Post Consolidation Estimates of ACH Scale Economies, Technical Change and Cost Efficiency

Paul W. Bauer, Economic Advisor
Federal Reserve Bank of Cleveland

and

Patrick Higgins, Research Assistant
Federal Reserve Bank of Cleveland

Abstract

Federal Reserve banks have been processing Automated Clearing House (ACH) payments, a consumer oriented product intended to be an electronic substitute for paper checks, since the mid-1970s. Previous studies by Humphrey (1982, 1984, 1985), Bauer and Hancock (1995), and Bauer and Ferrier (1996) have found strong evidence for increasing returns to scale (falling unit costs as volume transacted increases). Employing data from the early 1990s, the latter two studies also found evidence of rapid technical change (falling unit costs as better technology is implemented over time). Using data through 2000, we find that the ACH cost function still exhibits both of these properties, yet there are some signs that the technology may be maturing. The rate of technical change, while still rapid, appears to have slowed. Scale economies, at least as estimated by the GLS technique, still appears to exist far beyond the current range of output. Both of these findings have important implications for Reserve Banks that are facing increasing competition from private sector providers.

Introduction

Federal Reserve banks have been processing Automated Clearing House (ACH) payments, a consumer oriented product intended to be an electronic substitute for paper checks, since the mid-1970s. Previous studies by Humphrey (1982, 1984, 1985), Bauer and Hancock (1995), and Bauer and Ferrier (1996) have found strong evidence for increasing returns to scale (falling unit costs as volume transacted increases). The latter two studies also found evidence of rapid technical change (falling unit costs as better technology is implemented over time).

Our goal in this paper is to determine whether these two properties continue to hold. Specifically, we examine how the 1996 switch to FedACH, a centralized ACH application software used to process ACH payments, altered ACH's costs. Understanding the cost implications of that process suggests how costs might be impacted by the consolidation in 2001 of operations support to two Federal Reserve sites and could aid in planning further consolidation efforts. Also, our estimates of marginal costs should be of use in setting prices for this service and for predicting the long run market structure for this service as Reserve Banks face increased competition from private sector providers. Finally, our results are a rare chance to closely study productivity change in a service industry. Measuring output for some service industries is difficult, making estimates of productivity gains a challenge. Because ACH transactions are a natural measure of output and the Federal Reserve keeps detailed cost data, we can shed some light on this important sector of the economy.

We find that although the ACH cost function still exhibits both scale economies and rapid technical change, there are some signs that the technology may be maturing. Although scale economies appear to exist far beyond the observed range of output, the rate of technical change appears to have slowed since 1998. Both of these findings have important implications for

Reserve Banks that are facing increasing competition from private sector providers.

In the next section, we describe what an ACH transaction is and the market structure for this service. We then describe our empirical model and the data we employ to obtain our cost function estimates. After analyzing our results, we consider some of the implications of these findings for the Federal Reserve and ACH transactions.

The ACH System

The ACH system is a value-dated electronic funds transfer system. The principal participants in an ACH transaction are the customer initiating the transaction (typically, consumers, corporations, or government entities), that customer's bank, the bank of the party who is paying or receiving funds, that party, and the ACH processor. Either credit transfers or debit transfers may be made using an ACH system. With credit transfers, such as payroll, the payor's bank initiates the transfer. The funds received by the payee's bank are generally provisional until the morning of the business day following the settlement day, which is one or two days after the ACH payment is processed. Federal Reserve banks can revoke the payments if the sending bank does not have sufficient funds in its account to fund them on the settlement day. With debit transfers, such as mortgage payments, the payee's depository institution initiates the transfer. A relatively new use of ACH has been the processing of one-time payments, such as the checks that merchants truncate at the point-of-sale or, occasionally, as "e-checks" for internet transactions.

ACH transactions offer several key advantages over paper instruments. First, the payor knows exactly when the funds will be removed from his account, and likewise the payee knows exactly when the funds will be deposited to his account. Second, ACH transactions offer

convenience, particularly for recurring payments, because a paper check need not be written, delivered, and cashed. Third, the total costs to all parties are much cheaper for ACH transactions than for paper checks.[1] Fourth, there can be accounting efficiencies for payors and payees that track invoices electronically if ACH transfers are used instead of paper checks.[2]

In 2000, the last year of our sample, the ACH system processed 6.9 billion transactions with a total value of $20.3 trillion. Figure 1 plots the growth in ACH volume from 1990 to 2000, a period when volume growth averaged 13.4 percent per year. Over this period, unit cost declined from about $0.04 to $0.01, an average of 12.3 percent per year (see Figure 2).

For many years, the Federal Reserve had about an 80 percent share of the ACH market, but those days are gone. Competition from private sector operators is expected to provide much more intense competition for ACH transactions in the coming years. Given the increasingly competitive ACH market, the Federal Reserve has tried to become more responsive to market pressures. Currently the largest private sector operator is the Electronic Payments Network (EPN) with about 15 percent of the inter-bank ACH market.[3]

Pricing to meet the competition is critically important for the ACH service because about

---

[1]Using data from 1987, Humphrey and Berger (1990) estimated that the average "social cost" of an ACH item including payor, payee, and bank costs was $0.29, versus $0.79 for a check. More recently, Wells (1994) using a similar methodology but incorporating additional cost information and data from 1993 estimated that the social cost of payments by check ranged between $2.81 and $3.12 whereas social costs for payments by ACH ranged between $1.18-$1.50. These later estimates suggest that the difference in social costs between checks and ACH have grown over time. As electronic data interchange (EDI) gains adherents, not only are paper checks replaced with less costly ACH transactions, there are also cost savings for both the payor and the payee by allowing invoices to be tracked electronically.

[2]See Knudsen, Walton, and Young (1994) for a discussion of the potential benefits of financial electronic data interchange, the combination of electronic remittance data and electronic funds transfers, for business payments.

[3] See "Cheap and Efficient," U.S. Banker, November 2001 (http://www.us-banker.com/usb/articles/usbnov01-01c.shtml).

80 percent of FedACH origination volumes come from the 50 largest originating institutions. Given this high concentration of originations the loss of just a few of them would result in a significant loss in volume. The presence of scale economies would then lead to higher unit costs, making it harder to recover the full economic costs of providing the service without increasing fees. Consequently, volume loss could create a very un-virtuous circle. Receipt volumes are less at risk because they are more evenly spread among various endpoints. The largest 50 institutions account for only 27 percent of receipts, and it takes the largest 300 FedACH customers to get to 50 percent of total receipts. In such an environment, obtaining accurate estimates of scale economies and technical change is critical to correctly price the product, directly enhancing economic efficiency.

Empirical Model

In order to estimate the magnitude of scale economies, technological change for ACH processing we fit a cost function. A cost function approach is selected because cost minimization is a better assumption for Reserve Banks' behavioral objective than profit maximization. Under the MCA of 1980, the Reserve Banks are required to recover their costs of providing financial services that also have private sector competition. The costs recovered must include not only the actual production costs, but also a set of imputed costs called the "private sector adjustment factor" (PSAF) that includes financing costs, return on capital, taxes, and certain other expenses that would be incurred if a private business firm provided the service. Although the Reserve Banks must price their services to recover actual costs plus PSAF, our cost functions only model actual activity production costs.

A cost function maps the least cost method of processing each level of transfers when

inputs, such as labor and equipment, can vary freely. The cost function is useful because characteristics of the technological constraints facing the firm can be derived from it, such as estimates of scale economies and technological change, as will be explained more fully below. The effects of lower input prices can be obtained directly.

We employ a translog function as it provides a good local approximation to any arbitrary twice differentiable and linearly homogeneous cost function. Because this approximation becomes less precise the further one moves away from the center of mass of the data, we also estimate a hybrid-tranlog cost function, which has additional Fourier transform terms to improve its fit over a wider range of output. Even with these extra terms, there is no avoiding the fact that the fit for the smallest and largest firms will be less reliable, and estimates beyond the range of observed data will be less precise because there are fewer observations in these regions. Nonetheless, in this application we will see that the fit of these two models is very similar.

Our general translog cost function can be written as,

$$\begin{aligned}\ln C_{it} = {} & \beta_0 + \beta_y \ln y_{it} + 1/2\, \beta_{yy} (\ln y_{it})^2 \\ & + \sum_{k=1}^{K} \gamma_k \ln w_{kit} + 1/2 \sum_{k=1}^{K} \sum_{j=1}^{K} \delta_{kj} \ln w_{kit} \ln w_{jit} \\ & + \sum_{k=1}^{K} \delta_{yk} \ln y_{it} \ln w_{kit} \\ & + \sum_{j=T_0+1}^{T} \phi_j YR_j + u_i + v_{it},\end{aligned} \tag{1}$$

where y_{it} is the number of ACH items processed at site i in period t, w_{it} is a vector of K input prices for site i in period t, YR_j $(j=1991,\ldots 2000)$ is a set of T-1 yearly indicator variables (one for every year except the first), u_i is the cost inefficiency of the i-th site, and v_{it} term allows for statistical noise. Hybrid-translog specification is similar except that it includes a set of Fourier expansion terms to improve fit away from the sample mean,

$$\begin{aligned}\ln C_{it} = {} & \beta_0 + \beta_y \ln y_{it} + 1/2\, \beta_{yy} (\ln y_{it})^2 \\ & + \sum_{k=1}^{K} \gamma_k \ln w_{kit} + 1/2 \sum_{k=1}^{K}\sum_{j=1}^{K} \delta_{kj} \ln w_{kit} \ln w_{jit} \\ & + \psi_1 \cos(z_{it}) + \psi_2 \cos(2z_{it}) + \psi_3 \cos(3z_{it}) \\ & + \omega_1 \sin(z_{it}) + \omega_2 \sin(2z_{it}) + \omega_3 \sin(3z_{it}) \\ & + \sum_{k=1}^{K} \delta_{yk} \ln y_{it} \ln w_{kit} \\ & + \sum_{j=T_0+1}^{T} \phi_j YR_j + u_i + v_{it},\end{aligned} \tag{2}$$

where the z_{it} is output normalized over the range $(0, 2\pi)$.

To increase statistical efficiency, we estimate the cost function jointly with input share equations that are derived using Shephard's Lemma and the relevant cost function.[4] Estimation of both the cost function and the input share equations provides additional degrees of freedom and statistical precision. The resulting system of cost and share equations is estimated using the iterative seemingly unrelated regression technique.

[4]See Diewert (1982) for a discussion of Shephard's Lemma.

We impose the usual symmetry and linear homogeneity in input prices required by economic theory. Symmetry restrictions follow from assuming that the cost function is twice differentiable in input prices. This forces $\delta_{kj} = \delta_{jk}$ for every *k* and *j*. Linear homogeneity in input prices means that proportional changes in input prices affects the level of cost equi-proportionally but not the cost-minimizing mix of inputs.[5] Linear homogeneity restrictions is the result of defining the cost function as the minimum cost of producing a given output level when faced with a given set of input prices.

We use two alternative econometric assumptions to measure facility-level cost efficiencies because empirical measurements of cost efficiency in the financial services industry often vary significantly depending on the methodology employed. By estimating the model two ways, we can explore how sensitive our results are to alternative assumptions.[6] Both the generalized least squares (GLS) and the Within estimation techniques use the longitudinal aspect of the data and avoid assuming a specific distribution for the inefficiency term.[7] Basically, these techniques use repeated observations over time to identify firm-specific, time-invariant inefficiencies, u_i. Iterative seemingly unrelated regression techniques are used to estimate the system of cost and input share equations using longitudinal data. With the GLS technique, the inefficiency terms are calculated by using the average of the residuals for each facility *i*, $\hat{u}_i$. The most efficient currency handling facility in the sample is assumed to be fully efficient, and the inefficiency of every other facility *i* is measured by the proportionate increase in predicted costs

[5]Mathematically, linear homogeneity requires that $\lambda C(y, w) = C(y, \lambda w)$ for all $\lambda > 0$.

[6]See, for example, Ferrier and Lovell (1990), Bauer et al. (1993), and Berger (1993).
[7] See Schmidt and Sickles (1984).

above the predicted cost of the most efficient facility, or $\hat{u}_i - \min_j \hat{u}_j$. For GLS estimators to be consistent, the density of the inefficiency disturbances must be nonzero in the neighborhood of (0, ω) for some ω>0. In other words, as the number of facilities in the sample increases, the probability that a facility lies near the frontier approaches 1. Our efficiency measure is calculated using the log difference in average residuals between the facility on the cost frontier and the actual average residual of facility *i*, or exp($\min_j \hat{u}_j - \hat{u}_i$). This measure is bounded between 0 and 1, with the most efficient facility having an efficiency value equal to 1.

The Within technique differs from the GLS in that a dummy variable is included in the cost function for every site but one. Which one is omitted does not affect the final results for, as with the GLS technique, cost efficiency is measured relative to the best performing site. Only the method for obtaining the $\hat{u}_i$'s differ, not the way they are employed afterward. Including these dummy variables allows for the possibility that relative site efficiencies are correlated with the other regressors in the cost function. The GLS technique, by contrast, assumes that $\hat{u}_i$'s are independent of the other regressors.

In analyzing our empirical results, we employ a decomposition first employed by Bauer and Hancock (1995). For a given site, this technique enables one to determine the source of deviations in unit cost from the sample mean. We can rewrite the ratio of a site's average unit costs to that of the average unit costs for the entire sample as,[8]

[8]A modification of this decomposition could be employed to explore the sources of changes in unit costs over time.

(3)

$$\ln\left[\frac{(\overline{C}_i / \overline{y}_i)}{(\overline{C} / \overline{y})}\right] = \ln\left[\frac{C(\overline{y}_i, \overline{w}_i)\, exp(u_i + \overline{v}_i)}{\overline{y}_i}\right] - \ln\left[\frac{C(\overline{y}, \overline{w}) exp(\overline{u} + \overline{v})}{\overline{y}}\right]$$

Using the cost function defined in equation (1), equation (3), can be rearranged into the following expression,

$$\ln\left[\frac{(\overline{C}_i / \overline{y}_i)}{(\overline{C} / \overline{y})}\right] =$$

$$\left[\beta_y(\ln \overline{y}_i - \ln \overline{y}\ ) + 1/2\, \beta_{yy}(\ln \overline{y}_i^2 - \ln \overline{y}^2) - (\ln \overline{y}_i - \ln \overline{y}\ )\right]$$

$$+\left[\sum_{k=1}^{K} \gamma_k (\ln \overline{w}_{ki} - \ln \overline{w}_k) + 1/2 \sum_{k=1}^{K}\sum_{j=1}^{K} \delta_{kj} (\ln \overline{w}_{ki} \ln \overline{w}_{ji} - \ln \overline{w}_k \ln \overline{w}_j\right]$$

$$+\left[\sum_{k=1}^{K} \theta_k (\ln \overline{y}_i \ln \overline{w}_{ki} - \ln \overline{y} \ln \overline{w}\right]$$

$$+[u_i - \overline{u}]$$

(4) $$+\left[\overline{v}_i - \overline{v}\right]$$

where the expressions on successive lines can be defined as the effects resulting from differing scale effects, input price effects, interaction effects between processing volumes and input prices, cost inefficiency, and a random effect. While these effects are in logarithmic differences, they can be roughly interpreted as the percentage difference in costs stemming from the various effects.[9] Equation (4) provides a convenient framework for quantifying the source of cost savings over time. A similar decomposition is derived for the hybrid-translog (see authors for details).

Data

We collected quarterly data from 1990:I to 2000:IV on activity production costs, ACH processing volumes, and input prices for the 12 Federal Reserve Districts. Note that our units of observation are slightly different than earlier studies. Humphrey (1982, 1984, 1985) used cross-sectional data from various years from 1977 to 1982 and included every processing site with positive volume. Bauer and Hancock (1995) looked at 11 District Banks and the Los Angeles branch of the Federal Reserve Bank of San Francisco because those were the locations where most of the processing occurred in their sample from 1979 to 1994. New York was omitted because processing there was done by the New York Automated Clearing House Association (NYACH) over most of this period. Finally, Bauer and Ferrier (1996) compared cost performance across District Banks for ACH, Fedwire, and paper check, and hence combined the data for Los Angeles with San Franciso's while still omitting observations on New York. Like Bauer and Ferrier (1996), we aggregate up to the District level, but we include New York because in 1996 it began processing its volume on FedACH and we wanted to explore how this shift affected costs. To test the

[9]For the exact percentage difference, the antilog of each expression minus one should be utilized.

sensitivity of our results, we estimate our models with and without New York. Our estimates of scale economies and technical change are not greatly affected.

Our primary data source is the Federal Reserve's quarterly Planning and Control System (PACS), a time series of functional cost accounting reports collected in order to monitor costs and improve resource allocation within the System. These data are supplemented by price index information from the Commerce Department's Bureau of Economic Analysis, the Labor Department's Bureau of Labor Statistics, ACCRA Cost of Living Index, and R.S. Means Company. For interest rates required by the Jorgenson and Hall (1967) equations to measure capital services, we used Moody's Aaa corporate bond rate.

Production costs for processed ACH items are included in total costs, but certain overhead expenses, such as special district projects, were excluded. Our output measure is the number of ACH payments originated and received rather than the site-specific measure used in previous studies that employed the number of payments processed at each site. Prior to 1996, only the number of items processed was reported. After 1996 this statistic is no longer reported, but our new definition of output, the sum of the number of originations and receipts divided by two, is reported. There are two key differences in these two measures of output. The first definition includes 1/3 of addemdum records, additional information about the transaction increasingly used in electronic data interchange, whereas the latter one does not. The second key difference is that with the first definition, inter-District items get counted more than once because each District "processes" the item. In the latter definition, every item is only counted once as an origination and once as a receipt.

We splice the data together in two steps because there is the additional wrinkle that

neither definition of output was collected in 1996, the year when most of the accounting changes were being implemented. We do have data on originations for 1996, just not receipts. We assume that the ratio of originations to receipts is roughly constant in the short run and use their ratio from 1997 to obtain an estimate of 1996 receipts.

To extend the current definition of output back to the 1990 to 1995 period, we assume that the ratio of addendum items/total items and the ratio of inter-District/total items are relatively constant over this period. If this is the case, then first definition = constant x second definition. In order to estimate this constant, we use data from 1990 to 1995 to forecast the old definition of output for 1996 and use data from 1996 to 2000 to forecast the new definition of output for 1995. This gives us two years, 8 quarters of observations, with overlapping estimates of the two output definitions. For each site we take the average ratio of the old to the new definition of output and use that estimate of the constant to convert that site's output from 1990 to 1995 to the new definition of output. While the output series calls to mind something the Dr. Frankenstein might concoct, the result is an output series for each site that looks reasonably coherent over time. To test the robustness of our results, we estimate the various models over various sub-periods in order. We have complete results for 1990 to 1995 and 1997 to 2000, although in this paper we focus on the results for 1990 to 2000. Full details on the splicing techniques and the sub-period results are available from the authors.[10]

Buildings, materials, communication, and labor are inputs for ACH processing. For the

[10]To determine whether this new output measure affects our cost function estimates, we re-estimated Bauer and Ferrier's (1996) model using 1990 to 1994 data and the results were extremely close to theirs. This is not surprising given the correlation between the two output measures is greater than 99 percent over this period when the two output series are both available.

1990 to 2000 period, labor expenditures—salaries, retirement, and other benefits—accounted for 18.3% of total costs.[11] To calculate the price of labor, we divide the expenditures for labor, including salaries, retirement, and other benefits by the number of employee hours spent processing ACH transactions.

Over the 1990 to 2000 period, buildings' total cost share is only 2.4%, 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 private-sector adjustment factor (PSAF), rather than in activity production costs. Cost accounting information is supplemented by annual replacement-cost indexes, by city, available from the R.S. Means Company.[12] Square-foot replacement costs, adjusted by the depreciation rate, are used to price maintenance and building expenses for each site.

Expenditures for materials (office equipment and supplies, printing and duplicating, data processing, and computers) account for 48.0% of costs over this period. The service price for 'materials' is constructed by supplementing cost-accounting expenditure data with indexes for information and processing equipment.[13] For computer hardware, an estimate of the service value, or price, of machines is constructed using formulas derived by Hall and Jorgenson (1967). Data system support services are primarily for in-house product-specific software development. Utilizing expenditures for labor and hours worked in the data support services area of each Reserve

---

[11]Labor's contribution to ACH processing costs has declined from 33 percent in 1979 when Humphrey (1981) first estimated an ACH cost function.

[12]Data on replacement costs for buildings were taken from Volume 27, Number 1, *Means Construction Cost Indexes*, Means (2001).

[13]The implicit deflator for information processing and related equipment, from the Bureau of Economic Analysis, was used for data processing and computer hardware.

Bank, we construct a service price for data system support services. For the service price of supplies—printing, and duplicating, office supplies, and office equipment—we use the implicit price deflator for gross domestic product. We constructed a Theil-Tornquist (TT) (Theil, 1968, Tornquist, 1936) price index for 'materials' that used the expenditures and prices for the components of 'materials'—data processing, data systems support, and office supplies and equipment.

Communications expenditures--the expenses associated with data and other communications, shipping, and travel--made up 31.4 percent of total costs. The implicit price deflator for communications equipment purchases by non-residential producers is used for data and other communications. The fixed-weight aircraft price index aircraft for private purchases of producers' durable equipment is used to measure time series variation in shipping and travel expenditures, while the transportation component of the ACCRA cost of living index is used to measure cross sectional variation. Unfortunately data for some cities and some time periods are not available for the ACCRA data. Thus, data for cities near Federal Reserve Bank cities and interpolation were both used to fill in the gaps. A price index for communications expenditures is constructed using the expenditure shares and price indexes for data and other communications, and for shipping and travel. The index, taken from Severance-Lossin (1996), is designed to be used with panel data, and combines the desirable features of the Caves, Christiansen and Diewert (1982) index and the TT index. In particular the index is superlative for the translog cost function, and does not require revision as new data comes in, or a restricted time period is examined.

One last difference between this study and previous ones needs to be discussed at this point. We do not include any environmental variables because none exist that are available throughout

the whole sample. To explore how this may affect our results, we estimate the model using data from 1990 to 1995 with and without these variables. Over this time period, the inclusion or exclusion of environmental variables did not materially affect our estimates of scale economies or technical change.

Results

Our central focus is on estimating how much scale economies and technical change have contributed to the rapid unit cost declines for ACH processing from about $0.04 in 1990 to $0.01 in 2000. As the earlier studies have pointed out, this is a tricky endeavor because of the high correlation between output growth and the passage of time. We explore the robustness of our results by estimating translog and hybrid- translog versions of cost function using both GLS and Within estimation techniques. Further perturbations involve estimating the over three sample periods, 1990 to 2000, 1990 to 1995, and 1997 to 2000. In our sub-samples we omit data from 1996 because that is the year when the ACH data reporting underwent the most wrenching changes. We also estimate the models with and without New York. Mercifully, we do not inflict upon the reader all 24 permutations of the results. Mainly we will discuss the full sample GLS/Translog results, noting when necessary how the results change as various as aspects of the model are altered.

Technological Change

Figure 3 plots estimates of technological change obtained using the various estimation models described above. All of the technological change indexes were normalized to equal one in 1990. Being heavily dependent on computers and telecommunications equipment, it is not surprising that technical change accounts for a large share of the unit cost declines from 1990 to 2000. It is equally clear that estimates of technical change are not sensitive to whether the translog

or hybrid-translog functional form is employed, but they are sensitive to whether the GLS or Within estimation technique is employed (that is, whether the cost efficiency terms are independent of the other regressors or not). Technical change is estimated to be much more rapid with the GLS approach, falling almost 70 percent other things held equal, as opposed to "only" falling about 50 percent with the Within. As we will see, because each technique is seeking to explain the source of the same overall unit cost decline for ACH processing, to accommodate more rapid estimates of technical change, the GLS technique estimates less scale economies.

The transition to the new FedACH processing platform appear to have hit hardest in 1995. The Within estimates actually suggest that technical change caused the cost function to rise 6 to 7% depending on the functional form employed after having fallen about 8.5% on average over the previous four years. Qualitatively similarly, the corresponding GLS estimates suggest that technical change over this period slowed to 2.4 to 3.5% in 1995 from about 13% in the previous four years. These transition costs appear to have been well worth while. Over the next three years, the Within estimates suggest that technical change averaged 12% per year, while the GLS estimates suggest 15%.

By 1998, the one-time gains of the reorganization appear to have been reaped as the rate of decline in costs due to technical change appears to have slowed markedly. The slowdown is clearer for the Within estimates. From 1990 to 1998, technical change progressed at about 8 percent for both the translog and hybrid translog functional form, but from 1998 to 2000, the estimated rate actually rose 1% and 2% for the translog and hybrid-translog, respectively. For the GLS results there is also evidence of a slowdown: From 1990 to 1998 the estimated rate is about 12.7% for both functional forms, but for 1998 to 2000 the estimated rate of technical change slowed to 3.0% and

2.4% for the translog and hybrid-translog, respectively. If this slower rate of technical change continues over time, it could make cost recovery more challenging.

Scale Economies

Figure 4 plots estimates of the average cost function for the year 2000. Again, qualitatively the results are the same across the various models. All exhibit declining average cost functions throughout the full range of observed output indicating the presence of scale economies. Note particularly how similar the translog and hybrid-translog models tend to be in this application. Cost elasticities measure the effect of a 1 percentage point increase in output on total cost. For example, a cost elasticity of 0.7 means that if output increases 1%, costs would rise only 0.7%. A cost elasticity *less* than one indicates that are *scale economies* (that is, that average cost falls as output increases). Alternatively, a cost elasticity *greater* than one indicates that there are *scale diseconomies* (that is, that average cost rises as output increases).[14] As expected, the Within estimates attribute more of the unit cost decline to scale economies rather than to technical change. Over the full sample, the volume-weighted cost elasticity for the GLS/translog is 0.655 and 0.666 for the GLS/hybrid-translog. These are fairly significant scale economies, but not nearly as large as those estimated with the Within technique, which range from 0.194 to 0.390.

While the GLS estimates suggest that there remains unexploited scale economies for even the largest sites, the Within estimates suggest that the largest site may have reached scale efficiency and that no further cost savings can be had from increasing volume. In general we have more faith in the GLS results, so we do not believe that this is actually the case, but only time will tell as more

[14] Returns to scale was originally defined with respect to the production frontier. Duality theory results in the cost elasticity=1/RTS.

quarters of data are logged enabling further research.

Marginal Costs

Under the Monetary Control Act (MCA) of 1980, the Federal Reserve must set fees to recover all direct and indirect costs actually incurred providing payment services, including "interest on items credited prior to actual collection, overhead, and an allocation of imputed costs which takes into account the taxes that would have been paid and the return on capital that would have been provided had the services been provided by a private business firm." Thus, the MCA requires that total revenues raised from providing payment services will match the total costs incurred in production.

Bauer and Hancock (1995) discuss some pricing strategies the Reserve Banks could follow. The Federal Reserve charges a mixture of fixed and variable fees in order to recover the full economic costs of providing this service. The variable fee is intended to recover at least the marginal cost of production whereas the fixed fees makes up the shortfall as a result of scale economies (average costs exceeding marginal costs).

Site-specific estimates of marginal costs and the cost elasticities for the GLS/translog model are presented in Table 1 using data from 2000:IV. The table also report upper and lower confidence bounds for marginal costs and the site specific cost elasticities. The confidence bounds are relatively tight and the cost elasticities only range from 0.576 to 0.658. Table 2 reports the same statistics for the Within/translog model. Although the volume-weighted averages are very similar to those of the GLS/translog model, the site-specific estimates are much more variable. In 2000:IV, the cost elasticities range from about 0.34 to 1.1 and the upper and lower bounds of the marginal cost estimates are much wider. In addition to having less apparent statistical precision (because of

the wider upper and lower bounds) the Within/translog estimates of the average cost function have more curvature (in Figure 4, the Within average cost curves are higher for low output levels of output and lower for high output levels).

To save space we do not report the corresponding tables for the GLS/hybrid and Within/hybrid models because in each case the results are very similar to the estimator's translog results, as one would suspect given the close pairings revealed in Figure 4.

Cost Efficiency

Estimates of site-specific cost efficiency are plotted in Figure 5. As in previous papers, the magnitude of these estimates, 0.824 overall for the GLS/translog model, is similar to those found for private firms in the financial services industry. The pattern for the efficiency estimates is fairly similar across the models and with those reported by Bauer and Ferrier (1996). The correlation is around 0.7 for the various series.

Relative site performance can be best understood by looking at Table 5, which decomposes the unit cost differential observed across sites into their various sources using the technique described in the model section using the parameter estimates from the GLS/translog model. Given the log linear nature of the cost functions used in this paper, it is most straightforward to consider the decomposition in logarithmic units. To convert these numbers to percentage differences from the average Reserve Bank, one would need to take the antilog of the numbers reported here, subtract 1, and multiply by 100. For numbers near zero, just multiplying by 100 gives the approximate percentage of this site from the sample mean (this is a first order Taylor approximation). For example, using the approximation from 1990 to 2000 unit costs at the first site are about 4.1% lower than the average site (100x-0.041 = -4.1%). The precise (to 2 significant

digits answer) would be 100 [exp(-0.041)-1] = -0.040.

Average unit costs varied quite a bit across sites in this time period. For example, unit costs are highest at the second site, roughly 36% above the overall mean, primarily because of this site's relatively low level of estimated cost efficiency, about 28.6% higher than the overall average. Input prices also play a role. Input prices lead unit costs at this site to be about 8% higher than the overall average, the largest value for any site. Scale economies and interaction effects do not seem to play much of a role in explaining why this site's unit costs differ from the average.

Alternatively consider the sixth site, which has the lowest unit costs. Its lower than average unit costs, 29.9%, appear to be due to scale economies resulting from its larger than average volume, 17%, but also as a result of its better than average cost efficiency performance, 10%. Because the estimates of scale economies and cost efficiency are similar for the GLS/hybrid results, the corresponding decomposition for that model is similar and we do not reported them here.

Given that the estimates for scale economies and cost efficiency are somewhat different for the Within/translog estimator, it is not surprising, the corresponding decomposition, Table 6, also tells a bit different story. Again just looking at the extreme unit cost performers in Table 6, the second site's high unit costs are again explained by low estimated cost efficiency, 33.4%, and slightly higher than average input prices, 8%.[15] But source of the sixth's site superior unit cost performance no longer appears to come from a combination of better-than-average cost efficiency and scale economies. With the Within/translog estimates, this site now has costs about 13% higher than the average site, but because of the higher estimated scale economies of this model the Total

[15] Note that the estimated values for the Direct Input Price component of the decomposition are virtually the same in Tables 5 and 6. This is no typo. The relevant estimated coefficients for the two models are actually that close. The same holds for the Effect of Interaction Components, which are also very close in numerical value..

Output effect, 35.9%, is more than enough to offset this shift.

Conclusion

Using data from 1990 through 2000, we find that the ACH cost function continues to experience both rapid technical change and significant scale economies, yet there are some signs that the technology may be maturing. Technical change has slowed some since 1998. These results are generally robust across a variety of perturbations to the model such as employing a hybrid-translog functional form instead of a translog, estimating the cost function parameters using a Within estimator instead of GLS, and excluding New York or keeping this site in the sample.

Understanding the how costs are likely to evolve over the next several years should aid Reserve Banks in the increasingly competitive market for ACH services. Given magnitude of scale economies, the consolidation of customer support that was implemented in 2001 certainly appears to be well advised. Further opportunities for consolidation might lead to other cost saving opportunities. The magnitude of scale economies also suggests that Reserve Banks should be aggressive in trying to attract not just new volume but also in trying to retain current volume—particularly that of large originators given the high share of volume. In the long run, retaining volume can benefit all Reserve Bank customers because with scale economies larger volume leads to lower unit costs.

There are several possible strategies that Reserve Banks could pursue. Currently both originators and recipients roughly split the cost of paying for this service. Depending on the demand elasticities, costs may be more effectively recouped by shifting more of the burden to recipients. Another strategy would be to offer volume discounts, which Reserve Banks already

offer, but might need to be employed more aggressively.

References

ACCRA Cost of Living Index, 1990 - 2000 (Volume 23 No. 1 through Volume 33 No. 4), Alexandria, VA.

Bauer, Paul B., Allen N. Berger, and David B. Humphrey (1993) "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 University Press, Oxford), pp. 386-413.

Bauer, Paul W. and Gary Ferrier (1996) "Scale Economies, Cost Efficiencies, and Technological Change in Federal Reserve Payments Processing," *Journal of Money, Credit and Banking,* 28 (4), pp. 1004-1039.

Bauer, Paul W. and Diana Hancock, "Efficiency and Technical Progress in Check Processing*," Economic Review*, Federal Reserve Bank of Cleveland, 1995, vol. 29, no. 3, pp. 24-38.

Berger, Allen N. "'Distribution-Free' Estimates of Efficiency in the U.S. Banking Industry and Tests of the Standard Distributional Assumptions." *Journal of Productivity Analysis* 4 (September 1993), 261-292.

Diewert, W.E. "Duality Approaches to Microeconomic Theory," in *Handbook of Mathematical Economics,* Volume II, eds. K.J. Arrow and M.D. Intrilligator, North-Holland, 1982, pp.535-599.

Caves, Douglas W.; Christensen, Laurits R.; Diewert, W. Erwin, "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers," *Economic Journal*, March 1982,. 92, iss. 365, pp. 73-86

Ferrier, Gary and C.A.K. Lovell (1990) "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence," *Journal of Econometrics*, 46, pp. 229-245.

Hall, R.E. and D.W. Jorgenson. "Tax Policy and Investment Behavior," *American Economic Review,* 57, 391-414, 1967.

Humphrey, David B. "Scale Economies at Automated Clearinghouses." *Journal of Bank Research* 12 (Summer 1981), 71-81 (b).

_____. "Costs, Scale Economies, Competition, and Product Mix in the U.S. Payments

Mechanism," *Staff, Board of Governors of the Federal Reserve System*, April 1-18, 1982.

_____. "The U.S. Payments System: Costs, Pricing, Competition and Risk," *Monograph Series in Finance and Economics*, Monograph 1984-1/2.

_____. "Resource Use in Federal Reserve Check and ACH Operations after Pricing," *Journal of Bank Research*, vol. 12 (1985), pp. 45-53.

_____. "Delivering Deposit Services: ATMs Versus Branches," *Economic Quarterly,* Federal Reserve Bank of Richmond, Spring 1994, pp. 59-81.

_____, and A. N. Berger. "Market Failure and Resource Use: Economic Incentives to Use Different Payment Instruments," in *The U.S. Payment System: Efficiency, Risk and the Role of the Federal Reserve*, ed. by D. B. Humphrey, Kluwer Academic Publishers, 1990, pp. 45-86.

Knudsen, S.E., J.K. Walton, and F.M. Young, "Business-to-Business Payments and the Role of Financial Data Electronic Interchange, *Federal Reserve Bulletin*, Board of Governors of the Federal Reserve System, April 1994, pp. 269-278.

Means, R.S. *Means Construction Cost Indexes,* Volume 27, Number 1, January 2001, R.S. Means Company, Kingston, Massachusetts.

Severance-Lossin, Eric, "A Multilateral Index for Panel Data," working paper, Oct. 16, 1996.

Schmidt, Peter and Robin C. Sickels (1984) "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics,* 2, pp. 367-374.

Theil, H. "On the Geometry and the Numerical Approximation of Cost of Living and Real Income Indices," *De Economist*, vol. 116 (1968), pp. 677-689.

Tornquist, L. "The Bank of Finland's Consumption Price Index," *Bank of Finland Monthly Bulleting*, vol. 10 (1936), pp. 1-8.

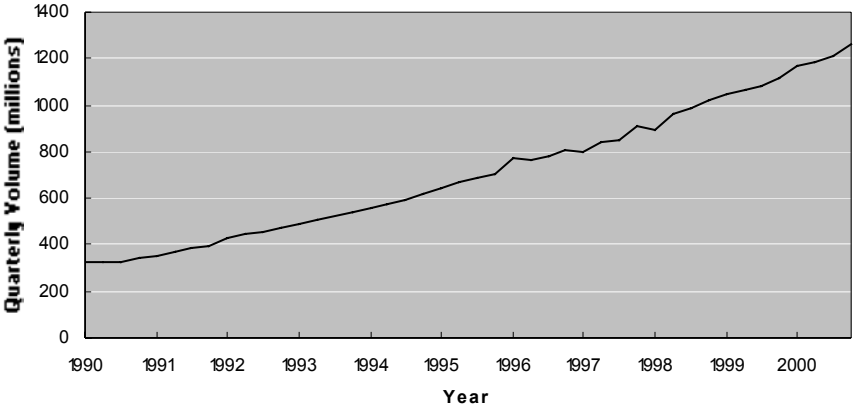
Walker, D.A. "Economies of Scale in Electronic Funds Transfer Systems," *Journal of Banking and Finance,* 2, 1978, pp. 65-78.

Wells, K.E. "The Social Costs of Paper and Electronic Payments," Master of Arts thesis, Virginia Polytechnic Institute, 1994

U.S. Department of Commerce, Office of Building Economics, 1978-1994, *Survey of Current Business,* U.S. Government Printing Office, Washington, D.C.

Zimmerman, G.C. "The Pricing of Federal Reserve Services Under MCA," *Economic Review,* Federal Reserve Bank of San Francisco, Winter 1981, pp. 22-40.

Figure 1, TotalACH Volume



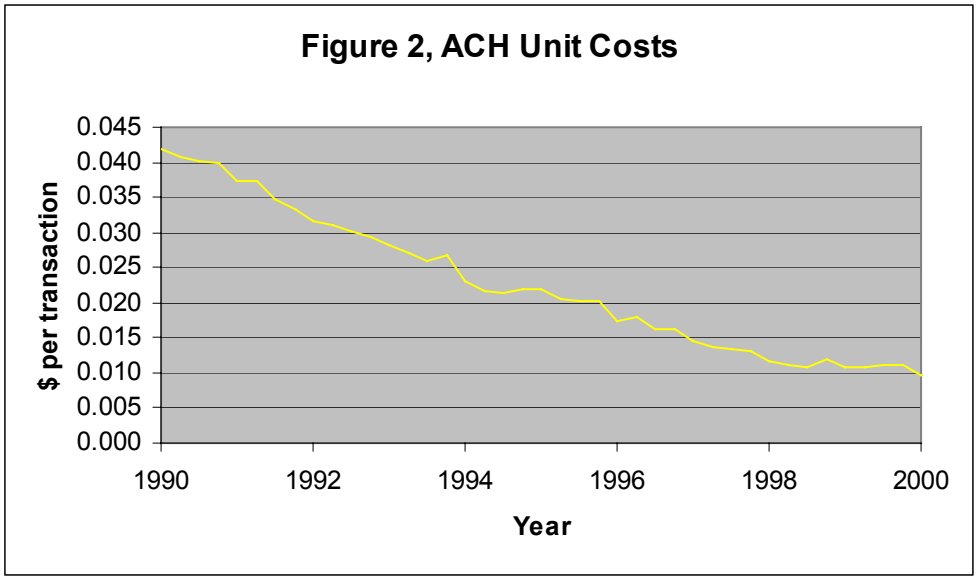


Figure 2, ACH Unit Costs
$ per transaction
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
1990
1992
1994
1996
1998
2000
Year

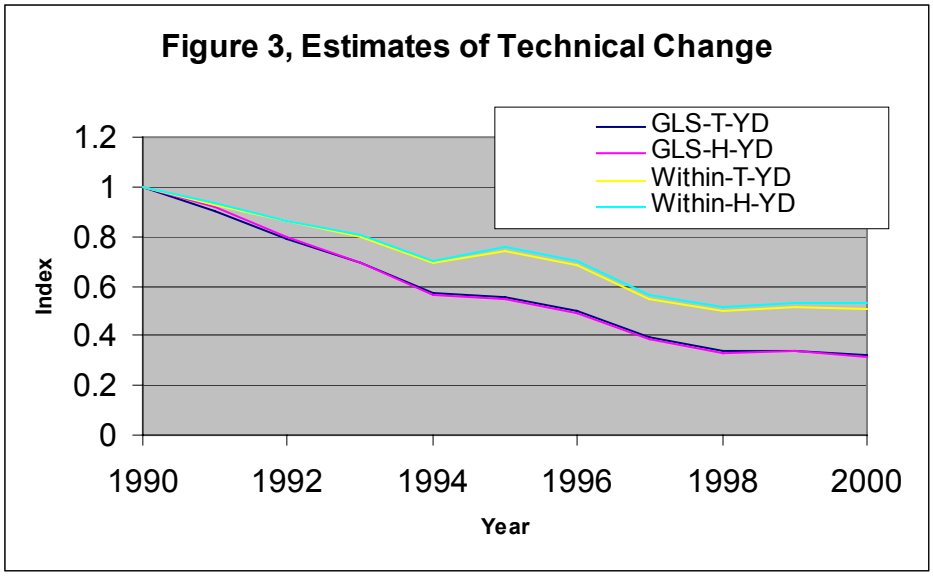


Figure 3, Estimates of Technical Change
GLS-T-YD
GLS-H-YD
Within-T-YD
Within-H-YD
1.2
1
0.8
0.6
0.4
0.2
0
Index
1990
1992
1994
1996
1998
2000
Year

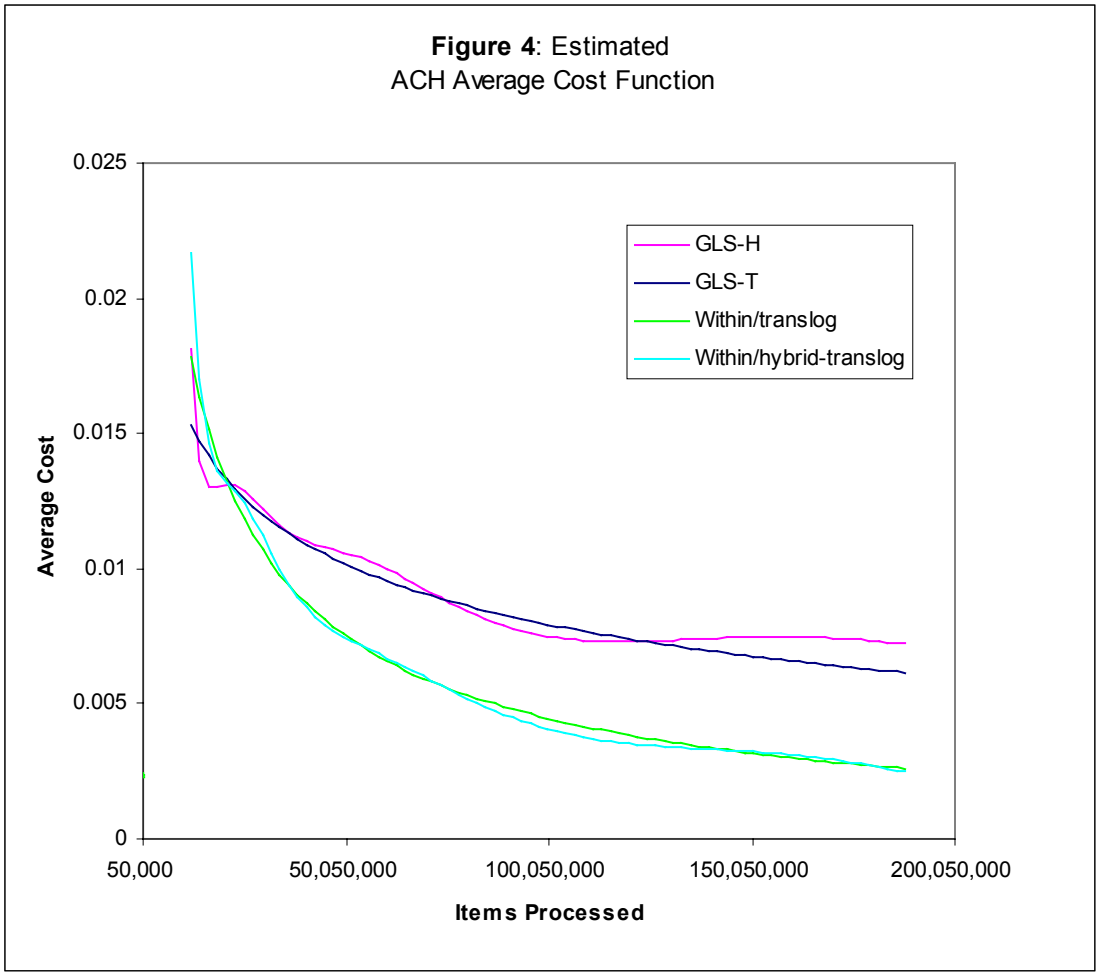


Figure 4: Estimated
ACH Average Cost Function
0.025
0.02
0.015
0.01
0.005
0
Average Cost
50,000
50,050,000
100,050,000
150,050,000
200,050,000
Items Processed
GLS-H
GLS-T
Within/translog
Within/hybrid-translog

Table 1,
GLS/translog

ACH, 2000:IV

Site	Unit Cost	Marginal Cost	Upper	Lower	Cost Elasticity
1	0.0091	0.0059	0.0063	0.0054	0.647
2	0.0156	0.0096	0.0108	0.0083	0.615
3	0.0096	0.0061	0.0067	0.0055	0.631
4	0.0046	0.0028	0.0032	0.0025	0.614
5	0.0072	0.0044	0.0050	0.0039	0.614
6	0.0112	0.0067	0.0077	0.0057	0.600
7	0.0091	0.0052	0.0063	0.0041	0.576
8	0.0131	0.0086	0.0092	0.0081	0.658
9	0.0141	0.0087	0.0097	0.0077	0.621
10	0.0112	0.0070	0.0077	0.0063	0.628
11	0.0132	0.0084	0.0091	0.0077	0.639
12	0.009	0.005	0.007	0.004	0.585
Volume weighted Average	0.0103	0.0063			0.611

Table 2, ACH,
Within/translog 2000:IV

District	Unit Cost	Marginal Cost	Upper	Lower	Cost Elasticity
1	0.0091	0.0040	0.0076	0.0004	0.4380
2	0.0156	0.0133	0.0172	0.0094	0.8562
3	0.0096	0.0034	0.0067	0.0000	0.3494
4	0.0046	0.0034	0.0043	0.0024	0.7352
5	0.0072	0.0054	0.0069	0.0038	0.7460
6	0.0112	0.0123	0.0175	0.0070	1.0978
7	0.0091	0.0054	0.0245	-0.0137	0.5980
8	0.0131	0.0093	0.0116	0.0070	0.7088
9	0.0141	0.0078	0.0110	0.0046	0.5519
10	0.0112	0.0044	0.0079	0.0010	0.3951
11	0.0132	0.0045	0.0098	-0.0007	0.3441
12	0.0091	0.0084	0.0163	0.0004	0.9195
wtd avg.	0.0103	0.0070			0.685

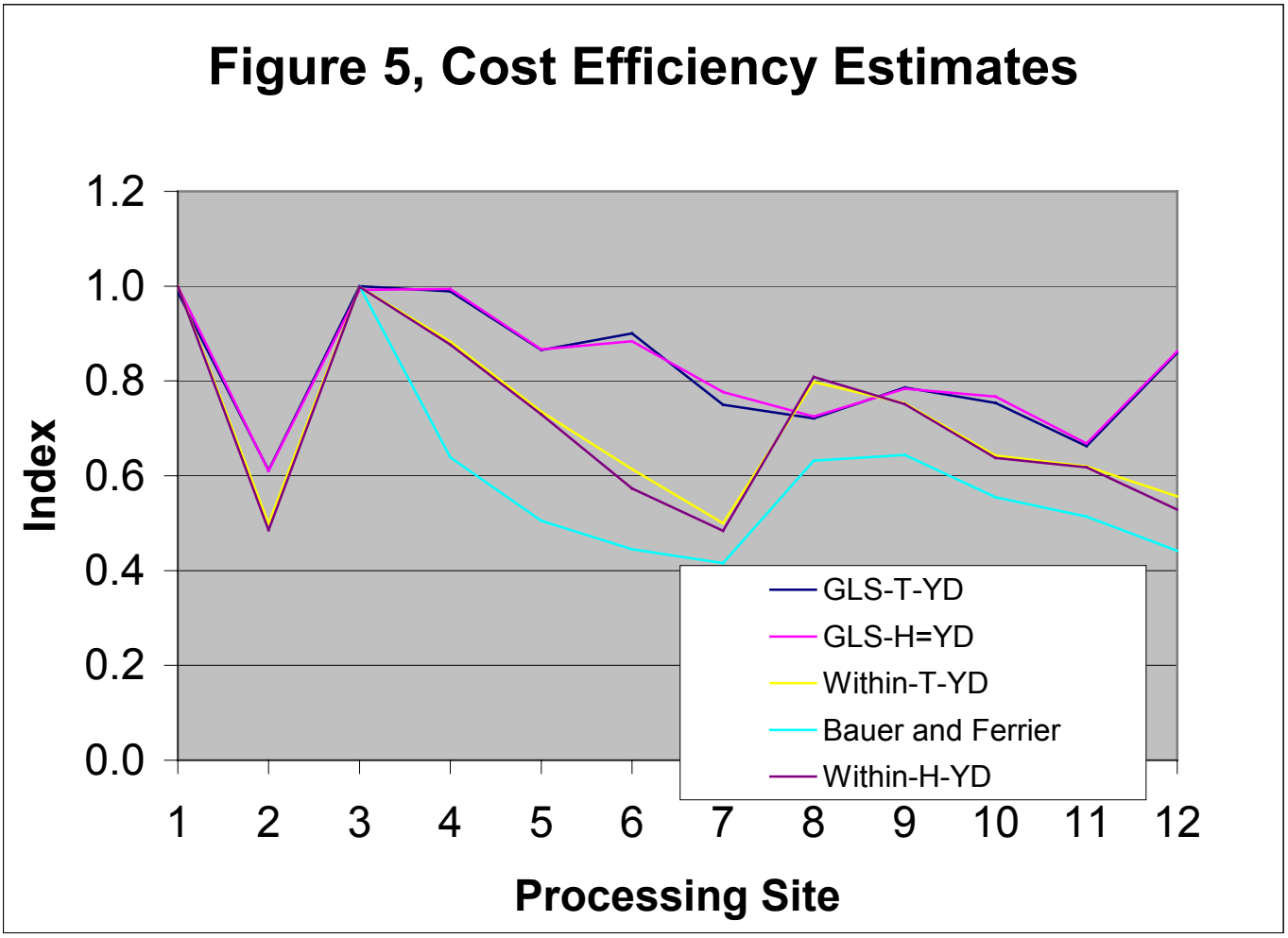


Figure 5, Cost Efficiency Estimates
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Index
1
2
3
4
5
6
7
8
9
10
11
12
Processing Site
GLS-T-YD
GLS-H=YD
Within-T-YD
Bauer and Ferrier
Within-H-YD

Table 5, Unit Cost Decomposition, GLS/translog

Processing Site	Unit Cost ($)	Logarithmic Differences from Sample Mean				
		Unit Cost	Cost Efficiency	Total Output	Direct Input Price	Effect of Interactions
s1	0.0195	-0.041	-0.192	0.107	0.045	-0.001
s2	0.0291	0.360	0.286	-0.006	0.080	0.000
s3	0.0184	-0.095	-0.206	0.119	-0.009	0.000
s4	0.0170	-0.179	-0.194	0.033	-0.017	0.000
s5	0.0186	-0.089	-0.061	-0.001	-0.027	0.000
s6	0.0150	-0.299	-0.100	-0.170	-0.028	-0.001
s7	0.0173	-0.159	0.082	-0.248	0.007	0.000
s8	0.0271	0.289	0.121	0.202	-0.035	0.001
s9	0.0225	0.104	0.035	0.092	-0.024	0.000
s10	0.0220	0.083	0.077	0.018	-0.013	0.000
s11	0.0267	0.273	0.206	0.077	-0.011	0.000
s12	0.0159	-0.246	-0.054	-0.224	0.032	0.000

Average over 1990:I to 2000:IV

Table 6, Unit Cost Decomposition, Within/translog

Processing Site	Unit Cost ($)	Logarithmic Differences from Sample Mean				
		Unit Cost	Cost Efficiency	Total Output	Direct Input Price	Effect of Interactions
s1	0.0195	-0.041	-0.360	0.226	0.045	-0.001
s2	0.0291	0.360	0.334	-0.012	0.080	0.000
s3	0.0184	-0.095	-0.359	0.251	-0.009	0.000
s4	0.0170	-0.179	-0.236	0.070	-0.017	0.000
s5	0.0186	-0.089	-0.053	-0.003	-0.027	0.000
s6	0.0150	-0.299	0.128	-0.359	-0.027	-0.001
s7	0.0173	-0.159	0.332	-0.524	0.007	0.000
s8	0.0271	0.289	-0.135	0.427	-0.035	0.002
s9	0.0225	0.104	-0.077	0.195	-0.024	0.000
s10	0.0220	0.083	0.082	0.039	-0.013	0.000
s11	0.0267	0.273	0.118	0.163	-0.011	0.000
s12	0.0159	-0.246	0.226	-0.473	0.032	0.000

Average over 1990:I to 2000:IV