payments system refers to such activities as the provision of currency and coin, processing and clearing of checks, providing for settlement of checks and other types of payments, and wire transfers of funds. See Board of Governors of the Federal Reserve System (1984).

■ 2 See Garbade and Silber (1979), Niehans (1971), and Brunner and Meltzer (1971).

■ 3 See Garbade and Silber (1979) and Humphrey (1980).

This has allowed unit cost performance measures (total service costs divided by service volume) to be calculated for each service offered.

This article examines the costs of providing check-processing services at 47 Federal Reserve offices (District Banks, branch offices, and regional check-processing centers) from 1983:IQ to 1990:IVQ by estimating a multiproduct cost function using an econometric frontier approach. After briefly discussing the advantages and disadvantages of the Federal Reserve's unit cost measures, I demonstrate how they can be decomposed into separate effects related to differences in cost efficiency, output mix, input prices, and environmental variables (these control for various site-specific characteristics) using estimates derived from the cost function.[4] The cost-function approach provides much more complete information about the sources of office performance than do unit cost measures, but it is more difficult and time-consuming to calculate.

In order to explore how the cost frontier may have shifted over time in response to technological and regulatory changes, the article also presents estimates of technical progress, as measured by whether the cost of producing a given level of output declines over time. This technique provides valuable insights into the technological constraints faced by the Federal Reserve.

It should be remembered, however, that research such as this is a continuing process and that a more complete understanding of the production and cost efficiencies associated with check processing will require multiple investigations. Consequently, the numerical estimates presented here must be interpreted with caution, understood in the context of stated caveats, and viewed as only a partial effort to model one aspect of the payments system.

Section I describes the central bank's provision of check-processing services and summarizes some previous studies of the payments system. Section II then discusses how the econometric frontier approach is used to estimate the multiproduct cost function, and explains how a unit cost measure of performance can be decomposed into its various components. After describing the data employed in the study, I analyze estimates of cost efficiency, scale economies, and technical change. Unit costs are decomposed for each office using the estimated multiproduct cost frontier. The final section considers the future of Federal Reserve check processing in light of new technologies, such as same-day settlement and electronic check truncation.[5]

I. Background

Description of Check Processing

Check processing is, in some ways, a fairly straightforward operation: A payor writes a check to a payee, who deposits it at his bank or other depository institution. This is all most of us ever think about, and if the payor and the payee are customers of the same bank (which occurs about 30 percent of the time), this is almost the end of the story. For these "on-us" items, the only step left is for the bank to debit the payor's account and credit the payee's account. But if both parties have accounts at different banks, then the payee's bank must forward the check to the payor's bank —a situation that occurs roughly 45 billion times a year. For these items, a bank can send checks directly to the payor's institution or route them indirectly through a local clearinghouse, a correspondent institution, or a Federal Reserve office. The Fed processes about 35 percent of these interbank checks.

In the relatively rare event that the check is returned for insufficient funds (less than 1 percent of checks), the process repeats itself, only in reverse. The return process is more labor intensive and costly. In contrast to forward volumes, the Federal Reserve handles the vast majority of payment system return items. This lack of private-sector competition suggests that the Fed's prices for handling returned checks may be too low, a subject discussed in more detail below.

Thus, the central bank provides two types of check-processing services, forward items and return items, and has a separate price schedule for each. Although the end result is the same for all checks, this description fails to reveal the myriad products offered by a typical processing center. Items can be differentiated by the location of the payor bank, the times of presentment and settlement, and the amount of presorting performed by the institution submitting the checks.

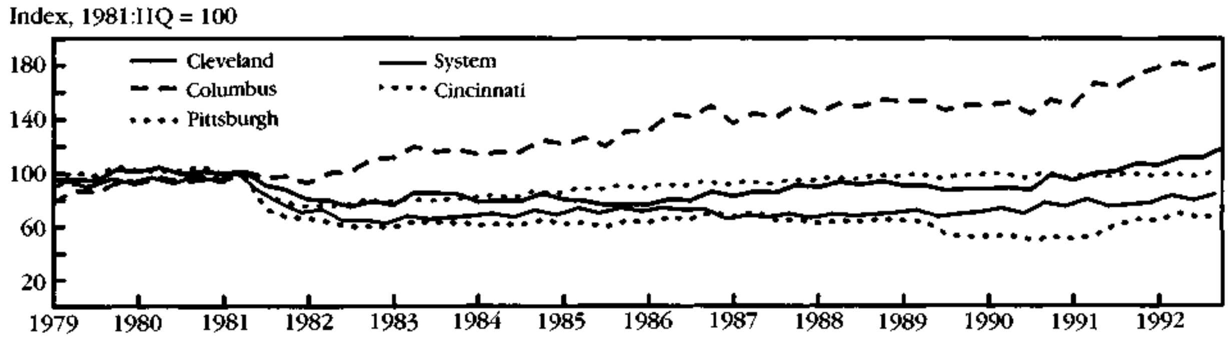
Costs can vary significantly as a result of these product characteristics. Fine-sort items, for instance, are fully presorted by the submitting

■ 4 Output mix includes the effects of scale economies, whether average cost rises or falls as output expands, and the effects of the relative production of the various outputs. Cost efficiency determines how closely firms operate to the cost frontier.

■ 5 Under same-day settlement, banks will have access to funds on the same day they are deposited, as long as the checks are presented before 8:00 a.m. Electronic check truncation refers to sending only an electronic image of the check, rather than the check itself, through the settlement process.

FIGURE 1

Check-Processing Volumes



SOURCE: Author's calculations.

institution and use only the Federal Reserve's transportation, settlement, and adjustment services, meaning that they cost very little to handle. At the other extreme, an item can be submitted without any presorting during the peak period of check processing (in Cleveland, from 10:00 p.m. to 1:00 a.m., but this varies significantly across offices), when the check reader-sorters are operated at close to maximum capacity. The incremental costs of these items are much higher.

Sorting checks and forwarding them to payor institutions (or returning them to depositing institutions) involves a variety of resources, or inputs. Transit (transportation and communication) is required to get the items to the processing site and on to their final destination once sorted. At the processing site, which must meet certain security standards, labor employs a variety of capital goods (mainly high-speed sorters and computers) to sort the checks and keep track of the settlement operation.

Monetary Control Act

While not changing the physical process of check clearing in the United States, the MCA altered the institutional environment profoundly. Federal Reserve payment services prior to passage were available at no charge, but only to member banks. The MCA required the Fed to begin charging for its services and to offer them to all depository institutions, including those that are not members of the System.[6] Based on guidelines established by the Board of Governors, prices for each payment service are designed to recover direct and indirect costs as well as a markup (known as the Private Sector Adjustment Factor [PSAF]) that imputes other costs typically incurred in the private sector. In check processing, each Federal Reserve District offers a slightly different mix of products, and District Banks have some flexibility in pricing.

Although the MCA increased the number of institutions that could employ the Federal Reserve's payment services, a large drop in volume was expected because fees were imposed on previously "free" services. When pricing was implemented, the Fed's share of interbank check processing fell from approximately 45 percent in 1981 to 38 percent in 1982. Currently, the System processes about 35 percent of all interbank items.

The drop in volume that immediately followed pricing can be seen in figure 1. In the first year, systemwide and Fourth District processing volumes plunged 15 and 18 percent, respectively. However, not all Fourth District offices experienced similar declines: In Pittsburgh, check volume dived almost 40 percent and has grown relatively little since, yet in Columbus, check volume recovered within the first year and expanded rapidly thereafter.[7]

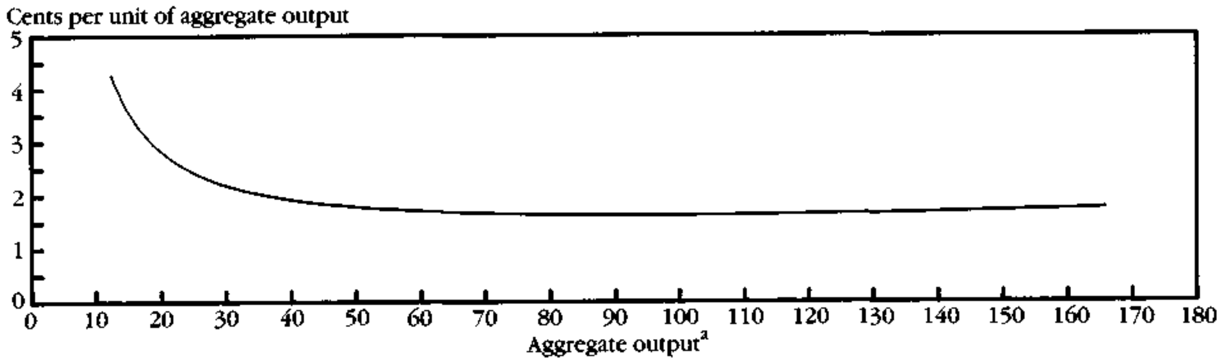
Even with this overall drop in Federal Reserve volume following the onset of pricing, the national allocation of resources improved because banks that already owned their own reader-sorters frequently found it less expensive to process more of their own checks and even to offer the service to others. Pricing boosted the efficiency of resource allocation in

6 These firms include savings and loans, credit unions, and foreign banks.

7 Unusual local economic factors accounted for much of the check volume growth in Columbus.

FIGURE 2

GLS Estimates of Ray Average Cost



a. Millions of items per quarter.
SOURCE: Author's calculations.

another way. Humphrey (1981) estimates scale economies in check processing at 36 Federal Reserve Banks and branches for the 1974–76 period. He concludes that 78 percent of checks deposited with the Fed during that time were processed at offices with significant scale diseconomies, a finding he attributes to a lack of market competition.

The increase in competition after pricing began led to greater cost-control incentives and improved resource allocation. By 1982, constant (rather than decreasing) returns to scale were the rule in Federal Reserve check-processing operations (see Humphrey [1980, 1985]).

II. Frontier Estimation and Unit Cost Decomposition

The cost function, $C(y,w,z)$, for a firm simply yields the minimum cost of producing any specified level of outputs (y) given technological constraints, input prices (w), and environmental effects (z). Foreshadowing our results somewhat, figure 2 plots the generalized least squares (GLS) estimates of the ray average cost function for check processing using the System's averages for output mix, input prices, and site-specific characteristics.[8] The curve indicates the lowest ray average cost that can be achieved for a given level of output, provided the site is operated efficiently.[9]

The concept of a frontier is quite natural in the context of a cost function. Even allowing for random events that may lead to temporarily lower or higher unit costs, we would expect most offices to operate on or above the cost function. In the context of this theoretical construct, there are many ways for things to go wrong and only one way to get them *exactly* right. Thus, observed costs will tend to be above the corresponding ray average cost curve.

The cost function is a particularly useful concept because many characteristics of the technological constraints facing the firm can be derived from it. For example, from figure 2 we can see that for low levels of output, check-processing services face scale economies—that is, ray average costs fall as both outputs are increased proportionally. For the average mix of outputs, the advantages of running a larger operation are almost exhausted after about 105 million aggregate items per quarter. Increasing the level of output from about 76 million items per quarter (the mean value from 1983 to 1990) to the level required for scale efficiency (holding the output mix constant) lowers ray average costs only 2.6 percent. Once these levels of output are reached, we will see that cost efficiency (the ratio of the cost on the frontier to observed cost) becomes a more important consideration.

■ 8 Ray average cost is defined as

$$C(\lambda y, w, z)/\lambda = C(y, w, z) / \sum_{i=1}^{M} y_i .$$

Although the denominator appears to be arbitrarily summing over the various outputs, since the output mix is held constant for this calculation, the rather arbitrary output aggregator function imposes no additional restrictions.

■ 9 Holding the mix of outputs constant by increasing them proportionally is extremely restrictive. In the results section, I demonstrate how the scale-efficient level of output depends crucially on the mix of outputs.

Given this demonstration of the usefulness of cost functions, there is one problem—these functions must be estimated from data generated by sites in operation. A number of empirical techniques have been developed to estimate frontier cost functions. Generally, they can be divided into two classes: 1) estimators based on econometric techniques, such as maximum likelihood estimation and panel data estimation, or 2) estimators based on linear programming techniques, such as data envelopment analysis.[10] In this paper, I report only estimates derived from the GLS approach.[11]

Econometric Techniques

Broadly speaking, econometric techniques employ a specific (although flexible), functional form for the cost function and impose some additional assumptions about the statistical properties of the inefficiency terms. As a category, these techniques assume a compound error term that comprises both cost inefficiencies and statistical noise. Within the category, the techniques differ in the assumptions used to decompose this error term to obtain estimates of cost efficiency.

All of the econometric techniques impose an explicit functional form for the cost function. The translog functional form is employed because it is a second-order approximation to any cost function about a point of approximation (here, the sample mean). Essentially, this means that it can model many different possible relationships among outputs, inputs, and environmental factors, depending on its parameter values. The translog cost function can be written as

$$(1)\quad \ln C_{it} = \beta_o + \sum_{m=1}^{M} \beta_m \ln y_{mit} + 1/2 \sum_{m=1}^{M}\sum_{l=1}^{M} \beta_{ml} \ln y_{mit} \ln y_{lit} + \sum_{k=1}^{K} \gamma_k \ln w_{kit} + \sum_{m=1}^{M}\sum_{k=1}^{K} \theta_{mk} \ln y_{mit} \ln w_{kit} + 1/2 \sum_{k=1}^{K}\sum_{l=1}^{K} \delta_{kl} \ln w_{kit} \ln w_{lit} + \sum_{m=1}^{L} \lambda_m \ln z_{mit} + \sum_{j=1984}^{1990} \phi_j D_{jt} + u_i + v_{it},$$

where y is a vector of M outputs, w is a vector of K input prices, z is a vector of L environmental variables, D is a set of $T-1$ dummy variables (one for every year except the first), u $(u \geq 0)$ measures cost inefficiency, and v represents statistical noise.

Estimation of this function involves finding the values of the parameters that best fit the observed data given the imposed assumptions. Equation (1) is estimated, along with the corresponding equations for input shares, imposing the usual mathematical restrictions of symmetry and linear homogeneity in input prices. The symmetry constraints come from assuming that the cost function is twice differentiable, so that

$$(2)\quad \frac{\partial^2 C}{\partial w_i \partial w_j} = \frac{\partial^2 C}{\partial w_j \partial w_i}$$

and

$$\frac{\partial^2 C}{\partial y_i \partial y_j} = \frac{\partial^2 C}{\partial y_j \partial y_i}.$$

This forces $\delta_{kl} = \delta_{lk}$ and $\beta_{kl} = \beta_{lk}$, for every k and l. Linear homogeneity in input prices, $t \cdot C(y, w) = C(y, t \cdot w)$, stems from defining the cost function as yielding the minimum cost of producing a given output level when faced with a particular set of input prices. Proportional changes in input prices affect only the cost level, not the cost-minimizing input bundle. This property imposes constraints on all parameters related to the $\ln w_{kit}$'s:

$$(3)\quad \sum_k \gamma_k = 1$$

and

$$\sum_k \theta_{mk} = \sum_k \delta_{kl} = 0,\ \forall\ l, m.$$

■ **10** A more detailed description of the techniques employed in this paper can be found in Bauer and Hancock (1993). For a thorough treatment of these two classes of techniques, see Greene (1993) and Ali and Seiford (1993).

■ **11** Bauer and Hancock (1993) report estimates using a variety of econometric and linear programming techniques. Here, I choose to concentrate on one set of results in order to provide a sharper focus.

The use of longitudinal data often allows us to avoid assuming a specific distribution for the inefficiency terms. Repeated observations over time identified site-specific, time-invariant inefficiencies.[12]

For the GLS technique, the inefficiency terms are calculated by using the average of the residuals by site, $\hat{\alpha}_i$. The most efficient site in the sample is taken to be the best estimate of where the cost frontier lies and is thus assumed to be fully efficient. The inefficiency of the i^{th} site is measured by the proportionate increase in predicted costs over the predicted costs of the most efficient site. An index bounded by zero (costs are incurred, but no output is produced) and one (a site on the cost frontier) can be calculated as

$$(4)\quad \exp(\min_j \hat{\alpha}_j - \hat{\alpha}_i).$$

The GLS technique runs an iterative, seemingly unrelated regression (ITSUR) on the system of cost and $K-1$ input share equations using panel data. One of the share equations, which are derived using Shephard's lemma, must be dropped in order to avoid singularity of the system.[13] However, since the estimates are obtained using ITSUR, the numerical estimates are the same no matter which one is dropped.

Unit Cost Decomposition

For the moment, assume that only one output is produced. In this case, unit cost is just C/y, where C is observed cost and y is observed output. If we wanted to compare one site to the average of all sites, we could do so by taking the ratio of that site's unit costs to the overall average. This would readily tell us whether a site's costs were above or below average, but we would not know why.

Using the definition of the cost function and the error specification developed for the GLS estimation technique, we can rewrite the ratio of a site's average unit costs to the overall average unit costs as follows in order to derive a more informative set of measures.[14]

$$(5)\quad \ln[(\overline{C_i/y_{1i}})/(\overline{C/y_1})] = 1/T\sum_{t=1}^{T} \ln[C(y_{it}, w_{it}, z_{it}) \exp(u_i + v_{it})/y_{1it}] - 1/TN\sum_{T=1}^{T}\sum_{i=1}^{N} \ln[C(y_{it}, w_{it}, z_{it}) \exp(u_i + v_{it})/y_{1it}].$$

Equation (5) can be rearranged to

$$(6)\quad \ln[(\overline{C_i/y_i})/(\overline{C/y})] = \{\bar{u}_i - \bar{u}\} + \left\{\left[\sum_{m=1}^{M} \beta_m(\overline{\ln y_{mi}} - \overline{\ln y_m}) + 1/2\sum_{m=1}^{M}\sum_{l=1}^{M} \beta_{ml}(\overline{\ln y_{mi}\ln y_{li}} - \overline{\ln y_m \ln y_l})\right] + \left(\overline{\ln y_{1i}} - \overline{\ln y_1}\right)\right\}$$
$$+ \left\{\sum_{k=1}^{K} \gamma_k(\overline{\ln w_{ki}} - \overline{\ln w_k}) + 1/2\sum_{k=1}^{K}\sum_{l=1}^{K} \delta_{kl}(\overline{\ln w_{ki}\ln w_{li}} - \overline{\ln w_k \ln w_l})\right\}$$
$$+ \left\{\sum_{m=1}^{M}\sum_{k=1}^{K} \theta_{mk}(\overline{\ln y_{mi}\ln w_{ki}} - \overline{\ln y_m \ln w_k})\right\} + \left\{\sum_{m=1}^{L} \lambda_m(\overline{\ln z_{mi}} - \overline{\ln z_m})\right\},$$

where the expressions in braces can be defined as effects resulting from differing cost efficiencies, outputs, input prices, the interaction of outputs and input prices, and environmental effects.[15]

12 See Schmidt and Sickles (1984) for further explanation. Berger (1993) contains some possible extensions.

13 For more details on the treatment of the share equations, see Bauer, Ferrier, and Lovell (1987).

14 Although any two observations could be chosen to compare unit costs, comparing the sample mean for the i^{th} site to the overall sample mean causes the term involving statistical noise, v, to drop out.

15 I derive the decomposition for the general case when there are M outputs and arbitrarily use the first output (forward items) as the denominator in the construction of unit costs. Empirically, the resulting measure of unit cost is highly correlated with the Federal Reserve's measure because forward processing appears to account for more than 80 percent of the costs of processing services, but has the advantage that this specification can be exactly decomposed into the various effects described below.

While these are logarithmic differences, as long as the numerical values are close to zero, they can be roughly interpreted as the percentage difference in costs stemming from these various effects.[16]

Clearly, unit costs provide a useful measure of a site's relative ability to produce a given level of output at the lowest possible cost, because it summarizes the overall effect of a variety of cost factors. Once the trouble and expense of collecting the data have been incurred, the unit cost measures are easy to calculate. On the other hand, the cost-function approach imposes greater structure and requires more effort to calculate, but it also provides a much more detailed set of information.

Now one must explicitly consider the complications posed by the presence of multiple outputs. The Federal Reserve constructs unit cost measures for each of its services and then weights them by cost shares to obtain an overall measure of performance across service lines. A potential problem is that the accounting rules employed to allocate the costs of joint inputs (those used to produce more than one service, like computer systems) may not accurately reflect the flow of services from these inputs to the various services. This will cause the calculated unit cost measures to be biased up or down, depending on whether the service in question receives more or less of its share of costs associated with the joint inputs. In the case of some joint inputs, there may be no simple accounting rule that could accurately allocate their costs because of nonlinear technological relationships among the various outputs and inputs.

Rather than relying on arbitrary accounting rules, the cost-function approach allows the data (combined with the imposed assumptions) to allocate costs to the various outputs by finding the parameters that best fit the cost model. Marginal costs for each of the outputs can then be readily calculated by differentiating the estimated cost function. For pricing and output decisions, marginal costs should be more relevant than unit costs.

III. Data Construction

Quarterly data for the 1983–90 period were collected on total costs, check volume, input prices, and environmental variables for 47 Federal Reserve check-processing sites.[17] The primary data source was annual functional cost accounting reports, which are prepared by the Federal Reserve via its Planning and Control System to monitor costs and improve resource allocation within the System. These data were supplemented by other cost and revenue figures, information from occasional Federal Reserve surveys, price index data from the Commerce Department's Bureau of Economic Analysis and the Labor Department's Bureau of Labor Statistics, and pricing data from industry sources.

Production costs for forward items, return items, and adjustments were included in total costs, but certain overhead expenses, such as special District projects, were excluded. The two measures of output were the total number of forward items and return items processed at each site. Reflecting the earlier discussion of the vast array of products offered by the various offices, this measure is at best an approximation. Some of the environmental variables discussed below attempt to adjust for the different product mixes across offices. Inputs to the check-processing function fall into the categories of buildings, materials, transit, and labor. Labor expenditures—salaries, retirement, and other benefits—accounted for 47.1 percent of total costs in 1990:IVQ.

Buildings' total cost share was only 5.6 percent in 1990:IVQ, in part because the interest expenses associated with the acquisition of buildings are not represented in the cost-accounting framework (these are included in the PSAF rather than in direct and indirect costs).

Expenditures for materials (office equipment and supplies, printing and duplicating, data processing, computers, and check reader-sorters) accounted for 29.8 percent of total costs in 1990:IVQ. Transit expenditures—the expenses associated with data and other communications, shipping, and travel—made up just over 17.5 percent.

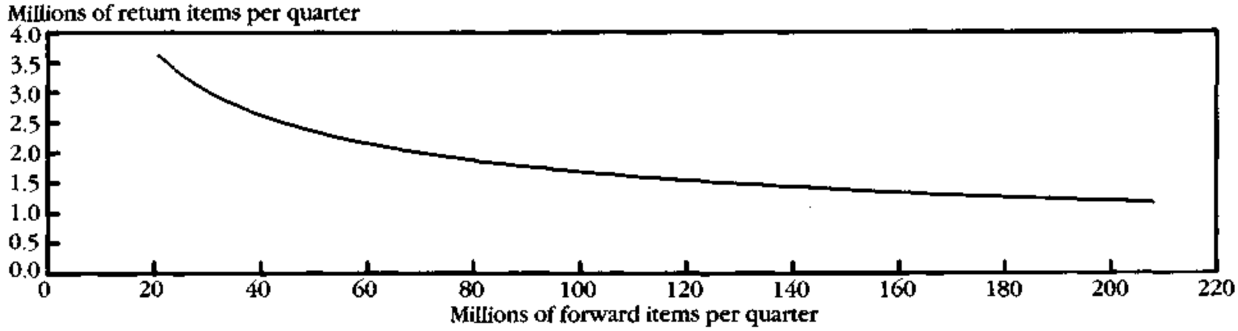
Environmental variables, which control for a variety of site-specific characteristics, include the item-pass ratio, the number of endpoints, the machine error rate, and the type of machine used. The item-pass ratio, defined as the average number of times a check must pass through a reader-sorter, is a measure of the exogenous check-sort pattern and has been found in previous studies to influence costs significantly. The number of endpoints is the number of locations

■ 16 For the exact percentage difference, one must take the antilog minus one.

■ 17 For complete details, see Bauer and Hancock (1993). The New York check-processing operation was omitted because it was closed in 1988.

FIGURE 3

M1 Locus



SOURCE: Author's calculations.

to which checks must be sorted and delivered. The machine error rate is the number of incoming errors per 100,000 checks at each office and is largely a matter of poor MICR (magnetic ink character recognition) encoding. The last environmental variable indicates whether the site used IBM or Unisys machines and allows for differences in maintenance expenses, failure rates, and downtime.

IV. Empirical Results

Scale Efficiency

A scale-efficient office operates at the output level at which ray average costs are minimized for its output mix or, equivalently, at the output level at which cost elasticity equals one (that is, a 1 percent increase in output would cause costs to rise by 1 percent).[18] Conversely, a scale-inefficient office operates at an output level larger or smaller than the scale-efficient level. If the office processes less than the scale-efficient volume, the cost elasticity is less than one (a 1 percent increase in output would raise costs by less than that amount), meaning that the office could achieve lower unit costs by boosting output. Alternatively, a scale-inefficient office that processes more than the scale-efficient volume has a cost elasticity greater than one, and unit costs can be lowered by reducing output. Thus, estimates of cost elasticities yield direct estimates of scale efficiency.

When multiple outputs are produced, the mix of outputs must also be considered when examining scale efficiency. The *M* locus (see figure 3) is defined as the set of all outputs with unitary cost elasticities.[19] For any level of forward items, the *M* locus reveals the corresponding level of return items required to achieve scale efficiency. A site operating below (above) the *M* locus experiences scale economies (diseconomies). The estimated *M* locus indicates that a site processing a large number of return items relative to forward items reaches scale efficiency at a lower level of forward items.

It may not be possible for every office to achieve scale efficiency, despite the best efforts of managers. The volume of checks and return items processed at an office depends on the size of the market and the prices charged. The economic size of managers' payments markets is outside their control, and although managers may have some authority over prices, their need to recover costs may prevent them from setting a price low enough to attract a scale-efficient volume of output. In short, even the best-run office will be scale inefficient if it is in a market too small to achieve scale efficiency.

Figure 2 demonstrated that the ray average cost curve for check and return processing was U-shaped (meaning that at low levels of aggregate output, ray average cost falls as outputs are increased proportionally, but that scale econo-

■ **18** Cost elasticity is defined as $\partial \ln C(\lambda y, w, z) / \partial \ln \lambda |_{\lambda=1}$. This turns out to be identical to the sum of the cost elasticities with respect to each output.

■ **19** The *M* locus in figure 3 is drawn with the input prices equal to their values at the sample mean. Unfortunately, the estimated *M* locus becomes increasingly speculative as it moves away from the output ratio found at the sample mean.

mies are exhausted at some point, so further increases result in higher ray average costs). In table 1, we present estimates of cost elasticities using site-specific characteristics for 1990:IVQ. Most sites are fairly close to achieving scale efficiency—given their individual output mixes, input prices, and environmental variables. Even so, these estimates suggest that the average office could lower its costs about 12 percent if it could generate scale-efficient volumes. Full scale efficiency would require the average office to increase its scale of operations significantly. However, as revealed by the ray average cost function in figure 2, most of the gains occur before 76 million items per quarter are processed.

Although some smaller offices appear to be operating in the output range where further scale economies could be exploited in the future, additional float costs that are not incorporated in our model may make this infeasible. As items from more-distant banks are processed, additional shipping costs and delays will be incurred that may outweigh the associated cost savings.

Estimates of marginal cost, or the incremental cost of processing one more item, can provide additional information for pricing. One of the beneficial outcomes of competitive markets is that competition forces prices to be set equal to marginal costs. In other words, the price that consumers pay for a good or service equals a firm's incremental cost of producing it. If the Federal Reserve set its actual prices for forward and return items in 1990:IVQ to equal estimated marginal costs, those prices would have averaged \$0.009 and \$0.643 per item, respectively (see table 1). In practice, prices are based on accounting data, and the Federal Reserve's calculated unit costs for forward and return items averaged \$0.0135 and \$0.159, respectively.

Neither unit costs nor marginal costs can be directly used for pricing because they fail to account for several characteristics (such as the time the checks are submitted for processing), yet to the extent that pricing is based on unit costs, the estimated marginal costs imply that forward items could be regarded as overpriced, whereas return items could be regarded as underpriced. Even though the Federal Reserve sets prices to recover costs, econometric estimates indicate that the accounting data appear to assign too much of the costs to forward items and too little to return items. Market conditions are consistent with econometric estimates, since there are no entry barriers into either market, yet the Federal Reserve faces little competition for return items, and there are many private-sector competitors for forward items. Clearly, this is an issue that requires further study.

Cost Efficiency

Cost inefficiencies appear to raise costs more than does scale inefficiency. If all offices could be operated on the cost frontier, costs could be lowered by about 23.5 percent. Table 2 compares GLS estimates of the cost efficiencies calculated using the multiproduct cost function employed here with the single-product estimates reported in Bauer and Hancock (1993). Overall, the results change remarkably little when return items are treated as a separate output: On average, estimated cost efficiency rises only 3.5 percent. However, one site with a relatively large number of return items (FR27) saw its estimated efficiency increase by 16.3 percentage points.

Many of the top-ranked offices were in the same Federal Reserve District, indicating that management differences may be important. Aside from superior managerial skill, estimated cost efficiency could vary across sites because some Districts may focus on cost performance while others may stress customer service, which is largely uncontrolled for in this study. For example, one District may specify precisely how checks must be submitted and refuse to accept them otherwise, while another may accept checks in any form but charge higher fees for packages that require more attention. The former District will appear to be more efficient than the latter, other things being equal, because it receives the checks exactly as it wants them. However, the latter District receives a higher fee by providing a service desired by its customers.

Unit Cost Decomposition

The average unit cost measure for each of the 47 offices over the 1983–90 period relative to the overall sample mean is presented in table 3, along with estimates of each of the component effects. Unit costs vary substantially across offices, from −0.388 to 0.309, or from about a third below to a third above the overall average. The largest single component appears to be the cost-efficiency effect, with a correlation between it and unit cost of more than 80 percent.

TABLE 1

Estimates of Marginal Costs and Cost Elasticities, 1990:IVQ

	Cost Elasticities			Marginal Costs	
Office	Forward Items	Return Items	Overall	Forward Items	Return Items
FR1	0.381	0.568	0.949	0.007	0.738
FR2	0.400	0.568	0.968	0.010	0.797
FR3	0.416	0.439	0.855	0.007	0.655
FR4	0.392	0.565	0.958	0.011	1.011
FR5	0.464	0.333	0.797	0.010	0.618
FR6	0.483	0.240	0.723	0.009	0.520
FR7	0.405	0.501	0.906	0.007	0.657
FR8	0.359	0.670	1.029	0.009	0.865
FR9	0.444	0.436	0.880	0.009	0.584
FR10	0.418	0.443	0.860	0.009	0.817
FR11	0.469	0.424	0.893	0.012	0.525
FR12	0.450	0.388	0.838	0.007	0.486
FR13	0.391	0.590	0.982	0.010	0.907
FR14	0.382	0.629	1.011	0.011	0.902
FR15	0.412	0.506	0.918	0.009	0.700
FR16	0.429	0.421	0.850	0.008	0.631
FR17	0.396	0.526	0.922	0.008	0.767
FR18	0.533	0.219	0.752	0.015	0.433
FR19	0.489	0.162	0.651	0.010	0.561
FR20	0.439	0.556	0.996	0.011	0.466
FR21	0.436	0.495	0.932	0.009	0.473
FR22	0.381	0.548	0.929	0.006	0.662
FR23	0.407	0.519	0.926	0.012	1.077
FR24	0.439	0.500	0.939	0.012	0.673
FR25	0.509	0.198	0.706	0.006	0.296
FR26	0.492	0.303	0.795	0.008	0.331
FR27	0.375	0.708	1.083	0.011	0.653
FR28	0.493	0.402	0.896	0.011	0.346
FR29	0.502	0.362	0.864	0.013	0.389
FR30	0.393	0.519	0.912	0.005	0.531
FR31	0.452	0.130	0.582	0.006	0.601
FR32	0.371	0.528	0.899	0.009	1.258
FR33	0.427	0.400	0.827	0.006	0.570
FR34	0.413	0.470	0.882	0.006	0.531
FR35	0.453	0.430	0.883	0.010	0.547
FR36	0.473	0.341	0.814	0.010	0.534
FR37	0.366	0.604	0.971	0.010	1.131
FR38	0.485	0.382	0.867	0.013	0.496
FR39	0.473	0.312	0.784	0.008	0.480
FR40	0.415	0.463	0.878	0.009	0.805
FR41	0.480	0.378	0.859	0.010	0.457
FR42	0.448	0.467	0.915	0.009	0.481
FR43	0.403	0.576	0.979	0.009	0.631
FR44	0.442	0.445	0.887	0.007	0.475
FR45	0.450	0.491	0.941	0.013	0.601
FR46	0.421	0.378	0.799	0.006	0.671
FR47	0.414	0.447	0.861	0.009	0.895
Average	0.433	0.446	0.880	0.009	0.643

SOURCE: Author's calculations.

The output and input price effects can also exert a significant influence on some offices' unit costs, but the correlations with unit costs are much lower. In fact, the correlation between unit costs and the input price effect is negative, hinting that some input quality may vary across sites and that higher-priced inputs may be more productive. By construction, input prices and estimates of cost efficiency are uncorrelated, so in this case, the unit cost measures have revealed an issue that requires further study.

The environmental effects tend to be minimal across all sites except FR25, which serves an unusually small number of endpoints. The interactive effect is slight for all offices, with the largest estimated effect shifting relative unit costs only about 6.8 percent.

Productivity Growth

Including year dummies in the cost function allows us to estimate whether it shifts down (or up) over time as a result of changes in technology or in the regulatory environment. Estimates of a technical change index are presented in figure 4. For 1983, the index equals 100; for later years, it rises or falls depending on the behavior of the estimated cost function. As of 1990, costs had risen about 8.7 percent. Most of the upward shift that occurred in 1989 appears to have stemmed from transitory costs related to the implementation of regulations designed to post checks more quickly, since costs fell sharply in 1990.

Measured productivity growth in check processing has been anemic for two main reasons: 1) some of the cost savings have been plowed back into producing a higher-quality product (such as expedited funds availability), and 2) even though prices for computer equipment and other office machinery have fallen precipitously over the last 10 years, the price of high-speed check reader-sorters has remained roughly unchanged in real terms. Apparently, the limit of how quickly paper checks can be read and sorted has nearly been reached, and further advances will have to await the increased use of electronics in collecting checks.

TABLE 2

GLS Cost Efficiency Estimates, 1983–90 Average

Office	Single Product GLS	Single Product Rank	Multiproduct GLS	Multiproduct Rank	Change in Efficiency	Change in Rank
FR1	0.864	7	0.884	8	0.020	1
FR2	0.535	46	0.604	45	0.069	−1
FR3	0.997	2	0.998	2	0.002	0
FR4	0.587	42	0.615	44	0.028	2
FR5	0.625	37	0.708	25	0.083	−12
FR6	0.669	25	0.687	29	0.018	4
FR7	0.634	31	0.689	28	0.056	−3
FR8	0.633	32	0.696	26	0.063	−6
FR9	0.656	28	0.673	33	0.017	5
FR10	0.629	35	0.632	43	0.003	8
FR11	0.581	43	0.668	34	0.086	−9
FR12	0.765	11	0.861	9	0.095	−2
FR13	0.717	16	0.736	19	0.019	3
FR14	0.713	19	0.718	24	0.005	5
FR15	0.715	17	0.737	18	0.022	1
FR16	0.714	18	0.721	22	0.007	4
FR17	0.754	12	0.728	20	−0.026	8
FR18	0.647	29	0.645	39	−0.002	10
FR19	0.645	30	0.679	32	0.034	2
FR20	0.707	21	0.770	15	0.063	−6
FR21	0.683	24	0.694	27	0.011	3
FR22	0.919	5	0.928	5	0.009	0
FR23	0.530	47	0.585	46	0.055	−1
FR24	0.612	39	0.661	36	0.048	−3
FR25	0.802	9	0.837	10	0.035	1
FR26	0.880	6	0.927	6	0.047	0
FR27	0.627	36	0.790	14	0.163	−22
FR28	0.738	15	0.798	13	0.060	−2
FR29	0.615	38	0.665	35	0.050	−3
FR30	0.969	3	0.961	3	−0.008	0
FR31	0.693	22	0.720	23	0.027	1
FR32	0.630	34	0.636	42	0.006	8
FR33	0.939	4	0.939	4	0.000	0
FR34	1.000	1	1.000	1	0.000	0
FR35	0.711	20	0.756	16	0.045	−4
FR36	0.660	27	0.685	31	0.025	4
FR37	0.574	44	0.656	38	0.082	−6
FR38	0.610	40	0.644	40	0.034	0
FR39	0.792	10	0.815	12	0.023	2
FR40	0.557	45	0.567	47	0.011	2
FR41	0.747	13	0.820	11	0.074	−2
FR42	0.742	14	0.721	21	−0.021	7
FR43	0.668	26	0.745	17	0.078	−9
FR44	0.843	8	0.914	7	0.071	−1
FR45	0.610	41	0.659	37	0.049	−4
FR46	0.630	33	0.685	30	0.055	−3
FR47	0.689	23	0.638	41	−0.051	18
Average	0.708		0.742		0.035	

SOURCE: Author's calculations.

TABLE 3

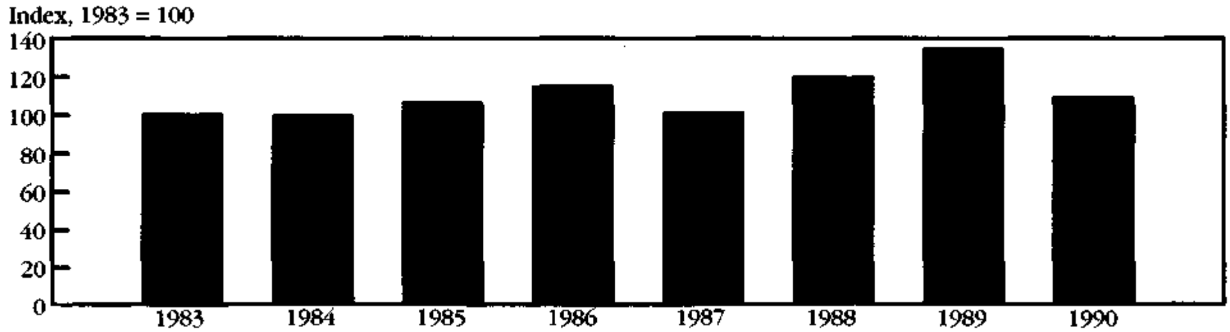
Unit Cost Decomposition[a]

Office	Unit Cost ($)	Logarithmic Differences from Sample Means, 1983–90					
		Unit Cost	Cost Efficiency	Total Output	Direct Input Price	Interactive Effect	Environmental Effect
FR1	0.015	−0.179	−0.185	−0.054	0.017	−0.006	0.048
FR2	0.021	0.176	0.195	0.001	0.061	0.003	−0.084
FR3	0.013	−0.349	−0.307	−0.110	0.038	0.005	0.024
FR4	0.023	0.242	0.178	−0.163	0.225	0.008	−0.005
FR5	0.018	0.022	0.037	−0.136	0.213	−0.022	−0.070
FR6	0.018	0.028	0.067	0.010	0.012	−0.006	−0.054
FR7	0.016	−0.087	0.064	−0.133	0.068	0.000	−0.086
FR8	0.022	0.212	0.054	−0.122	0.211	0.019	0.050
FR9	0.019	0.047	0.087	0.018	−0.046	0.006	−0.018
FR10	0.020	0.133	0.151	−0.119	0.086	0.002	0.013
FR11	0.020	0.131	0.096	0.174	−0.044	0.004	−0.099
FR12	0.014	−0.221	−0.159	−0.038	0.083	−0.002	−0.106
FR13	0.018	0.023	−0.002	−0.148	0.223	0.001	−0.051
FR14	0.018	0.029	0.022	0.053	−0.167	−0.023	0.143
FR15	0.017	−0.036	−0.003	−0.125	−0.029	0.001	0.119
FR16	0.015	−0.145	0.019	−0.085	−0.084	0.000	0.005
FR17	0.018	−0.017	0.009	−0.155	0.073	0.000	0.055
FR18	0.024	0.304	0.130	0.278	−0.159	0.030	0.026
FR19	0.021	0.182	0.078	0.054	0.076	−0.029	0.002
FR20	0.018	−0.007	−0.047	0.149	−0.183	0.004	0.070
FR21	0.016	−0.112	0.057	0.005	−0.156	0.002	−0.020
FR22	0.012	−0.376	−0.234	−0.191	0.078	0.002	−0.031
FR23	0.022	0.193	0.227	−0.081	0.113	−0.001	−0.066
FR24	0.022	0.219	0.106	0.188	−0.167	−0.002	0.094
FR25	0.014	−0.217	−0.130	0.134	0.404	−0.068	−0.556
FR26	0.016	−0.124	−0.232	0.149	−0.136	0.020	0.075
FR27	0.020	0.120	−0.073	0.357	−0.233	−0.047	0.117
FR28	0.020	0.134	−0.083	0.356	−0.270	0.020	0.111
FR29	0.024	0.309	0.100	0.375	−0.275	0.028	0.080
FR30	0.012	−0.388	−0.269	−0.193	0.065	0.002	0.008
FR31	0.019	0.042	0.019	−0.027	0.024	−0.003	0.029
FR32	0.020	0.122	0.145	−0.288	0.219	0.012	0.034
FR33	0.014	−0.269	−0.245	−0.165	0.101	0.003	0.038
FR34	0.013	−0.352	−0.308	−0.179	0.124	0.001	0.010
FR35	0.018	0.025	−0.029	0.117	−0.189	0.008	0.119
FR36	0.021	0.155	0.070	0.105	−0.151	0.016	0.115
FR37	0.021	0.162	0.113	−0.093	0.034	0.001	0.107
FR38	0.021	0.141	0.132	0.029	0.076	−0.004	−0.092
FR39	0.016	−0.086	−0.103	0.017	−0.004	0.004	0.001
FR40	0.019	0.038	0.258	−0.126	0.031	0.001	−0.126
FR41	0.017	−0.051	−0.111	0.102	−0.058	0.008	0.008
FR42	0.017	−0.036	0.018	0.187	−0.338	0.019	0.077
FR43	0.019	0.046	−0.014	0.152	−0.150	−0.018	0.076
FR44	0.013	−0.288	−0.218	0.001	−0.005	0.003	−0.069
FR45	0.023	0.259	0.109	0.133	−0.130	0.002	0.144
FR46	0.016	−0.104	0.069	−0.252	0.217	−0.007	−0.132
FR47	0.017	−0.049	0.142	−0.162	0.101	0.003	−0.133

a. Office mean relative to overall sample mean.
SOURCE: Author's calculations.

FIGURE 4

Technical Change Index



SOURCE: Author's calculations.

V. Conclusions and Prospects for the Future

This study finds scale economies to be sufficiently large to enable most offices to proportionally increase their forward and return volumes, yet still lower their ray average costs by roughly 12 percent. Costs appear to increase much more rapidly as more return items are processed than as more forward items are processed. Although there appear to be opportunities for most offices to improve their performance by further exploiting scale economies, costs could be lowered even more (up to 23.5 percent overall) if all offices could operate closer to the cost frontier.

It is necessary to keep in mind three important caveats. First, some offices may be located in areas where it may not be possible to expand output enough to achieve scale efficiency. Second, the cost-efficiency measure is relative to the most efficient office observed in the sample. Third, although I use the concise term "cost efficiency," the concept is more fully described as "once factors included in the cost function are controlled for, there remain unexplained cost differences across processing sites." Every effort has been made to control for the factors that affect the costs of check-processing offices, but no one can hope to account for every factor that *might* significantly affect costs. Future research will extend the analysis by trying to control for product quality in a more detailed way.[20]

The multiproduct cost-efficiency estimates for the 47 offices covered here are highly correlated with earlier single-product estimates presented in Bauer and Hancock (1993). On average, cost efficiency rose only 3.8 percentage points when returns were treated as a separate output. However, one office that processed an atypically high level of returns had its cost-efficiency index increase by 16.3 percentage points. The overall level of cost efficiency is roughly the same as that found for private financial institutions, using similar estimation techniques.[21]

In the single-product setting, unit cost measures provide an easily calculated overall indicator of relative office performance. Unfortunately, they do not reveal the sources of superior or inferior performance. In the multiproduct setting, unit cost measures could be biased if the costs of joint inputs are misallocated across services. The cost-function approach overcomes both of these drawbacks, but requires imposing a number of potentially restrictive assumptions. The decomposition of unit costs reveals that, for this sample, cost efficiency tends to be the largest single component, but considerable office-specific variation results from the other components. Only the interaction effect between output levels and input prices is consistently small in magnitude for all offices.

■ **20** Product quality can affect cost efficiency measures because it is expensive to provide higher quality. If output is not adjusted for product quality, sites providing lower-quality output will, other things being equal, appear to be more cost efficient.

■ **21** For example, see Bauer, Berger, and Humphrey (1993), Ferrier and Lovell (1990), Fried, Lovell, and Vanden Eeckaut (1993), and Mester (1993), to name just a few. While these studies examine the cost efficiency of producing outputs other than check-processing services, their estimated efficiency levels suggest that the Federal Reserve pursues its behavioral goal about as well as private financial institutions pursue theirs.

The net effect of technological and regulatory changes seems to have shifted the multiproduct cost frontier up slightly over time, a finding that supports the prevailing view that much greater use of new technologies, such as check truncation and imaging, will be required to achieve significant technical change in check processing. This finding is also consistent with earlier work by Bauer and Hancock (1993).

In the coming years, check processing at the Federal Reserve will face a number of new challenges, since volume is likely to rise less rapidly, and may even fall. One cause is mergers and acquisitions in the financial service sector, which have resulted in more on-us items that can be cleared internally. Other causes include bilateral agreements among banks to swap checks directly, the emergence of private nationwide check processors, same-day settlement, and technological advances such as electronic check presentment and the shift to electronic payments.

The introduction of pricing, the evolution of technology, and the consolidation of the banking industry during the past few years have led to many changes in the check-processing market. Moreover, increased competition between bankers and nonbank providers of financial services, along with more competition between checks and other payment media, indicates that more changes will follow. In the future, market forces will largely determine the number and location of check-processing sites across the country. Research studies can contribute to a more complete understanding of developments in this dynamic payment service.

References

Ali, A., and L. Seiford. "The Mathematical Programming Approach to Efficiency Analysis," in H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, eds., *The Measurement of Productive Efficiency: Techniques and Applications.* Oxford: Oxford University Press, 1993.

Bauer, P. W., A. N. Berger, and D. B. Humphrey. "Efficiency and Productivity Growth in U.S. Banking," in H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, eds., *The Measurement of Productive Efficiency: Techniques and Applications.* Oxford: Oxford University Press, 1993.

Bauer, P. W., G. Ferrier, and C. A. K. Lovell. "A Technique for Estimating a Cost System That Allows for Inefficiency," Federal Reserve Bank of Cleveland, Working Paper 8704, May 1987.

Bauer, P. W., and D. Hancock. "The Efficiency of the Federal Reserve in Providing Check Processing Services," *Journal of Banking and Finance,* vol. 17 (April 1993), pp. 287–311.

Berger, A. N. "Distribution-Free Estimates of Efficiency in the U.S. Banking Industry and Tests of the Standard Distributional Assumptions," *Journal of Productivity Analysis,* vol. 4, no. 3 (September 1993), pp. 261–92.

Board of Governors of the Federal Reserve System. *The Federal Reserve System: Purposes and Functions.* Washington, D.C.: Board of Governors, 1984.

Brunner, Karl, and Allan H. Meltzer. "The Uses of Money: Money in the Theory of an Exchange Economy," *American Ecoomic Review,* vol. 61, no. 5 (December 1971), pp. 784–805.

Ferrier, Gary D., and C. A. K. Lovell. "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence," *Journal of Econometrics,* vol. 46, nos. 1/2 (October/November 1990), pp. 229–45.

Fried, H. O., C. A. K. Lovell, and P. Vanden Eeckaut. "Evaluating the Performance of U.S. Credit Unions," *Journal of Banking and Finance,* vol. 17 (April 1993), pp. 251–65.

Garbade, Kenneth D., and William L. Silber. "The Payment System and Domestic Exchange Rates: Technological versus Institutional Change," *Journal of Monetary Economics*, vol. 5, no. 1 (January 1979), pp. 1–22.

Greene, W. H. "The Econometric Approach to Efficiency Analysis," in H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, eds., *The Measurement of Productive Efficiency: Techniques and Applications.* Oxford: Oxford University Press, 1993.

Humphrey, David B. "Costs, Scale Economies, Competition, and Product Mix in the U.S. Payments Mechanism," Board of Governors of the Federal Reserve System, Research Papers in Banking and Financial Economics, No. 37, May 1980.

________. "Economies to Scale in Federal Reserve Check Processing Operations," *Journal of Econometrics*, vol. 15, no. 1 (January 1981), pp. 155–73.

________. "Resource Use in Federal Reserve Check and ACH Operations after Pricing," *Journal of Bank Research*, vol. 16, no. 1 (Spring 1985), pp. 45–53.

Mester, L. J. "Efficiency in the Savings and Loan Industry," *Journal of Banking and Finance*, vol. 17 (April 1993), pp. 267–86.

Niehans, Jürg. "Money and Barter in General Equilibrium with Transaction Costs," *American Economic Review*, vol. 61, no. 5 (December 1971), pp. 773–83.

Schmidt, P., and R. Sickles. "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics*, vol. 2 (1984), pp. 367–74.