working paper

9810

Optimal Employment of Scale Economies in the Federal Reserve's Currency Infrastructure

by Paul W. Bauer, Apostolos Burnetas, Viswanath CVSA, and Gregory Reynolds

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 9810

OPTIMAL EMPLOYMENT OF SCALE ECONOMIES IN THE FEDERAL RESERVE'S CURRENCY INFRASTRUCTURE

by Paul W. Bauer, Apostolos Burnetas, Viswanath CVSA, and Gregory Reynolds

Paul W. Bauer is at the Federal Reserve Bank of Cleveland; Apostolos Burnetas and Viswanath CVSA are at the Weatherhead School of Management, Case Western Reserve University; and Gregory Reynolds is at Hewitt Associates.

Working papers of the Federal Reserve Bank of Cleveland are preliminary materials circulated to stimulate discussion and critical comment on research in progress. They may not have been subject to the formal editorial review accorded official Federal Reserve Bank of Cleveland publications. The views stated herein are those of the authors and are not necessarily those of the Federal Reserve Bank of Cleveland or of the Board of Governors of the Federal Reserve System.

Working papers are now available electronically through the Cleveland Fed's home page on the World Wide Web: http://www.clev.frb.org.

July 1998

Optimal Employment of Scale Economies in the Federal Reserve's Currency Infrastructure

Paul W. Bauer
Financial Services Research Group
Federal Reserve Bank of Cleveland

Apostolos Burnetas
Weatherhead School of Management
Case Western Reserve University

Viswanath CVSA
Weatherhead School of Management
Case Western Reserve University

Gregory Reynolds
Hewitt Associates

Abstract

This paper investigates whether the Federal Reserve might lower its currency processing costs by reallocating high-speed currency sorting volume among its processing sites. Although scale economy estimates from Bauer, Bohn, and Hancock (1998) suggests that consolidation might permit some processing cost savings, it can be very expensive to ship currency due to security and insurance requirements and these costs increase rapidly as currency is transported further and further from a given processing site. Given estimates of currency shipping costs and scale economies for high-speed sorting, our model determines the distribution of sorting volumes across possible processing sites that minimizes the Federal Reserve's overall costs. These cost savings are achieved while leaving service levels to depository institutions roughly constant. The sensitivity of our results is explored by employing a range of estimates for shipping costs and scale economies.

August 1998

DO NOT QUOTE WITHOUT AUTHORS' PERMISSION

1. Introduction

The Federal Reserve's involvement in the provision of paper currency dates back to its founding in 1913. In fact, the major impetus for the creation of the Federal Reserve system was to furnish an "elastic currency" in an attempt to make financial crises, such as the Panic of 1907, less likely. While much has changed in the financial landscape over the last 80 years, currency remains an integral part of the U.S. payments system.

For many types of transactions, currency has withstood the onslaught of checking accounts and credit cards, and will likely hold an edge over debit and smart cards for some types of transactions for the foreseeable future. The reasons for this are simple: currency offers finality, anonymity, and familiarity at a reasonably low cost for small-value transactions, setting a high hurdle for other payment instruments. This is good news for the Treasury, because the 15.5 billion Federal Reserve notes (with a value of $344.5 billion in 1993) are backed by Treasury debt. The $20 billion a year in interest payments that the backing of these notes kick off is indirectly remitted to the Treasury.[1] Ultimately, the interaction of payment instrument providers, payors, payees, and regulators will determine the market shares of various payment instruments, but it is likely that currency will continue to retain a significant share of transactions.

The Federal Reserve has a responsibility to operate its service efficiently. Even if there were no evolution in the market as a result of new payments instruments, the advent of nationwide branching would have necessitated a reexamination of the currency infrastructure.[2] This paper studies whether a redeployment of resources can lower the Federal Reserve's costs while still providing depository institutions (DIs) roughly the same level of service. More precisely, we explore the optimal solution to two optimization models for the Federal Reserve. The first starts with the current 37 processing sites and determines how the processing volume should be allocated among them.[3] In other words, if we limit our consideration only to the locations where the Federal Reserve already has processing sites, how should processing volume be allocated to minimize overall costs (processing and shipping costs). In the other scenario, we optimize the same model, but this one starts with the cities that anchor Rand McNally's 46 Major Trading Areas (MTAs), except Honolulu.[4] This is essentially a "green field" approach: if we are allowed to consider the 46 largest metropolitan areas, where would the Federal Reserve choose to locate processing sites to minimize overall costs? This scenario provides us with an estimate of the maximum possible cost savings from reallocating currency volume.

---

[1] Anywhere from half to two-thirds of US currency is held overseas for reasons that are unlikely to be affected by payment instrument innovation, so this revenue source will likely continue for the foreseeable future, although the European Community's new "Euro" could provide some stout competition. Foreign holdings generally involve the higher denominations, mostly $100's.

[2] The first step in this process has already started with the announcement of the Uniform Cash Access Policy (UCAP) in April of 1996. This policy is designed to achieve a uniform and consistent level of cash access service across the nation in the distribution of currency. See the Federal Register notice dated April 25, 1996 for a detailed discussion of the UCAP policy.

[3] The cost implications for other Federal Reserve services remaining at these sites is not considered here.

[4] These 46 MTAs include the 37 cities with Federal Reserve processing sites (except Helena) plus 9 others. While Honolulu is also designated an MTA, no sites in Hawaii or Alaska are considered in our analysis.

Like the ultimate question to life, the universe, and everything[5], the appropriate objective function for the Federal Reserve to pursue in the provision of currency has not been explicitly stated. While it is widely acknowledged that the Federal Reserve should run its currency operations in a cost-efficient manner and the Uniform Cash Access Policy (see footnote 2) specifies the level of service that DIs should receive from existing Federal Reserves sites, many aspects of the Federal Reserve's behavioral objective have not been explicitly defined. For instance, how many processing sites should the Federal Reserve operate? More sites would mean a higher level of service to DIs, perhaps boosting the use of currency and indirectly reducing the borrowing needs of the Treasury. Alternatively, fewer sites would lower service levels, but might lower costs more than enough to offset any such losses to the Treasury.

Because such issues are far beyond the scope of this paper, our model seeks to provide DIs roughly their current service level. By adopting this constraint, minimizing the Federal Reserve's costs is equivalent to maximizing social welfare (in this limited context). In the model this is accomplished by having the Federal Reserve continue to pay for shipping currency to and from the location when a processing site is closed. This accounts for most of the social costs such a closing would impose on third parties.[6] Whether the Federal Reserve actually picks up these costs is a policy matter that we will not consider here. Another policy matter that we do not address is the cost implications for other Federal Reserve services remaining at these sites.[7]

Our key finding is that while the Federal Reserve may be able to save almost 20 percent of its controllable costs by reallocating volume, most of the $5 million dollars cost savings can be obtained by reallocating processing volume without closing any processing sites. In addition, only a handful of Federal Reserve sites appear to be candidates for closing, a decision which would require examining a number of issues that we avoid here such as transition costs and the impact on other Federal Reserve services. Finally, our "green field" approach suggests that the current geographic distribution of Federal Reserve sites is very close to the optimum. The most significant departures are that our model would not operate a site in Helena and would open sites in Phoenix and Milwaukee.

The rest of the paper is organized as follows. A brief description of the Federal Reserve's currency service is presented in Section 2. A mathematical programming model that determines the optimal cash processing volumes and shipment schedule to minimize overall costs is developed in Section 3. This model is subsequently solved using estimated data on the demand for cash and the processing and shipping costs. The results are presented in Section 4, together with a discussion on the robustness of the solution

---

[5] See "The Hitchhiker's Guide to the Galaxy," by Douglas Adams.

[6] This is not exactly what the Federal Reserve Bank of Cleveland did when the Pittsburgh branch's high-speed sorting was moved to Cleveland. In this case, paying and receiving has been maintained, so far, in Pittsburgh. Our model assumes that these operations are also removed from any site that is closed.

[7] If a service is removed from a location, the building space previously allocated to that service and possibly some overhead expenses must be recovered from the remaining services. This could have an adverse effect on "priced" services that have to recover their full economic costs through user fees.

within a range of estimates for the shipping costs. Some conclusions and possible extensions to the model are discussed in Section 5.

2. Overview of Currency Service

DIs deposit their excess currency at the nearest Federal Reserve currency processing site, where it is counted and verified. An important part of the verification process is separating fit notes from worn notes and possible counterfeits. Processing sites hold fit notes (non-worn processed notes as well as new notes) until a DI requests a cash order. Unfit notes are destroyed and their face value is deducted from the total amount of Federal Reserve notes outstanding. Counterfeits are turned over to the Secret Service for further investigation.

Paying and Receiving operations involve having a highly secure area for DIs to deposit and receive currency deposits and orders. This area separates the DIs' armored car personnel from the Federal Reserve employees. Security cameras cover every angle. From the receiving area, currency deposits are transferred to the vault until ready to be processed on the high-speed currency sorters. These machines can sort tens of thousands of notes per hour. Fit notes are sorted into one bin, counterfeit notes are sorted into another, while unfit notes are automatically shredded and discarded.

The interaction of processing scale economies and shipping costs is the key component of our model. Economists define minimum efficient scale (MES) as the lowest level of output at which average cost reaches its minimum. Processing costs would be minimized if all sites operated at this level of output. To make our model computationally feasible, we employ a piecewise linear approximation to the translog cost function estimated by Bauer, Bohn, and Hancock (1998). Both the translog and piecewise linear average cost functions are plotted in Figure 1 in order to demonstrate that very little information about the estimated average cost function is lost in the piecewise approximation.[8] Note that MES is achieved at 218 million items per quarter.

If geography—and hence shipping costs—were not an important factor, then one could determine an upper bound on the number of sites by dividing system-processing volume by MES. If shipping costs are low or if MES is achieved at a low level of output (i.e., full scale economies are achieved at a low level of output), then this approach would still be roughly correct. However, when MES occurs at a relatively large level of output compared to total system processing volume, then the number of processing sites will depend on both the degree of scale economies and how expensive it is to ship currency.

We were fortunate to obtain estimates of shipping costs from a recent Federal Reserve Currency Infrastructure study.[9] Because transportation costs are so crucial to our results—and can only be imperfectly estimated—we solve our model using these estimates and then two more times, with transportation costs alternatively 10 percent higher and then 10 percent lower than these best estimates.

[8] Bauer, Bohn, and Hancock (1998) actually estimate a cost function with three outputs, fit notes, unfit notes, and shipments, but for tractability we hold the ratio of these three outputs constant across processing sites. This means that we only have to track processed notes (fit + unfit notes).

[9] See Report of the Cash Infrastructure Task Force (1998), forthcoming.

3. Model Description

In this section we develop an integer linear programming model to determine the allocation of cash processing volumes and shipment schedules among processing sites that minimizes a total measure of processing and shipping costs of the Federal Reserve. The model is then used to explore the tradeoff between scale economies in processing costs on the one hand and transportation costs on the other.

We assume that all the locations considered can be used for cash processing (for some of the MTA cities, this would mean that processing facilities would have to be constructed). We then allow the model to determine whether some or all the unprocessed cash collected in one site should be shipped to other sites for processing. We also allow for fit cash at a site to be shipped to other sites for distribution to the local depository institutions. The main assumptions adopted for the modeling are the following:

(1) Currency bills are differentiated according to their denomination, in terms of shipping costs as well as the proportion destroyed during currency processing. The model presented in this section can generally be applied for an arbitrary type of bill category. However, in the numerical computations of the Section 4 we only use two types, namely bills of low ($1 bills) and high (all other denominations) value.

(2) The costs associated with shipping cash (fit or unprocessed) are proportional to the volume of shipped cash. The unit transportation cost between two sites depends on the sites chosen, as well as the type of bill shipped. The unit costs of shipping fit and unprocessed currency are the same.[10] Finally, the cost of shipping new cash to a site is directly proportional to the volume of new cash shipped, with the unit shipping costs dependent on the destination site.

(3) The costs associated with currency processing are a function of the processed volume, independent of the type of bills processed. This function can generally be different for each processing site, although in the current application all sites have identical processing costs. In this our approach takes a decidedly long run perspective. The specific form of the cost function used in this paper is discussed in detail in the cost function subsection.

(4) Finally, we assume that there is no restriction on the amount of cash that can be stored at any site. Again, our approach takes a longer term perspective because some Federal Reserve sites are starting to face some vault capacity constraints and might have to be expanded if processing volumes at those sites were to be significantly increased.

The following notation will be used in the subsequent development.

System Parameters

N = Number of cash processing sites.

b = Number of generic note types.

---

[10] See Cash Infrastructure Task Force Report (1998) for more details.

c_{ijk} = Unit transportation cost for currency notes of type k (processed or unprocessed) shipped from site i to site j, for $i = 1,\ldots,N$, $j = 1,\ldots,N$, k=1,…b. This parameter represents all transportation related costs, including the cost of paying and receiving.

p_{jk} = Unit cost of new cash of type k delivered to site j, $j = 1,\ldots,N$, k=1,…b. This parameter includes all costs due to shipping, printing, paying and receiving of new cash.

d_{ik} = Demand for currency of type k at site i.

s_{ik} = Supply of unprocessed currency of type k at site i.

u_i = Cash processing capacity at site i.

α_{ik} = Proportion of unprocessed cash usable after processing (yield) at site i.

Decision Variables

v_{ik} = Volume of cash of type k to be processed at site i.

v_i = Total volume of cash to be processed at site i.

t_{ijk} = Volume of *unprocessed cash* of type k shipped from site i to site j, $i,j = 1,\ldots,N$, $i \neq j$, k=1,…,b.

m_{ijk} = Volume of *fit cash* of type k shipped from site i to site j (to satisfy site-j demand), $i,j = 1,\ldots,N$, k=1,…,b. Note that for i=j, m_{iik} denotes the amount of cash either processed at site i or sent new to site i , which is not shipped anywhere but instead is used to satisfy the demand at this site.

n_{ik} = Volume of new cash of type k sent to site i. $i = 1,\ldots,N$, k=1,…,b.

The Optimization Model

Let $f_i(v)$ denote the cost of processing cash volume v at site i.

The mathematical optimization model developed below determines the optimal currency volumes v_i, t_{ijk}, m_{ijk}, n_{ik}, to minimize the total processing and transportation cost, subject to capacity and demand satisfaction constraints.

$$\text{Minimize} \quad \sum_{i=1}^{N} f_i(v_i) + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{k=1}^{b} c_{ijk}t_{ijk} + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{k=1}^{b} c_{ijk}m_{ijk} + \sum_{j=1}^{N}\sum_{k=1}^{b} p_{jk}n_{jk}$$

$$\text{subject to} \quad (1) \quad \sum_{j=1}^{N} m_{jik} = d_{ik} \qquad \forall i = 1,\ldots,N,\ k = 1,\ldots,b$$

$$(2) \quad v_{ik} + \sum_{\substack{j=1\\ j\neq i}}^{N} t_{ijk} = s_{ik} + \sum_{\substack{j=1\\ j\neq i}}^{N} t_{jik} \qquad \forall i = 1,\ldots,N,\ k = 1,\ldots,b$$

$$(3) \quad \alpha_{ik}v_{ik} + n_{ik} = \sum_{j=1}^{N} m_{ijk} \qquad \forall i = 1,\ldots,N,\ k = 1,\ldots,b$$

$$(4) \quad v_i = \sum_{k=1}^{b} v_{ik} \qquad \forall i = 1,\ldots,N$$

$$(5) \quad 0 \le v_i \le u_i \qquad \forall i = 1,\ldots,N$$

$$m_{ijk}, t_{ijk}, v_{ik}, n_{ik} \ge 0$$

We next explain briefly the motivation for each of the three constraint groups.

(1) *Demand Constraints.* These constraints express the requirement that cash demand at each site must be satisfied. Note that the left hand side is equal to the total processed cash volume available in site *i*, including cash sent from other sites and cash remaining in the site after processing.

(2) *Unprocessed Cash Balance.* These constraints ensure that the amount of outgoing unprocessed cash from site *i* (either shipped to other sites before processing, t_{ij}, $j\neq i$, or remaining in the site for local processing, v_i) is equal to the incoming unprocessed cash to the site (shipped from private banks, s_i, or from other sites t_{ji}, $j\neq i$).

(3) *Fit Cash Balance.* These constraints ensure that a similar balance is satisfied for the incoming and outgoing fit cash at each site.

(4) The total volume of cash processed in a site is comprised of the volumes of different types.

(5) Processing capacity constraints.

The model above is a nonlinear programming problem, because the processing cost functions $f_i(v)$ are generally nonlinear. The method discussed next can be used to transform the original model into a mixed-integer linear programming problem, for the case where the processing costs can be approximated by piecewise linear functions.

The Processing Cost Function

The model can be made computationally tractable by assuming that the processing cost function $f_i(v)$ for cash processed at site *i* is a piecewise linear function of *v* , i.e.,

$$f_i(v) = f_{ij} + \theta_{ij}(v - w_{ij}), \text{ for } w_{ij} < v \le w_{i,j+1},\ j = 1,\ldots,r_i - 1, \tag{1}$$

where $w_{i1} = 0$, $w_{ir_i} = u_i$, and $f_i(0)$=0. According to this definition, the range [0, u_i] of possible processing volumes at site i can be divided into r_i-1 consecutive subintervals with endpoints w_{ij}, j=1,…r_i, such that the processing cost is linear with slope equal to θ_{ij} within the subinterval [w_{ij}, $w_{i,j+1}$]. In addition, since the cost is function is continuous in v for all v>0, it follows that $f_{ij} = f_i(w_{ij})$, therefore equation (1) is equivalent to

$$f_i(v) = \theta_{i1} w_{i1} + \sum_{l=2}^{j-1} \theta_{il} (w_{i,l+1} - w_{il}) + \theta_{ij} (v - w_{ij}), \text{ for } w_{ij} < v \le w_{i,j+1},\ j = 1, \ldots, r_i - 1. \quad (2)$$

The slopes θ_{ij} of the cost function in consecutive segments are increasing in j for any i, therefore, the cost is convex for v>0.

The functional form of the processing costs proposed in equations (1) and (2) is consistent with the empirical findings about economies of scale in cash processing[11]. By setting f_{i1}>0, the cost function becomes discontinuous at v=0, with f_{i0} representing the fixed cost of the cash processing operation at the site. The fixed cost is responsible for the scale economies in cash processing in this formulation. Indeed, the average processing cost per unit of cash volume as a function of v, corresponding to the cost function in (1), has a form similar to that in Figure 1, which was based on actual cost data. The initial decreasing segment of the average cost, which represents the scale economies, is due to the discontinuity jump at 0. On the other hand, the increasing segment, which represents scale diseconomies for larger volumes, is recovered in the theoretical model by the increasing slopes θ_{ij}. For these reasons, the processing cost function defined in (1) is very realistic given the actual cost data. In addition, the piecewise linear approximation can be made arbitrarily accurate by increasing the number of subintervals r_i.

The piecewise linear approximation model for $f_i(v)$ allows us to reformulate the original nonlinear programming problem using mixed-integer linear programming. To see this, note that the piecewise linear cost can be equivalently defined by the sequence of points $\{(w_{ij}, f_{ij}), j=1,\ldots,r_i\}$. Therefore, for any $v \in (0,u_i]$, $f_i(v)$ can be expressed as the solution to the following linear programming problem:

$$\begin{aligned} f_i(v) = \min \quad & \theta_{i1} z_1 + \sum_{l=2}^{r_i - 1} \theta_{il} (z_l - z_{l-1}) & \\ & z_l \le w_{i,l+1} & l = 1, \ldots, r_i - 1 \\ & 0 \le z_1 \le z_2 \le \ldots \le z_{r_i - 1} = v & \end{aligned}$$

Indeed, since θ_{ij} is increasing in j, it can be shown that if v is in the j^{th} subinterval, i.e., $w_{ij} < v \le w_{i,j+1}$ for some j, then the optimal solution to the above problem is $z_l = w_{i,l+1}, l = 1, \ldots, j-1$ and $z_j = z_{j+1} = \ldots = z_r = v$, with objective function value precisely equal to the processing cost as defined in equation (2).

---

[11] See Figure 1.

The above expression for the processing cost function at each site i must now be incorporated into the original problem of cost minimization. Recall that $f_i(v)$ is not continuous for v=0, with the discontinuity representing the fixed processing costs. To model the fixed cost, define a binary variable δ_i for each site i, such that δ_i=1 if v>0 and δ_i=0 if v=0. Variable δ_i is associated with keeping site i open for cash processing (δ_i=1) or not (δ_i=0).

The resulting mixed-integer linear programming problem is given below:

$$\text{Minimize} \quad \sum_{i=1}^{N} f_{i1}\delta_i + \sum_{i=1}^{N}\theta_{i1}z_{i1} + \sum_{i=1}^{N}\sum_{l=2}^{r_i-1}\theta_{il}(z_{il} - z_{i,l-1})$$

$$+ \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{k=1}^{b} c_{ijk}t_{ijk} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{b} c_{ijk}m_{ijk} + \sum_{j=1}^{N}\sum_{k=1}^{b} p_{jk}n_{jk}$$

$$\text{subject to} \quad \sum_{j=1}^{N} m_{jik} = d_{ik}$$

$$v_{ik} + \sum_{\substack{j=1\\ j\neq i}}^{N} t_{ijk} = s_{ik} + \sum_{\substack{j=1\\ j\neq i}}^{N} t_{jik}$$

$$\alpha_{ik}v_{ik} + n_{ik} = \sum_{j=1}^{N} m_{ijk}$$

$$v_i = \sum_{k=1}^{b} v_{ik}$$

$$v_i \le u_i\delta_i$$

$$z_{il} \le w_{i,l+1}$$

$$0 \le z_{i1} \le z_{i2} \le \mathrm{K} \le z_{i,r_i-1} = v_i$$

$$m_{ijk}, t_{ijk}, v_{ik}, n_{ik} \ge 0,\ \delta_i \in \{0,1\}_i$$

$$\text{Minimize} \quad \sum_{i=1}^{N} f_{i1}\delta_i + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{k=1}^{b} c_{ijk}t_{ijk} + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{k=1}^{b} c_{ijk}m_{ijk} + \sum_{j=1}^{N}\sum_{k=1}^{b} p_{jk}n_{jk}$$

$$\text{subject to} \quad \sum_{j=1}^{N} m_{jik} = d_{ik}$$

$$v_{ik} + \sum_{\substack{j=1\\ j\neq i}}^{N} t_{ijk} = s_{ik} + \sum_{\substack{j=1\\ j\neq i}}^{N} t_{jik}$$

$$\alpha_{ik}v_{ik} + n_{ik} = \sum_{j=1}^{N} m_{ijk}$$

$$v_i = \sum_{k=1}^{b} v_{ik}$$

$$v_i \le u_i\delta_i$$

$$z_{il} \le w_{i,l+1}$$

$$0 \le z_{i1} \le z_{i2} \le \mathrm{K} \le z_{i,r_i-1} = v_i$$

$$m_{ijk}, t_{ijk}, v_{ik}, n_{ik} \ge 0,\ \delta_i \in \{0,1\}_i$$

4. Computational Results

Two sets of data are used for the computations, hereafter referred to as the Current Processing Sites (CPS) and the Major Trading Areas (MTA) data set, respectively. The two sets differ in the number of sites considered as cash processing centers and distribution depots.

In the CPS set, the existing 37 cash processing sites of the Federal Reserve System are included in the set of possible processing locations (see Table 1 for a list of these sites). The results of solving this model will explore whether reallocating volume among the existing Federal Reserve sites can lower overall costs. Alternatively, considering the 46 site MTA set, which include all the 37 CPS sites except Helena, is more of a green field approach. This model explores how currency processing sites would be located if one were to start from scratch.

Recall that in both data sets the currency notes were divided into two types, generic note 1 ($1 bills) and generic note 2 (all other notes). The two generic notes differ in the unit transportation cost as well as in the yield of the cash processing operation. The justification for this breakup is that, because of cash transportation regulations, the insurance cost for shipping $1 bills is significantly lower than for other types of notes. Based on interviews with cash personnel, we assume that a bundle of $1 bills can be shipped at 1/10 the cost of a bundle of generic note 2's.

In addition, the yield of processing unfit cash is lower for $1 bills, because they appear to receive much higher abuse than other types of notes. On the other hand, differences in transportation costs and yields among notes of $5 and higher are not significant enough to justify differentiation.

The data on the demand for fit cash and supply of unfit cash in each of the current 37 processing sites are derived from average quarterly values for the year.[12] For the MTA cities we rely on estimates from the Cash Infrastructure Task Force. The volumes of generic notes 1 and 2 are assumed to represent 1/3 and 2/3, respectively, of both supply and demand. This ratio is based on the average of processing volumes observed across the Federal Reserve. The demand and supply volumes are presented in Tables 1 and 2 respectively, for both the CPS and MTA data sets.[13]

Regarding the processing costs, we use the average cost function presented in Figure 1. Let $g(v)$ denote the average cost per unit volume, represented by the solid curve in Figure 1 . The total processing cost as a function of v can then be expressed as $f(v)=v\, g(v)$ and it is the same for all sites. We employ a piecewise linear approximations to $f(v)$ based on 9 subintervals, with endpoints w_j (in 1,000 notes) and costs f_j (in $), for j=1,…,11 as in Table 3. The dashed curve in Figure 3 represents the average cost corresponding to the approximation of f. It can be seen that the 10-point approximation is extremely accurate in the entire range of processing volumes.

The cost of new cash delivered to site j, is taken p_{ij} =$41 per 1,000 notes, for i=1,2, j = 1,…,N.[14] The yield of the cash processing operations is equal to α_{1j}=60%, α_{2j}=70% for all sites.

Again, the unit shipping costs between any two sites is based on estimates from the Cash Infrastructure Task Force. Transportation costs increase with both volume and distance shipped. They also take into account the difference in costs between the two generic notes.

---

[12] We do not study the possible complications of seasonal fluctuations in the demand for currency across the various locations.

[13] There is a slight discrepancy in the total volumes between the two data sets because the supply and demand for currency at Phoenix is assumed to increase if a processing site is located there.

[14] This includes both the cost of printing the notes and shipping them to the processing site.

We solve the optimization model for three separate assumptions about transportation costs to determine the sensitivity of our results. After solving the model for our best estimates of transportation costs, we also estimated the model with the unit transportation costs uniformly lower (90%) and higher (110%). These cases are referred to as the low cost and the high cost case respectively.

The optimal solutions for the two data sets and the various cost cases are obtained employing CPLEX optimization software. The results are summarized in Table 4 for the 37-site (CPS) and in Table 5 for the 46-site (MTA) data sets, respectively. The tables include information on the overall cost for the three cost cases, as well as a baseline case where no shipments are allowed between processing sites, and the case where shipments are allowed but no sites are permitted to close. The no-shipment case corresponds to the current state of affairs in cash processing and serves as the basis of comparison for estimating cost savings through transportation. The site information in Tables 4 and 5 corresponds only to sites that are closed under at least one cost case. Sites not present in these tables remain open for cash processing in all cases. The tables also contain information on the breakdown between transportation, processing and new cash cost for the various cases. The controllable cost figures refer to the sum of transportation and processing costs, since only these two components of the total cost can be affected by reallocating cash volumes among sites. In contrast, the cost of new cash is not controllable, because the unit cost of new cash is assumed to be the same regardless of the site, and the total amount of new cash required is determined by the demand, supply and yield figures and is independent of possible reallocations. Therefore, for optimization purposes, the cost of new cash can be considered as a sunk cost of the currency operation.

A first observation from Tables 4 and 5 is that transportation and processing costs, as well as cost savings resulting from the transportation option, are very similar between the 37 and the 46-site data sets. Although the subsequent discussion refers to the first case only, it also applies to the second.

A comparison between the reference cost and the no shipments columns in Table 4 reveals that allowing for shipment of cash between processing sites results in total cost savings of approximately $5.2 million per quarter. This corresponds to savings in controllable costs of nearly 18%. With respect to the total cost, including the sunk cost of new cash, the savings are approximately 5%. Specifically, by spending $717,000 in transportation, a more efficient allocation of processing volumes can be achieved, which results in a $6 million savings in processing costs.

It is also evident (by comparing the last two columns of Table 4) that the major part of these savings (approximately $5 million) can be realized merely by allowing for cash shipments between sites without any sites being closed. Adding the possibility of site closings results in additional savings in the order of $200,000. Thus, reallocation of cash volume through cash shipments seems to be the critical factor in improving the system efficiency, whereas closing sites has a much smaller effect.

As the unit shipping costs increases from 90% to 110% of the reference case, both the total transportation cost and total processing cost increase. This is expected, because as transportation costs rise, fewer notes are shipped, resulting in smaller cost savings from exploiting scale economies in currency processing. However, the increase in controllable costs is about 1%, which indicates that the optimal cost is fairly robust with respect to shipping cost variations.

Regarding the behavior of site closings as a function of the unit cost, we can make several observations from Table 4. Consider for example the processing site in Oklahoma City. It is optimal for this site to be closed in the low cost case and open in the other two cases. This is intuitive, as increasing shipping costs tends to lower the volume of cash volume shipped, leading to more sites remaining open. On the other hand, for Salt Lake City, moving from normal to high shipping costs results in closing the site. This observation is counterintuitive if considered in isolation. However, the objective of the model is to minimize the cost across the entire system, and its volume can be more cheaply handled at a combination of other sites under this cost configuration (much of the volume goes to Kansas City).

Figures 2 and 3 present the proportion of the demand for fit cash in each site that is satisfied by cash processed locally, cash processed in other sites, and new cash shipped to the site, for generic notes 1 and 2 respectively. It follows from the graphs that most of the unfit cash shipped volumes correspond to generic note 1, whereas for generic note 2 the volume of notes shipped is minimal. This indicates that the unit shipping costs for generic note 2 is prohibitive compared to potential scale economies from reallocation. On the other hand, the lower shipping costs for generic note 1 makes the transportation option more viable.

Only five sites are closed under all three levels of transportation costs. Four others might also be candidates for closing. Of course, this study does not examine all the factors that must be considered before any sites are closed. For instance, this study takes a long run approach and assumes that all inputs adjust fully to the new processing volumes. This means that currency sorters are reallocated and vault space is constructed if necessary. Adding these additional costs into a present value analysis of the site-closing decision, may reveal that closing, or opening, a particular site is too costly. Also, at any sites that were closed there would be an impact on the cost of providing other Federal Reserve services that would have to be considered. Finally, because it would be costly to reopen a processing site, there is an option value for retaining them.

Lastly, as mentioned above, analogous conclusions can be made from the MTA data set. However, this case does have a number of interesting points. First, the biggest omission appears to be the lack of a processing site in Phoenix, a site that remains open under all three transportation cost scenarios. Second, in a handful of cases a nearby city is preferred to an existing Federal Reserve site. For example, Tampa Bay is preferred to Jacksonville in the MTA solution.

Only two of the 10 added sites, Phoenix and Milwaukee, remain open under all three transportation cost scenarios.[15] Consequently, although there may be some opportunities

[15] Because we employ a translog approximation to the underlying true functional form, our cost function may overstate the level of diseconomies of scale once MES is achieved. If so, Milwaukee's volume would most likely be sent to Chicago for processing.

for additional cost savings, it appears that the founders of the Federal Reserve System did a remarkably good job selecting processing sites shortly after the turn of the century that continue to serve well the currency processing needs of the 1990's.[16]

5. Summary

Our goal was to construct a model that we could use to determine the least-cost configuration of currency processing sites given the tradeoff between processing scale economies and transportation costs. We have several robust results. First, the Federal Reserve may be able to save up to about 20 percent of controllable costs by reallocating processing volumes. Second, most of these cost savings can be achieved without closing any processing sites by allowing mostly low denomination bills to be processed at sites currently below MES. Another important result is the finding that only a few processing sites appear to be candidates for closing. Finally, even when we adopt a "green field" approach and search for the optimal allocation of processing volume among the 46 MTAs, the current Federal Reserve sites generally remain open. Intriguingly, Phoenix and Milwaukee are the only non-Federal Reserve sites that remains open under all three shipping cost assumptions, justifying Phoenix's current consideration for some type of improved currency service.

A few caveats should be kept in mind. Some of these shortcomings can be surmounted in our future efforts, but others can not. First, because we did not consider the transition costs involved in closing processing sites, our results can only suggest situations that require additional study. Before actually closing a processing site, a cost-benefit analysis should be performed.

Also, short of actually performing the contemplated shipments, it is unlikely that our estimates of shipping costs will be improved. We tested the sensitivity of our results to this potential weakness by solving the model with transportation costs 10 percent higher and lower than our best estimates. If shipping costs rise substantially, then no currency will be shipped. At the other extreme, inexpensive transportation costs would allow processing volume to be reallocated to allow sites that remain open to operate near MES.

Another potential weakness is that the cost function estimated by Bauer, Bohn, and Hancock (1998) used data from a sample that included sites with both old and new high-speed currency sorting machines. When a large enough sample has been collected of the new sorters, which requires the further passage of time, the cost functions will be re-estimated and our models could be re-optimized.

On a related issue, the cost functions estimated by Bauer, Bohn, and Hancock (1998) incorporate site-specific environmental variables and allow for varying levels of cost efficiency. Including these factors would result in cost functions that vary across processing sites, which could affect our results. On the other hand, the sources of these varying productivity levels can be mitigated over time (by encouraging under-performing

---

[16] Of course, cities with currency processing sites have received at least a small boost to their economic vitality in that depository institutions located there would incur lower costs in shipping currency to and from their branches and the currency depot.

sites to adopt best-practice techniques) and our model is a “long run” model at heart, so an argument can be made for omitting potentially transient factors. However, one short run factor that we may want to include in future work is vault capacity, which varies significantly across processing sites.

Yet another weakness is that the introduction of new notes is somewhat crude. The Bureau of Engraving and Printing’s (BEP) average printing cost per thousand notes is known. The total cost of shipping new notes to Federal Reserve processing sites is also known, but not the specific costs of shipping to a particular site. In our model, the cost of new notes is just the sum of the average cost of printing new notes plus the average cost per note of shipping currency (BEP to Federal Reserve shipping costs/number of new notes). Allowing for differential shipping costs from the BEP to the various processing sites would give sites closer to the BEP sites in Washington, DC and Fort Worth, TX an advantage over ones located further away. We determined that refining this aspect of the model was not a high priority at this time.

Finally, while most of the cost saving can be obtained by reallocating volumes without closing any sites, a different objective function (for example one that specified either higher or lower service levels) might suggest more fundamental changes. Solving our model with this alternative behavioral objective could be employed to provide additional insight into such important policy questions.

References

Bauer, Paul, James Bohn, and Diana Hancock (1998), "Scale Economy, Technological Change and Cost Efficiency of Federal Reserve Currency Processing," working paper.

Cash Infrastructure Study (1998).

Table 1: Demand for Fit Cash (in 1,000 notes)					
37 Site Case (CPS)			46 site case (MTA)		
Site	**Note 1**	**Note 2**	**Site**	**Note 1**	**Note 2**
Atlanta	49722	100952	Atlanta	59299	118597
Baltimore	60628	123094	Baltimore-Washington	69977	139955
Birmingham	20322	41260	Birmingham	18938	37876
Boston	143476	291300	Boston-Providence	110942	221884
Buffalo	39483	80163	Buffalo-Rochester	41118	82236
Charlotte	89775	182271	Charlotte-Greensboro-Raleigh	87348	174695
Chicago	176208	357757	Chicago	101817	203633
Cincinnati	36244	73586	Cincinnati-Dayton	35337	70674
Cleveland	35025	71112	Cleveland	23056	46111
			Columbus	13537	27075
Dallas	43006	87316	Dallas-Fort Worth	54031	108061
Denver	33183	67372	Denver	26122	52244
			Des Moines-Quad Cities	13032	26065
Detroit	63652	129232	Detroit	71285	142569
El Paso	11072	22480	El Paso-Albuquerque	16588	33175
Helena	3830	7776			
Houston	33771	68566	Houston	35299	70599
			Indianapolis	21014	42028
Jacksonville	36360	73823	Jacksonville	10702	21404
Kansas City	16690	33886	Kansas City	12821	25641
			Knoxville	9200	18399
Little Rock	16054	32594	Little Rock	12942	25884
Los Angeles	196467	398887	Los Angeles-San Diego	165664	331328
Louisville	18028	36603	Louisville-Lexington-Evansville	22388	44776
Memphis	17510	35551	Memphis-Jackson	21627	43255
Miami	40520	82268	Miami-Fort Lauderdale	40929	81859
			Milwaukee	38501	77002
Minneapolis	39139	79464	Minneapolis-St Paul	35137	70274
Nashville	19129	38837	Nashville	9865	19730
New Orleans	37930	77009	New Orleans-Baton Rouge	31729	63458
New York	293172	595227	New York	365755	731511
Oklahoma City	18861	38293	Oklahoma City	11337	22674
Omaha	9089	18453	Omaha	7265	14529
Philadelphia	83083	168683	Philadelphia	66213	132427
			Phoenix	28303	56605
Pittsburgh	32025	65021	Pittsburgh	28475	56949
Portland	16379	33255	Portland	14476	28952
Richmond	50720	102976	Richmond-Norfolk	34777	69554
St. Louis	25760	52300	St. Louis	26865	53730
Salt Lake	12139	24646	Salt Lake City	12295	24591
San Antonio	25949	52685	San Antonio	19891	39782
San Francisco	103361	209854	San Francisco-Oakland-San Jose	95631	191263
Seattle	35207	71480	Seattle	31097	62194
			Spokane-Billings	9248	18495
			Tampa - St Petersburg - Orlando	34322	68643
			Tulsa	6664	13328
			Wichita	5735	11469
Totals	**1982969**	**4026028**		**2002859**	**4005714**

Table 2: Supply of Unfit Cash (in 1,000 notes)					
37 Site Case (CPS)			**46 site case (MTA)**		
Site	**Note 1**	**Note 2**	**Site**	**Note 1**	**Note 2**
Atlanta	46522	94453	Atlanta	55481	110962
Baltimore	58671	119120	Baltimore-Washington	67718	135436
Birmingham	17181	34882	Birmingham	16011	32021
Boston	132935	269899	Boston-Providence	102791	205583
Buffalo	38908	78995	Buffalo-Rochester	40519	81039
Charlotte	84448	171456	Charlotte-Greensboro-Raleigh	82165	164330
Chicago	162057	329024	Chicago	93639	187279
Cincinnati	29001	58881	Cincinnati-Dayton	28276	56551
Cleveland	37313	75757	Cleveland	24561	49123
			Columbus	12222	24443
Dallas	37696	76534	Dallas-Fort Worth	47359	94717
Denver	32445	65873	Denver	25541	51081
			Des Moines-Quad Cities	11156	22312
Detroit	53815	109261	Detroit	60268	120537
El Paso	14484	29407	El Paso-Albuquerque	21698	43397
Helena	3888	7894			
Houston	30770	62473	Houston	32162	64324
			Indianapolis	17995	35990
Jacksonville	41621	84503	Jacksonville	12250	24501
Kansas City	14855	30160	Kansas City	11411	22822
			Knoxville	9705	19411
Little Rock	15731	31938	Little Rock	12682	25363
Los Angeles	228842	464618	Los Angeles-San Diego	192963	385927
Louisville	16649	33802	Louisville-Lexington-Evansville	20675	41349
Memphis	17203	34927	Memphis-Jackson	21247	42495
Miami	58477	118727	Miami-Fort Lauderdale	59068	118136
			Milwaukee	35411	70823
Minneapolis	37833	76813	Minneapolis-St Paul	33964	67929
Nashville	20182	40975	Nashville	10408	20816
New Orleans	40774	82784	New Orleans-Baton Rouge	34109	68217
New York	243636	494655	New York	303956	607911
Oklahoma City	17975	36496	Oklahoma City	10805	21610
Omaha	7781	15799	Omaha	6220	12439
Philadelphia	88657	180001	Philadelphia	70656	141312
			Phoenix	37030	74059
Pittsburgh	27342	55512	Pittsburgh	24310	48621
Portland	14425	29287	Portland	12749	25497
Richmond	48119	97697	Richmond-Norfolk	32994	65988
St. Louis	24514	49770	St. Louis	25565	51131
Salt Lake	11274	22890	Salt Lake City	11419	22839
San Antonio	34621	70290	San Antonio	26538	53076
San Francisco	105750	214706	San Francisco-Oakland-San Jose	97842	195685
Seattle	33015	67030	Seattle	29161	58321
			Spokane-Billings	9389	18777
			Tampa - St Petersburg - Orlando	45729	91458
			Tulsa	6357	12714
			Wichita	5319	10637
Totals	**1929408**	**3917284**		**1949494**	**3898989**

Table 3: Endpoints of Cost Function Approximation

Sub-Interval Number j	Sub-Interval Starting Point w_j (1,000 notes)	Cost f_j ($)
1	0	100,166
2	8,913	125,687
3	28,000	183,207
4	48,000	245,717
5	98,000	406,406
6	158,000	612,088
7	218,000	833,646
8	318,000	1,238,865
9	418,000	1,688,538
10	788,000	3,726,716

Table 4: Optimal Solution Summary for CPS data set ($)					
Site	**Low Cost (90%)**	**Reference Cost (100%)**	**High Cost (110%)**	**All Sites Open**	**No Cash Shipments**
El Paso	OPEN	CLOSED	OPEN		
Helena	CLOSED	CLOSED	CLOSED		
Kansas City	OPEN	CLOSED	OPEN		
Little Rock	CLOSED	CLOSED	CLOSED		
Louisville	CLOSED	CLOSED	CLOSED		
Oklahoma City	CLOSED	OPEN	OPEN		
Omaha	CLOSED	CLOSED	CLOSED		
Portland	CLOSED	CLOSED	CLOSED		
Salt Lake	CLOSED	OPEN	CLOSED		
Generic Note 1	260,744	282,150	306,728	285,295	0
Generic Note 2	386,972	435,075	410,316	163,822	0
Total Trans. Cost	**647,717**	**717,227**	**717,044**	**449,117**	**0**
Variable Cost	19,825,154	19,826,972	19,795,879	19,563,818	25,070,764
Fixed Cost	3,004,980	3,004,980	3,105,146	3,706,142	3,706,142
Total Proc. Cost	**22,830,134**	**22,831,952**	**22,901,026**	**23,269,960**	**28,776,906**
Controllable Costs	**23,477,851**	**23,549,179**	**23,618,070**	**23,719,077**	**28,776,906**
New Cash Cost	**75,236,786**	**75,236,786**	**75,236,786**	**75,236,786**	**75,236,786**
Total Cost	**98,714,637**	**98,785,965**	**98,854,856**	**98,955,863**	**104,013,692**

Table 5: Optimal Solution Summary for MTA data set ($)

Site	Low Cost (90%)	Reference Cost (100%)	High Cost (110%)	All Sites Open	No Cash Shipments
Birmingham	CLOSED	CLOSED	OPEN		
Columbus	CLOSED	CLOSED	OPEN		
Des Moines-Quad Cities	CLOSED	CLOSED	CLOSED		
Indianapolis	CLOSED	OPEN	OPEN		
Jacksonville	CLOSED	CLOSED	OPEN		
Knoxville	CLOSED	CLOSED	OPEN		
Little Rock	CLOSED	CLOSED	CLOSED		
Nashville	CLOSED	CLOSED	OPEN		
Oklahoma City	CLOSED	CLOSED	CLOSED		
Omaha	CLOSED	CLOSED	CLOSED		
Portland	CLOSED	CLOSED	OPEN		
Salt Lake City	CLOSED	CLOSED	OPEN		
Spokane-Billings	CLOSED	CLOSED	CLOSED		
Tulsa	CLOSED	CLOSED	CLOSED		
Wichita	CLOSED	CLOSED	CLOSED		
Generic Note 1	245,753	255,839	276,592	302,860	0
Generic Note 2	784,430	816,519	586,122	334,037	0
Total Trans. Cost	**1,030,183**	**1,072,358**	**862,714**	**636,897**	**0**
Variable Cost	19,587,647	19,553,963	19,299,133	19,098,687	23,767,247
Fixed Cost	3,105,146	3,205,312	3,903,474	4,607,636	4,607,636
Total Proc. Cost	**22,692,793**	**22,759,275**	**23,202,607**	**23,706,323**	**28,374,883**
Controllable Costs	**23,722,976**	**23,831,633**	**24,065,321**	**24,343,220**	**28,374,883**
New Cash Cost	**76,088,222**	**76,088,222**	**76,088,222**	**76,088,222**	**76,088,222**
Total Cost	**99,811,198**	**99,919,855**	**100,153,543**	**100,431,442**	**104,463,105**

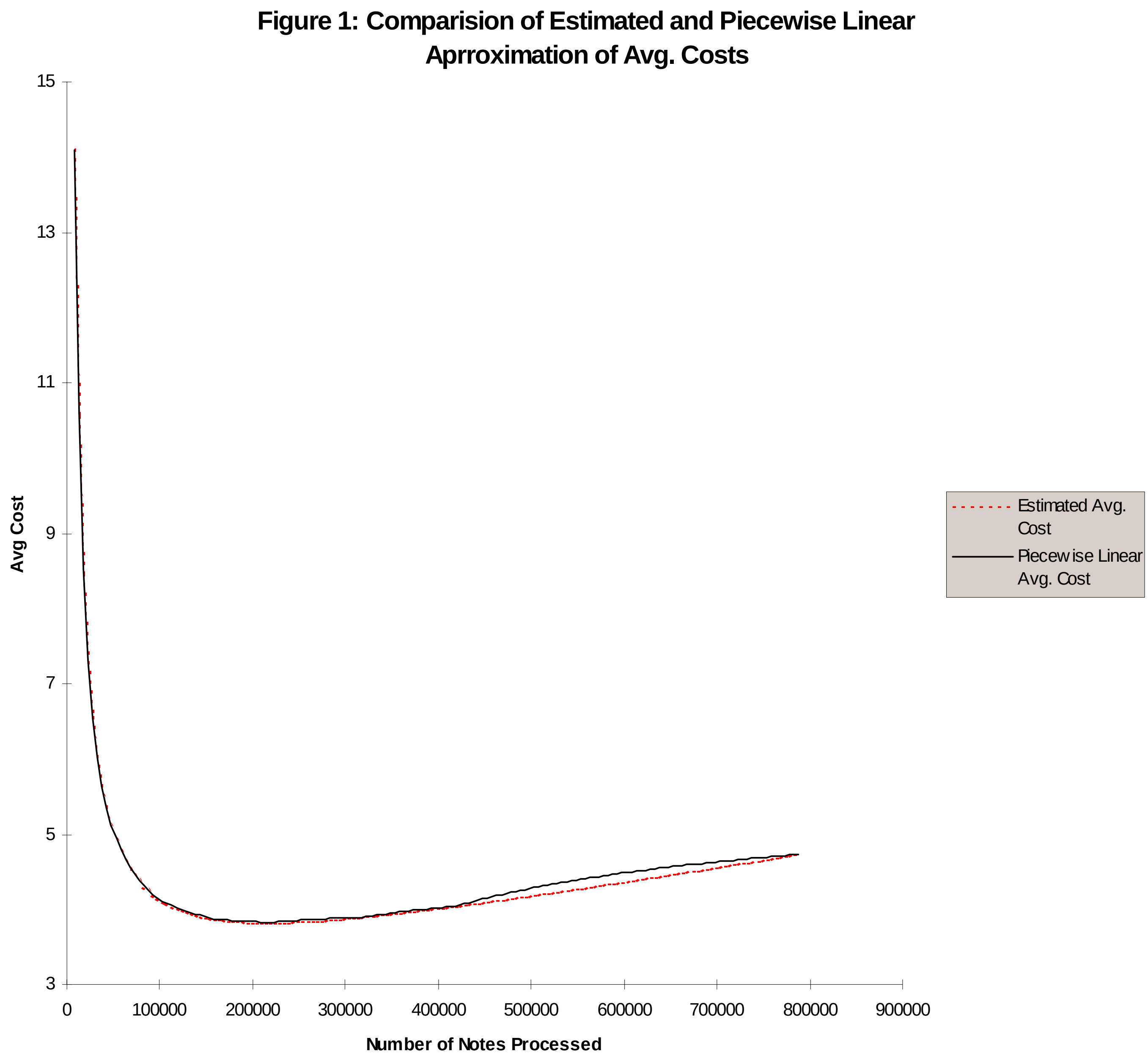
Figure 1: Comparision of Estimated and Piecewise Linear Aprroximation of Avg. Costs
15
13
11
9
7
5
3
Avg Cost
0
100000
200000
300000
400000
500000
600000
700000
800000
900000
Number of Notes Processed
Estimated Avg. Cost
Piecewise Linear Avg. Cost

Figure 2: Proportions of Fit Cash for Generic Note # 1

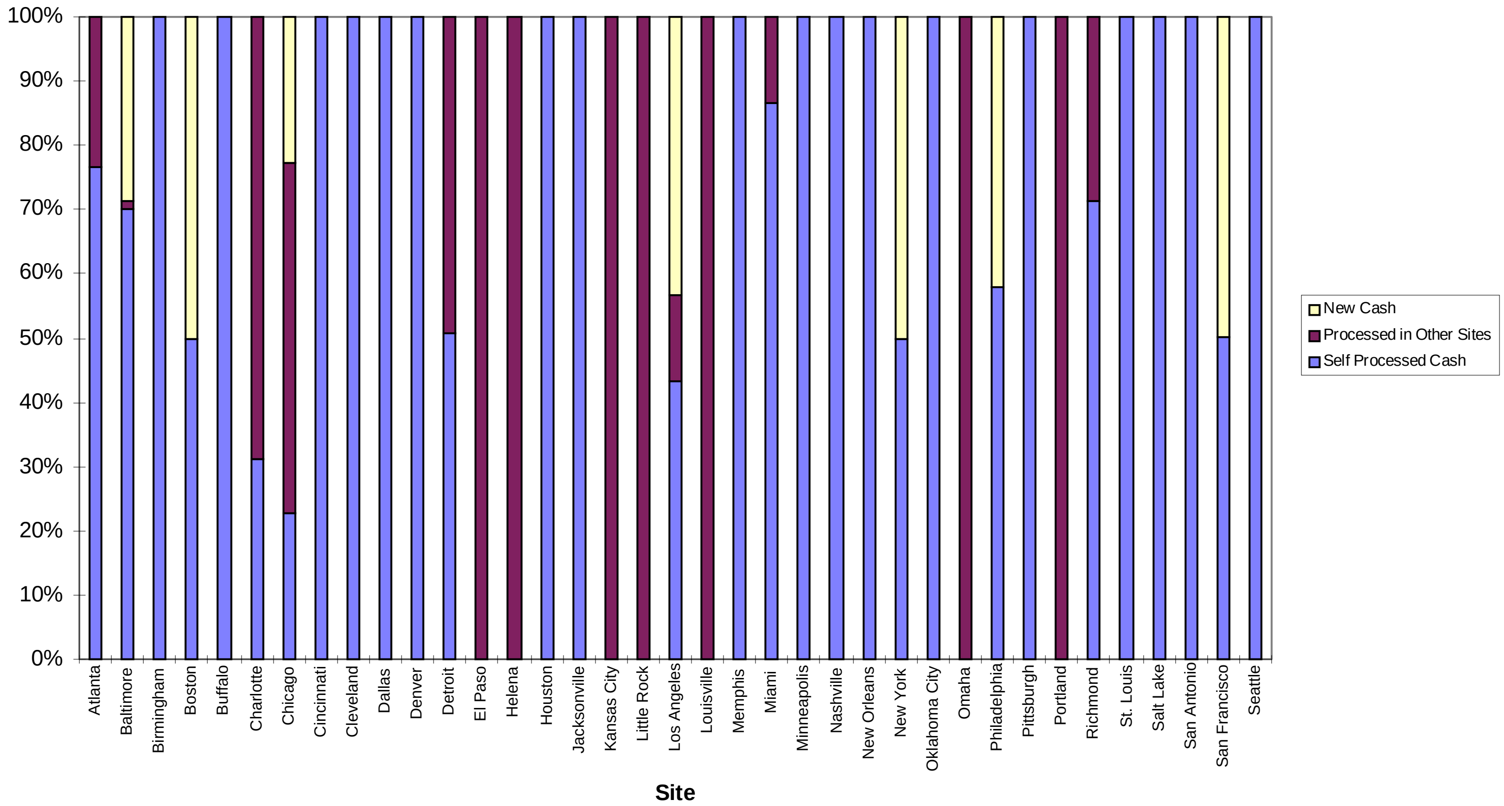

Figure 3: Proportions of Fit Cash for Generic Note # 2