Productivity Assessment of Multiple Retail Outlets

WAGNER A. KAMAKURA
University of Pittsburgh

THOMASZ LENARTOWICZ
University of Texas at Austin

BRAIN T. RATCHFORD
State University of New York at Buffalo

One of the main problems in assessing the relative efficiency of multiple retail outlets is the fact that customers are an essential component of the "production process" in retailing. Hence, the efficiency of a particular retail store depends both on the seller's and the buyers' efficiency. While there is often data on each store's inputs to the production of retail services, the contribution of each store's customers to this process is difficult to measure, and often goes unmeasured. In the absence of data on variation in customer efficiency across stores, it is difficult to make comparisons of the relative efficiency of the retailer's operation across stores. In this paper, we propose to solve the problem of missing data on customer inputs by treating stores as members of different latent classes, where the classes are defined by differences in customer inputs and other market characteristics. We consider the clusterwise estimation of multiple translog cost functions, which identifies sets of retail outlets operating under similar conditions and simultaneously estimates multiple cost functions for these classes. The efficiency of each outlet would then be evaluated relative to others in its class, which allows for a more equitable evaluation of each retail outlet, in comparison to other units operating under similar conditions.

We apply this approach in the evaluation of multiple branches from a commercial bank in Latin America, and compare the efficiency measures obtained from it with measures obtained from other methods, using the bank's central managers' classification of markets as a benchmark. Our results indicate that the clusterwise estimation of translog cost functions leads to a more equitable assessment of the branches, more in tune with the market differences perceived by the bank managers.
INTRODUCTION

This study was motivated by the problem faced by a major commercial bank in Latin America. This large multi-branch bank has major difficulties in evaluating branch performance. All evaluation criteria used in the past encountered strong resistance by branch managers; many of them argue against being compared to other branches that might have a differential advantage due to the type of customers served and differences in the mix of services provided. In response to these complaints, central management classified the branches into homogenous groups, based on a qualitative assessment of the branches and their markets, so that branches could be ranked within each of these homogenous groups. However, this classification did not stop the complaints because the selection of the “comparable” sets was based on subjective criteria and judgment, highly vulnerable to objections. Branch managers also complained that profitability ratios did not reflect the true efficiency of their branch operations because costs incurred in one branch might be generating revenues for other branches, and because the profitability of any particular branch is affected by the actions of central management.

The problems outlined in the above example are typical of the problems encountered in attempts to evaluate the relative performance of retail outlets. While some measure of the relative efficiency or productivity of the various outlets is required for addressing these problems, the usefulness of existing measures of retail productivity is hindered by a number of measurement problems (Achabal, Heineke and McIntyre, 1984). At the root of many of these measurement problems is that, while the customer participates in the production of retail services (Oi, 1992), differences between outlets in the nature and quality of the customer’s input are very difficult to measure, and often overlooked. In the banking case presented above, for example, differences in the literacy of customers between rich and poor areas imply that labor and other resource requirements will differ between these areas. Because of these differences, the production or cost function of a bank will differ between these areas, and this needs to be taken into account in assessing the relative efficiency of the bank’s outlets.

In this paper we propose a solution to this problem of missing data on customer inputs. Our proposal is to employ a clusterwise cost function estimation approach, which allows for different production frontiers, and for outlets to be members of latent classes defined by these frontiers. Efficiency comparisons are then made relative to other members of a given class. The different production frontiers result from the outlets servicing different customer populations, which give rise to differences in the services provided by them. The latent classes arise because the characteristics of different customer populations and services are not fully captured by the data. While the particular latent class estimation procedure which we employ has been used in other contexts (Wedel and Steenkamp, 1991), we believe that the application to productivity comparisons is novel, and that the identification of multiple efficiency frontiers leads to a more valid assessment of efficiency.

To validate our proposed procedure, we employ data on the branch banks mentioned above. We compare the results that our procedure provides with results obtained from two standard procedures for assessing relative efficiency, data envelopment analysis and the standard translog cost function. We show that our proposed clusterwise procedure gives
results which are generally congruent with an independent grouping of the banks provided by the bank’s managers, while the two other methods do not. We take this as evidence for the validity of our procedure.

Thus, the main objective of this paper is to explain and assess the clusterwise method for comparing the productivity of different outlets. To provide a setting for our discussion, the next section of this paper reviews the literature on retail productivity, and on methodologies for efficiency measurement. After the literature review, we describe the proposed clusterwise approach. We then present our empirical study on the bank data which compares the validity of the proposed approach against the competing methods. The paper closes with a set of conclusions and directions for further research.

PREVIOUS RESEARCH

Retail productivity is an important issue and a vast literature may be found on its definitions and measurements. A review of this literature shows that multiple methodologies have been applied to assess productivity of individual retail stores, groups of stores, and the retail industry, but that surprisingly little attention has been given to comparing the efficiency of retail outlets.

The most widely used conceptualization of productivity has been as the ratio of outputs to inputs; total input productivity is defined as the ratio of all outputs to all inputs, and partial or single input productivity is the ratio of all outputs to a single input (Ingene, 1982, Lusch and Moon, 1984). Good (1984) provides a list of possible measures of retail outputs and inputs. Outputs are usually measured as the number of transactions, physical units sold, value added, and sales. Inputs are measured as the hours of labor employed, number of employees, wages, salaries and benefits paid, area of selling place, inventory, and advertising.

Though simple to define, assessments of retail productivity based on simple ratios of outputs to inputs have been criticized for the following reasons: improper measurement of output (Achabal et al., 1984; Parsons, 1994, Oi, 1992), failure to account for changes in the quality of inputs or outputs over time or across stores (Doult, 1984; Good, 1984; Lusch and Moon, 1984; Nooteboom, 1985, Oi, 1992); failure to account for the consumer’s input to the process (Ingene, 1984; Oi, 1992); improper weighting of multiple inputs and outputs (Parsons, 1992), inability to separate differences in productivity from scale effects (Ratchford and Brown, 1985). In addition to these limitations, the traditional “ratio” approach to retail productivity presents other problems when the focus is the evaluation of the multiple outlets from a single retailer. These multiple outlets are typically located in different markets and serve a diverse population of customers, leading to distinct operational characteristics at each outlet. These differences are not taken into account by traditional productivity indices, leading to a biased assessment of the relative efficiency of multiple outlets.

Surprisingly, studies on the assessment of multiple outlets are quite rare in the marketing literature (see Ingene, 1982 for a review of the marketing literature on productivity). The main focus of previous research on retail productivity has been on the measurement and improvement of performance of an industry (Ingene, 1984; Lusch and Moon, 1984), a firm
as a whole (Weitzel, Schwartzkopf and Peach, 1989), or groups of firms (Good, 1984; Doutt, 1984; Stassen, Kelkar and Schulman, 1992). Some studies present important findings on which elements may improve the performance of store managers (Lusch and Jaworski, 1991), or about the factors affecting store sales performance (Weitzel et al., 1989), but do not address the question of how to properly evaluate the relative efficiency of multiple outlets. This scarcity of studies on the assessment of multiple outlets is not compatible with the importance of the topic, in view of its managerial implications: branch managers’ promotions and raises are often tied to company-wide competition and corporate offices must know how store managers are rated relative to other stores (Lusch and Dunne, 1990).

In contrast, the retail banking industry has shown strong interest in the productivity of multiple-outlets. U.S. banking deregulation regarding multi-branch banks and geographic constraints created new growth opportunities (Rose, 1987, Berger, Hanweek and Humphrey, 1987; Chelst, Schultz and Sanghvi, 1988). Consequently, many studies have compared the productivity of single versus multi-unit banks (Aly, Grabowska, Pasturka and Rangan, 1990; Rangan, Zardkoohi, Kolari and Fraser, 1989; Cebanoyan, 1988), and examined the characteristics leading to higher productivity. Some of the studies on how to evaluate and to improve the productivity of bank branches use the traditional ratio approach (Arvey, 1989), but the tendency is to use newer techniques that avoid the problems we had reviewed earlier. These newer methodologies for productivity assessment will be explained and evaluated next.

- RECENT APPROACHES TO PRODUCTIVITY ASSESSMENT

In this section, we provide a brief description of two recently developed approaches for productivity assessment, Data Envelopment Analysis (DEA) and translog cost functions. These will be used for a comparative assessment of our proposed approach.

Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a mathematical programming approach originally proposed by Charnes, Cooper and Rhodes (1978) to measure the relative efficiency of decision making units (DMU) producing multiple outputs from multiple inputs. In DEA, efficiency is measured by comparing the inputs of a decision making unit (e.g., a store) relative to the inputs needed by the most efficient unit (store) to produce an equivalent output: an efficiency ratio of 1 means that the store has maximum efficiency, ratios below this indicate how inefficient the unit is relative to the most efficient one. The mathematical programming procedure in DEA compares the outputs and inputs of several decision making units, and identifies the “most efficient” set of DMU’s to which any particular unit should be compared, based on the similarity of their input and output-levels. A virtual “efficient DMU” is constructed as a convex combination of the most efficient set, so that the relative
efficiency of the DMU in question can be calculated. A more detailed description of the structure of the DEA model is presented in Appendix A.

DEA has been extensively applied in a wide range of contexts, such as the evaluation of insurance companies (Mahajan, 1991), school districts (Bessent, Bessent, Kennington and Reagan, 1982), bank branches (Sherman and Gold, 1985; Ferrier and Lovell, 1990; Oral and Yolalan, 1990; Vassiloglou and Giokas, 1990; Aly et al., 1990) and university departments (Beasley, 1990). Extensive reviews of this approach can be found in Boussofiane, Dyson and Thanassoulis (1991) and Norman and Stoker (1990), among many others. A recent survey of the literature (Seiford, 1990) identified some 400 published studies in this area.

One of the most valuable features of DEA is the flexibility of its piecewise-linear frontier, which allows the evaluation of any decision making unit relative to a facet formed by other units operating with similar combinations of inputs and outputs. On the other hand, the deterministic nature of this linear-programming approach capitalizes on measurement errors in the inputs and outputs, and may make the evaluation sensitive to outliers. A single DMU operating at unusual circumstances may render all other DMU's inefficient or efficient.

**Translog Cost Function Estimation**

Another common approach to the assessment of productivity and production efficiency is the econometric modeling of the dual of the production frontier, for example, the indirect cost function for multi-product multi-input technologies. The most popular model for the estimation of this stochastic frontier is the translog cost function (Christensen and Greene, 1976; Caves, Christensen and Swanson, 1981). The logarithm of the total cost at outlet b is expressed as:

\[
\ln C_b = \alpha_o + \sum_j \alpha_j \ln y_{ib} + \sum_j \beta_j \ln w_{jb} + \sum_k \phi_k \ln z_{kb} + 1/2 \sum_i \sum_j \alpha_{ij} \ln y_{ib} \ln y_{ib} + 1/2 \sum_j \sum_j \beta_{jj} \ln w_{jb} \ln w_{jb} + \sum_i \sum_j \delta_{ij} \ln y_{ib} \ln w_{jb} + T_b + \epsilon_b 
\]

(1)

where:

- \(y_{ib}\) = units of output \(i\) produced by outlet (or branch) \(b\);
- \(z_{kb}\) = units of allocative input \(k\), not under direct control by the store manager, used by outlet \(b\);
- \(w_{jb}\) = unit price of input \(j\) at outlet (branch) \(b\);
- \(T_b\) = the addition to log-cost due to technical inefficiency of outlet \(b\), \((\Sigma_o T_b = 0)\);
- \(\epsilon_b\) = a normally-distributed random disturbance.

For multiple-input technologies, estimation is improved by simultaneously considering the following cost-share equations which result if cost-minimizing levels of each input are chosen (Caves, Christensen and Swanson, 1980):
\[ CS_{jb} = \beta_j + \sum_j \beta_{jj} \ln w_{jb} + \sum_j \delta_{jj} \ln y_{jb} + \mu_{jb} \] 

(1a)

where \( CS_{jb} \) is the cost share of input \( j \), out of the total costs for outlet \( b \). The translog function can be shown to be a second order Taylor series approximation to any arbitrary function which has desirable global properties (Caves and Christensen, 1980).

If estimated with standard econometric methods, the translog cost function in Equation 1 does not represent the "minimum" cost frontier. Rather, this function provides the "mean" or expected cost for any combination of input prices and output volumes. Therefore, \( T_b \) must be viewed as the technical inefficiency (in log-cost) relative to the "average" outlet operating at the same scale level. This view is compatible with our particular purpose of evaluating multiple retail outlets relative to each other. However, if one is interested in measuring technical inefficiency relative to the minimum cost frontier (see Ferrier and Lovell, 1990), \( T_b \) must be constrained to non-negative values, and assumed to be distributed across outlets as a truncated normal (Stevenson, 1980) or a Gamma (Greene, 1990).

An application of the translog cost function in the study of retail productivity can be found in Ratchford and Stoops (1988). A comprehensive review of the applications of the translog cost function in the banking industry can be found in Gilbert (1984) A few studies have also focused on comparing DEA and the translog approaches (Banker, Conrad and Strauss, 1986, Ferrier and Lovell, 1990).

Stochastic frontier estimation through the translog cost function alleviates the major problems of DEA (i.e., sensitivity to outliers and to measurement and specification errors in the inputs and outputs), because it allows for statistical error resulting from factors outside the firm’s control (Bauer, 1990). However, the translog function imposes a parametric structure on the production technology and on the distribution of efficiency (Ferrier and Lovell, 1990), implying that all outlets must use the same translog technology and face the same long-run cost function. This assumption of a common production technology restricts the usefulness of this approach for the assessment of multiple retail outlets, especially when one expects substantial differences in consumer knowledge, efficiency and shopping behavior across outlets.

**CLUSTERWISE TRANSLOG COST FUNCTION ESTIMATION**

Our proposed clusterwise approach for productivity assessment of retail outlets has the same stochastic nature of the translog cost function estimation, but allows for multiple translog technologies. In a sense, the clusterwise translog cost function approach represents a compromise between the flexible piecewise-linear deterministic frontier in DEA and the stochastic estimation of a single translog cost function. Rather than stipulating a single translog technology, we allow for multiple cost functions, so that a group of outlets will be assessed relative to a common cost function distinct from the ones used by other groups. Furthermore, these groups or clusters of outlets are identified at the same time their common translog cost function is estimated, so that they represent homogeneous groups in terms of their production function. A similar approach, based on finite-mixture theory, was
Productivity Assessment

proposed by Beard, Caudill and Gropper (1991) to identify two production technologies in the savings and loan industry.

This clusterwise approach is especially suitable for the assessment of retail outlets, where comparable groups are crucial to make the evaluation of each outlet more equitable. As pointed out by Parsons (1994), "the technology of a marketing activity typically depends on external factors that are unlikely to be constant over time and place."

Estimation of multiple translog functions for various clusters of retail outlets is possible when longitudinal data are available on the outputs and inputs for each outlet. For a branch b, the translog cost function at any given time period t is defined as:

$$\ln C_{bt} = \sum_g u_{bg} \ln C_{bgt}$$

where:

$$u_{bg} = \text{fuzzy weights assigning branch } b \text{ to each group } g = 1, \ldots, G, \text{ so that } \sum_g u_{bg} = 1, \text{ and}$$

$$\ln C_{bgt} = \alpha_{og} + \sum_i \alpha_{ig} \ln y_{ibt} + \sum_j \beta_{jg} \ln w_{jbt} + \sum_k \phi_{kg} \ln x_{kt} + 1/2 \sum_i \sum_j \alpha_{ijg} \ln y_{ibt} \ln y_{ibt} +$$

$$+ 1/2 \sum_j \sum_j \beta_{jjg} \ln w_{jbt} \ln w_{jbt} + \sum_i \sum_j \delta_{ijg} \ln y_{ibt} \ln w_{jbt} + \epsilon_{bgt}$$

We estimate the translog cost functions defined in Equations 2 and 3 for G fuzzy clusters, using the fuzzy clusterwise regression method developed by Wedel and Steenkamp (1991).\(^2\) As described in detail in their article, this method estimates clusterwise regression equations for a pre-specified number of groups, and simultaneously assigns each outlet b to the overlapping clusters g = 1, \ldots, G, using fuzzy weights \(u_{bg}\) (see Appendix B for a brief description of this procedure). Estimates of the contribution of technical inefficiency to the log-cost of each branch b are obtained as \(T_b = \sum_t \{\ln C_{bt} - \sum_g u_{bg} \ln C_{bgt}\}\). Alternatively, these estimates could also be obtained from other clusterwise regression models (DeSarbo and Cron, 1988; Wedel and Kistemaker, 1989).

Note that in the description of the clusterwise approach a time dimension has been added. In contrast to the other approaches, cross-sectional time-series data is strictly necessary, so that branches can be clustered into homogeneous groups in terms of their cost function.

**EMPIRICAL APPLICATION**

Our evaluation of the proposed approach for making efficiency comparisons was performed on data from 188 branches of a commercial bank within a large metropolitan area in Latin America. High inflation and banking regulations in this area create the need for frequent face-to-face banking transactions among all socioeconomic groups, making banks there more like typical retailers than they may be elsewhere. In this section, we will describe the problem faced by management of the bank, which we believe is typical of situations in which efficiency comparisons between outlets have to be made, and we will present the basic empirical estimates upon which our assessment of the clusterwise proce-
dure is based. The next section will present some tests of the validity of this procedure on
the data described in this section.

Our application is based on an actual case in which performance comparisons among out-
lets were a critical management problem. This particular bank currently uses a system of
performance indices to evaluate branches. This system has been criticized by branch man-
gagers as being unfair because it ignores the differences in the market conditions faced by
each branch. Many branch managers argue against being compared to other branches that
might have a differential advantage due to the type of customers served and differences in
the mix of services provided. Some branch managers, for example, complain that their cus-
tomers are less educated and poorer than in other regions, and thus require more help in fill-
ing out deposit slips, checks, etc. for transactions of smaller volume. Another problem with
using accounting ratios in evaluating individual branches is that transactions in one particu-
lar branch are often processed into accounts from another branch, so that costs incurred in
one branch might be generating revenues for another branch. Furthermore, the profitability
of any particular branch is affected by the actions of central management, who make the
decision on how to apply the funds generated from the various accounts in all branches.

In response to these complaints, central management classified the branches into homoge-
aneous groups, based on a subjective assessment of the market characteristics and type of
customer served, so that branches could be ranked within each of these homogenous
groups. The approaches compared in this study will also seek a comparison of each branch
with “comparable” sets of branches. However, the selection of the “comparable” sets will
be based on the relationships among the inputs and outputs used by each branch, rather than
on external criteria or judgment. We will use the managers’ classification of homogeneous
markets as a criterion for assessing the validity of the efficiency measures.

DESCRIPTION OF THE DATA

A variety of data about inputs and outputs was made available to us for each of the 188
branches of the bank in one major metropolitan area, in a period of 5 months. The unit of
the major part of the data is the local currency. In view of the high rate of inflation, this unit
was transformed to constant currency using a specific inflation index utilized in the country
for financial transactions. Because of the wide range of unit values across services and
branches, we use the monetary unit as the measure of output in each service, as suggested

For inputs, data were available on the total operational costs, wage rate and man-hours
of labor in each branch. However, two problems led to a change in the models to be tested.
First, labor represents more than 85% of all costs under control by the branch manager, and
the other portion of the operational costs is distributed into a wide variety of sources such
as materials, security, janitorial services that cannot be individually priced. Second, wage
rates suffered inordinate and irregular changes during the period of our study, because of
collective agreements to bring wages at par with an inflation rate greater than 25% per
month at the time of this study. For these reasons, we decided to focus on labor productivity
and to measure labor in physical units (man-hours).
The models tested in our illustration also include one fixed input (floor area), which is constant in the medium term, and is not exclusively controlled by the branch manager. We also had information on another fixed input, the total number of teller stations in each branch. However, a preliminary analysis showed that this variable was highly collinear with the floor area in the branches, thus adding little information to it. Hence number of teller stations was dropped from the final study.

The variables used in this study are summarized below:

1. Outputs:
   - **CASH**: volume of cash deposits.
   - **OTHER**: volume of other deposits, such as checks and money orders.
   - **PAYC**: volume of funds “in transit” in the branch, which already have their destination, like tax collections, payment of bills, or pay checks.
   - **MREV**: volume of service fees charged to customers by the branch to pay for their transfers, checkbooks, statements, etc., that may be waved by the manager of a branch to preferred customers.

2. Inputs
   - **LABOR**: total number of man-hours of direct (clerks) labor allocated at the branch.
   - **AREA**: size (in square meters) of the customer services area.

A summary of these data is shown in Table 1, according to the classification of similar markets by the bank's central management.

**TABLE 1**

<table>
<thead>
<tr>
<th></th>
<th>Industrial/Residential (Low)</th>
<th>Residential Commercial (Prime)</th>
<th>Financial District</th>
<th>Suburban Centers</th>
<th>Residential Commercial (Low)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LABOR</strong></td>
<td>4 188</td>
<td>4 210</td>
<td>6 570</td>
<td>6 635</td>
<td>3 704</td>
</tr>
<tr>
<td></td>
<td>(2 153)</td>
<td>(1 486)</td>
<td>(3 535)</td>
<td>(2 576)</td>
<td>(1 259)</td>
</tr>
<tr>
<td><strong>CASH</strong></td>
<td>2 316</td>
<td>2 887</td>
<td>3 949</td>
<td>3 776</td>
<td>2 285</td>
</tr>
<tr>
<td></td>
<td>(1 252)</td>
<td>(1 116)</td>
<td>(1 862)</td>
<td>(1 342)</td>
<td>(0 832)</td>
</tr>
<tr>
<td><strong>OTHER</strong></td>
<td>1 258</td>
<td>1 940</td>
<td>2 507</td>
<td>2 199</td>
<td>1 353</td>
</tr>
<tr>
<td></td>
<td>(0 768)</td>
<td>(0 768)</td>
<td>(1 332)</td>
<td>(0 797)</td>
<td>(0 499)</td>
</tr>
<tr>
<td><strong>PAYC</strong></td>
<td>0 892</td>
<td>0 630</td>
<td>1 892</td>
<td>1 368</td>
<td>0 385</td>
</tr>
<tr>
<td></td>
<td>(0 931)</td>
<td>(0 756)</td>
<td>(2 830)</td>
<td>(0 686)</td>
<td>(0 503)</td>
</tr>
<tr>
<td><strong>MREV</strong></td>
<td>0 866</td>
<td>0 644</td>
<td>1 295</td>
<td>1 332</td>
<td>0 549</td>
</tr>
<tr>
<td></td>
<td>(0 880)</td>
<td>(0 434)</td>
<td>(1 027)</td>
<td>(1 203)</td>
<td>(0 292)</td>
</tr>
<tr>
<td><strong>AREA</strong></td>
<td>656 3</td>
<td>646 9</td>
<td>874 1</td>
<td>988 3</td>
<td>568 8</td>
</tr>
<tr>
<td></td>
<td>(341 0)</td>
<td>(276 0)</td>
<td>(421 3)</td>
<td>(616 2)</td>
<td>(236 6)</td>
</tr>
<tr>
<td><strong>Number of Branches</strong></td>
<td>54</td>
<td>73</td>
<td>12</td>
<td>10</td>
<td>39</td>
</tr>
</tbody>
</table>

*Note: Measurement units for the output variables are disguised.*
A comment on the output measures is in order. The output measures should capture retail services provided to customers. The four measures do capture various aspects of services. CASH and OTHER obviously capture deposit services. Similarly PAYC captures the services provided by the bank to individuals and organizations in the collection and payment of bills, and MREV is related to the volume of specific services provided by the bank for which direct fees are charged. Because the level of service is likely to be related to the number of transactions, it is unfortunate that we did not have a measure of number of transactions or average transaction size. Similarly, while more service per transaction may be required by different customer types, we did not have a measure of this. One must note, however, that the clusterwise cost function estimation procedure provides a way of capturing the impact of these and other omitted service measures on the cost functions of branches operating under similar circumstances.

Clusterwise Translog Cost Function Estimation

Because there is only one input (labor), the wage component of lnC in Equation 1 can be eliminated by eliminating terms in lnw from both sides of the equation.\(^3\) The dependent measure in the translog becomes lnLABOR. Notice also that the share Equations 1a are not needed, because we use only one input. The translog cost function for our example then becomes:

\[
\ln \text{LABOR}_{bt} = \alpha_0 + \alpha_1 \ln \text{CASH}_{bt} + \alpha_2 \ln \text{OTHER}_{bt} + \alpha_3 \ln \text{PAYC}_{bt} + \alpha_4 \ln \text{MREV}_{bt} \\
+ \phi_1 \ln \text{AREA}_{bt} + \sum \delta_t \ln \text{MONTH}_t + 1/2(\alpha_{11} \ln \text{CASH}_{bt} - \ln \text{CASH}_{bt}) \\
+ \alpha_{12} \ln \text{CASH}_{bt} \ln \text{OTHER}_{bt} + \alpha_{13} \ln \text{CASH}_{bt} \ln \text{PAYC}_{bt} \\
+ \alpha_{14} \ln \text{CASH}_{bt} \ln \text{MREV}_{bt} + \alpha_{22} \ln \text{OTHER}_{bt} \ln \text{OTHER}_{bt} \\
+ \alpha_{23} \ln \text{OTHER}_{bt} \ln \text{PAYC}_{bt} + \alpha_{24} \ln \text{OTHER}_{bt} \ln \text{MREV}_{bt} \\
+ \alpha_{33} \ln \text{PAYC}_{bt} \ln \text{PAYC}_{bt} + \alpha_{34} \ln \text{PAYC}_{bt} \ln \text{MREV}_{bt} \\
+ \alpha_{44} \ln \text{MREV}_{bt} \ln \text{MREV}_{bt}) + T_b + \varepsilon_{bt}.
\]

We treat AREA as a fixed input that is constant on the short-run, and is not under the exclusive control by the branch manager. The variables MONTH\(_t\) (\(t = 2, 3, 4, 5\)) are dummy variables to account for seasonal variations, and distortions in the deflation of outputs.

The contribution of technical inefficiency to the log-labor \(T_b\) is computed as described earlier. We use a technical inefficiency ratio relative to the regression mean, \(T_b = \exp(T_b)\). Elasticities of labor requirements for a particular output can be obtained by:

\[
E_{bt} = \frac{\partial \ln \text{LABOR}_{bt}}{\partial \ln y_{bt}},
\]

where \(y_t\) represents the four types of outputs. Following Caves, Christensen and Swanson (1981), the returns to scale RTS\(_b\) for branch \(b\) are given by:

\[
\text{RTS}_{b} = (1 - \phi_1)[\Sigma E_{bt}]^{-1} 
\]
where \( \phi_i \) is the coefficient of \( \ln\text{AREA} \) in Equation 4. A branch operates under increasing, constant or decreasing returns to scale when \( RTS_b \) is greater, equal or less than one, respectively.

Estimates are obtained through a fuzzy-clustering regression procedure (Wedel and Steemkamp, 1991), leading to one translog cost function as in Equation 4 for each latent-class or segment. One of the main difficulties with this procedure, as with other clustering methods, is in selecting the number of clusters. For this decision, we followed the recommendations by Wedel and Steemkamp (1991), and estimated the clusterwise model for up to 7 clusters. The overall fit improved only marginally from 5-clusters (\( R^2 = 0.985 \) JRML = 1.953) to 6-clusters (\( R^2 = 0.989 \) JRML = 1.510), and the 7-cluster solution failed to converge.\(^4\) Furthermore, the 6-cluster solution was similar to the 5-cluster solution, except for an additional cluster that contained only a few branches with large fuzzy weights. Based on these results we chose the five-cluster solution.\(^5\)

The translog functions for each of the 5 fuzzy clusters are presented in Table 2. These functions define five production technologies that might be representative of groups of branches. As indicated in Equation 2, each branch’s cost function is a weighted average of these five functions, where the weights reflect the probability of cluster membership. However, our results indicate a limited amount of overlapping or “fuzziness” in cluster mem-

**Table 2**

Parameter Estimates for the Translog Functions

<table>
<thead>
<tr>
<th>Variables</th>
<th>Aggregate Model</th>
<th>Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.782</td>
<td>37.092</td>
</tr>
<tr>
<td>CASH</td>
<td>0.72</td>
<td>0.14</td>
</tr>
<tr>
<td>OTHER</td>
<td>-0.95*</td>
<td>-5.30*</td>
</tr>
<tr>
<td>PAYC</td>
<td>0.012</td>
<td>0.11</td>
</tr>
<tr>
<td>MREV</td>
<td>-0.70*</td>
<td>-0.60*</td>
</tr>
<tr>
<td>AREA</td>
<td>0.088*</td>
<td>0.099*</td>
</tr>
<tr>
<td>MONTH(_3)</td>
<td>-0.060*</td>
<td>-0.097*</td>
</tr>
<tr>
<td>MONTH(_4)</td>
<td>-0.006</td>
<td>-0.056*</td>
</tr>
<tr>
<td>MONTH(_5)</td>
<td>0.20</td>
<td>0.12</td>
</tr>
<tr>
<td>MONTH(_6)</td>
<td>0.188*</td>
<td>0.11*</td>
</tr>
<tr>
<td>cash(*)</td>
<td>0.005</td>
<td>-0.008</td>
</tr>
<tr>
<td>cash(*)</td>
<td>0.068</td>
<td>0.009</td>
</tr>
<tr>
<td>cash(*)</td>
<td>-0.035*</td>
<td>-0.027</td>
</tr>
<tr>
<td>cash(*)</td>
<td>-0.055</td>
<td>0.062*</td>
</tr>
<tr>
<td>cash(*)</td>
<td>0.042</td>
<td>0.033</td>
</tr>
<tr>
<td>cash(*)</td>
<td>0.009</td>
<td>0.36*</td>
</tr>
<tr>
<td>cash(*)</td>
<td>0.037*</td>
<td>0.003</td>
</tr>
<tr>
<td>pacy(*)</td>
<td>0.026*</td>
<td>-0.038*</td>
</tr>
<tr>
<td>pacy(*)</td>
<td>0.002</td>
<td>0.008</td>
</tr>
<tr>
<td>pacy(*)</td>
<td>0.030*</td>
<td>0.011*</td>
</tr>
<tr>
<td>R(^2)</td>
<td>0.91</td>
<td>0.98</td>
</tr>
</tbody>
</table>

**Note** * Statistically significant at 0.05

Average Fuzzy Weights: 0.210, 0.207, 0.222, 0.137, 0.223
bership; the entropy of classification across all 188 branches and 5 segments was equal to 76.64, compared to a maximum entropy of 303.57, showing an overall overlap of 25%. Table 2 also includes the estimates for the aggregate translog function (i.e., a single cluster).

In Table 2, the coefficients of the monthly dummy variables ($MOUTH_2 - MONTH_3$) indicate that, ceteris paribus, the usage of labor has increased over the 5-month period, especially in the last two months, which could have been caused by distortions in the deflation of outputs into constant currency. The coefficients for AREA in Table 2 represent the elasticity of labor requirements relative to the size of the branch. One would expect these estimates to be positive because larger branches would require more labor for maintenance. This happens for all clusters except cluster C; its negative and statistically significant estimate is counter intuitive, implying that larger branches within this cluster have slightly lower maintenance costs. The other results in Table 2 are more readily interpreted if they are translated into measures of technical inefficiency and elasticity, which we will do later.

**EVALUATION OF THE CLUSTERWISE PROCEDURE**

In this section, we will present several pieces of evidence that the clusterwise procedure worked as intended on our data set. Our tests, and the logic underlying them, are discussed in this section.

**Test of Need to Adjust for Unmeasured Customer Characteristics**

If the DEA and aggregate translog procedures based on a common production frontier adequately capture the effect of differences due to the unmeasured customer characteristics, or if the effect of these differences is small, there is no need for the clusterwise procedure. Assume that the clusters revealed by the clusterwise procedure are valid measures of the effect of differences between customers (this will be tested later). Given this assumption, we would expect differences in average efficiency across clusters as revealed by DEA and translog to be zero if they capture the effect of market conditions, significant if they do not. Hence we hypothesize:

**H1:** Given that the clusterwise procedure is valid, DEA and aggregate translog do not adequately adjust for differences in customer characteristics.

The results of this test are presented in the final two rows of Table 3, which presents the average technical inefficiency obtained from the aggregate translog function and the average efficiency ratio obtained from the application of a DEA model (see Appendix A for details) for each cluster. The estimates vary significantly across clusters, and indicate that Clusters B and D are less efficient than the other clusters. To the extent that these differences are due to differences in customer characteristics which are captured by our cluster
TABLE 3
Description of the Five Clusters
(Average Outputs and Inputs)

<table>
<thead>
<tr>
<th></th>
<th>Cluster A</th>
<th>Cluster B</th>
<th>Cluster C</th>
<th>Cluster D</th>
<th>Cluster E</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABOR</td>
<td>4,674</td>
<td>4,763</td>
<td>3,535</td>
<td>5,541</td>
<td>4,000</td>
</tr>
<tr>
<td>CASH****</td>
<td>2,909</td>
<td>2,819</td>
<td>2,352</td>
<td>3,172</td>
<td>2,570</td>
</tr>
<tr>
<td>OTHER***</td>
<td>1,845</td>
<td>1,693</td>
<td>1,503</td>
<td>1,910</td>
<td>1,549</td>
</tr>
<tr>
<td>PAYC***</td>
<td>0,972</td>
<td>0,669</td>
<td>0,577</td>
<td>0,912</td>
<td>0,837</td>
</tr>
<tr>
<td>MREV***</td>
<td>0,819</td>
<td>0,829</td>
<td>0,611</td>
<td>1,043</td>
<td>0,669</td>
</tr>
<tr>
<td>AREA***</td>
<td>720.2</td>
<td>664.8</td>
<td>599.4</td>
<td>642.4</td>
<td>708.0</td>
</tr>
<tr>
<td>Aggregate Technical Inefficiency**</td>
<td>0.99</td>
<td>1.06</td>
<td>0.96</td>
<td>1.11</td>
<td>0.96</td>
</tr>
<tr>
<td>DEA Efficiency***</td>
<td>0.81</td>
<td>0.76</td>
<td>0.82</td>
<td>0.75</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Notes: * Differences across clusters are significant at 0.001 level
** Differences across clusters are significant at 0.01 level
*** Differences across clusters are not significant at the 0.05 level

ing procedure, a need to adjust for these differences in making efficiency comparisons is indicated. Thus H1 is accepted.

Table 3 also presents mean values for the various inputs and outputs by cluster. The data in the table suggest that differences between clusters are reflected primarily in labor input, and that other variables tend not to differ significantly between clusters. Because the clustering is based on different shapes for the cost function rather than average values of inputs and outputs this result does not have a clear interpretation. However, it is consistent with the need to spend more time servicing customers in various clusters, which is consistent with differences in customer efficiency across clusters.

Using the procedure outlined earlier, the parameter estimates from Table 2 were also used to compute the technical inefficiencies of each branch relative to its cluster. The resulting technical inefficiencies were quite different from the measures of efficiency obtained from the aggregate translog function or from the DEA model; the correlation between the DEA efficiency ratios and clusterwise relative inefficiency across the 188 branches was -.30 (remember that the scales go in opposite directions), while the correlation of aggregate inefficiency with the clusterwise inefficiency was .44. These differences reflect a fundamental distinction between the clusterwise frontier and the ones implied by DEA and the aggregate translog function. This is further evidence in support of H1.

Tests of Convergent and Discriminant Validity

One aspect of construct validity is convergent validity, the degree to which different measures of the same concept are in agreement; another aspect is discriminant validity, the degree to which measures of the same concept give unique results (Campbell and Fiske, 1959). We have an independent classification of the banks into groups which have similar customer characteristics, the managers’ own classification of the, 188 branches by type of
market, based on subjective criteria. We employ the managers’ classification to assess the convergent and discriminant validity of our procedure in two ways.

First, if the clusterwise procedure is working to control for differences in customer input across markets as intended, it should eliminate differences in average productivity between the different market types as defined by the managers, which would provide evidence of convergent validity. Moreover, evidence of discriminant validity would be provided if other procedures, such as DEA and aggregate translog cannot eliminate differences in average productivity between market types, while the clusterwise procedure accomplishes this: the clusterwise procedure would be the only one of the procedures to eliminate differences in average productivity in the way one would expect given the managers’ classification. Hence, we hypothesize:

**H2:** Given that the clusterwise procedure is valid, this procedure will eliminate differences in average productivity between market types as classified by managers, while alternative methods will not eliminate these differences.

A second way of assessing the convergence between the clusterwise procedure and the managers’ independent classification would be to study the agreement between the groupings produced by the two procedures. Hence, we hypothesize:

**H3:** Given that the clusterwise procedure is valid, this procedure will produce the same grouping of banks as the managers’ classification.

Table 4 presents the evidence needed to test H2. This table presents average efficiency measures within each of the five market types defined by the managers. The first three rows

**TABLE 4**

<table>
<thead>
<tr>
<th></th>
<th>Industrial/ Residential (Low)</th>
<th>Residential Commercial (Prime)</th>
<th>Financial District</th>
<th>Suburban Centers</th>
<th>Residential Commercial (Low)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profit Margin per Man-Hour</td>
<td>13.27 (0.47)</td>
<td>15.69 (0.50)</td>
<td>14.17 (0.92)</td>
<td>15.04 (0.75)</td>
<td>12.96 (0.39)</td>
</tr>
<tr>
<td>Efficiency Ratio (DEA)*</td>
<td>0.765 (0.016)</td>
<td>0.830 (0.039)</td>
<td>0.841 (0.035)</td>
<td>0.718 (0.029)</td>
<td>0.770 (0.018)</td>
</tr>
<tr>
<td>Technical Inefficiency (Aggregate)*</td>
<td>1.017 (0.017)</td>
<td>0.973 (0.036)</td>
<td>1.032 (0.036)</td>
<td>1.079 (0.040)</td>
<td>1.031 (0.012)</td>
</tr>
<tr>
<td>Technical Inefficiency (Clusterwise)**</td>
<td>1.000 (0.006)</td>
<td>1.000 (0.021)</td>
<td>0.995 (0.014)</td>
<td>1.002 (0.010)</td>
<td>1.010 (0.006)</td>
</tr>
</tbody>
</table>

* Differences across market types are significant at the 0.01 level

** Differences across market types are not significant at the 0.00 level
of the table present three efficiency measures which are alternatives to our clustwise measure. These are a simple index of labor productivity (profit margin per man-hour), and measures based on the DEA and aggregate translog approaches described earlier.

The results in Table 4 indicate that, as hypothesized, the clustwise procedure eliminates differences in efficiency between the five market types, while the other three measures do not. For the clustwise procedure, the hypothesis that the mean relative inefficiency is the same across market types could not be rejected at even the .80 level. On the other hand, means for each of the other three measures differ significantly across market types. For DEA and aggregate technical inefficiency, the averages within each group imply that branches in the (low income) suburban centers tend to be less efficient than the ones located in the financial district or in the prime residential/commercial neighborhoods. The financial district is the banking center of this large metropolitan area. Branches in this area serve large investors who close their transaction by phone, and more informed customers who usually do not need much help in their banking transactions. The prime residential districts are also characterized by educated high-income customers. On the other hand, the customers of suburban centers tend to be less educated on average. They probably take more time and need more help in their banking decisions and transactions. By failing to correct for these differences in customer characteristics, an evaluation of the branch managers on the basis of the DEA or aggregate technical inefficiency results would favor managers from branches in certain markets (e.g., financial district, prime residential/commercial neighborhoods). Conversely, our clustwise procedure does correct for these differences. In sum, H2 is supported.

Though not part of our hypothesis test, Table 4 indicates that, while DEA and aggregate translog provide similar results, there is substantial disagreement between these measures and the measures of productivity based on accounting ratios. In fact, the correlation between the DEA efficiency ratios and the profit margin per man-hour across all 188 branches was relatively small (r = 0.31). We have argued earlier that profitability measures are probably not valid measures of relative efficiency for our data. This result illustrates that simple ratio measures of efficiency such as profit per man-hour will not necessarily provide the same results as more sophisticated measures.

Table 5 displays the average fuzzy weights for each type of market, as defined by the managers. This table provides the information needed to test H3. As can be seen from

<table>
<thead>
<tr>
<th>Fuzzy Clusters</th>
<th>Industrial/Residential (Low)</th>
<th>Residential/Commercial (Prime)</th>
<th>Financial District</th>
<th>Suburban Centers</th>
<th>Residential Commercial (Low)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A’</td>
<td>0.14</td>
<td>0.20</td>
<td>0.52</td>
<td>0.20</td>
<td>0.24</td>
</tr>
<tr>
<td>B’</td>
<td>0.15</td>
<td>0.15</td>
<td>0.19</td>
<td>0.23</td>
<td>0.39</td>
</tr>
<tr>
<td>C’</td>
<td>0.19</td>
<td>0.34</td>
<td>0.08</td>
<td>0.06</td>
<td>0.14</td>
</tr>
<tr>
<td>D’</td>
<td>0.17</td>
<td>0.11</td>
<td>0.12</td>
<td>0.39</td>
<td>0.09</td>
</tr>
<tr>
<td>E’</td>
<td>0.35</td>
<td>0.20</td>
<td>0.09</td>
<td>0.12</td>
<td>0.14</td>
</tr>
<tr>
<td>Total</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Notes: * Differences across market types are significant at the 0.01 level
** Differences across market types are significant at the 0.05 level
Table 5, there is a correspondence between the two classifications. Thus, branches in the financial district are more likely to belong to fuzzy cluster A; branches in the low residential/commercial areas are more likely to belong to cluster B; branches in the prime residential/commercial area to cluster C; branches in the suburban centers to cluster D; branches in the low industrial/residential to cluster E. The differences in fuzzy weights across the market types are all statistically significant at the 0.01 and 0.05 levels, as indicated in Table 5. The correspondence between the fuzzy weights and the independent classification of markets provides evidence of the convergent validity of our procedure, supporting H3.

Test of Face Validity

If the clusterwise procedure is valid, it should produce estimates of elasticities and returns to scale that agree with what one might expect from knowledge of the characteristics of the markets that form the basis for the managers' classification. Hence we hypothesize:

**H4:** Given that the clusterwise procedure is valid, this procedure will produce estimates of elasticities and returns to scale for each of the managers' market types that agree with one might expect for that type

The estimates of elasticity and returns to scale required to test H4 are presented in Table 6. Overall, the estimates in the table have face validity. Thus, as the last column of Table 6 shows, "cash deposits" is the most labor intensive output; a one percent increase in output requires a 0.29% increase in labor, compared to 0.09% for "managerial revenues." Since cash deposits are the most time consuming transactions in any branch, this result is reasonable. The variation in elasticities across market types, which is significant at the .05 level for each category, also agrees with what one might expect. In particular, the relatively high elasticity attached to other deposits reflects the complexity of non-cash transactions at the financial center banks, and the relatively high coefficient for managerial revenue for the suburban centers reflects their heavy use as a vehicle for routine financial

**TABLE 6**

<table>
<thead>
<tr>
<th>Output Category</th>
<th>Industrial/Residential (Low)</th>
<th>Residential Commercial (Prime)</th>
<th>Financial District</th>
<th>Suburban Centers</th>
<th>Residential Commercial (Low)</th>
<th>All Branches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash Deposits</td>
<td>0.21</td>
<td>0.30</td>
<td>0.34</td>
<td>0.29</td>
<td>0.34</td>
<td>0.29</td>
</tr>
<tr>
<td>Other Deposits</td>
<td>0.07</td>
<td>0.13</td>
<td>0.28</td>
<td>0.09</td>
<td>0.07</td>
<td>0.11</td>
</tr>
<tr>
<td>Pay &amp; Collect</td>
<td>0.12</td>
<td>0.09</td>
<td>0.08</td>
<td>0.11</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td>Mgr Revenue</td>
<td>0.09</td>
<td>0.10</td>
<td>0.09</td>
<td>0.13</td>
<td>0.06</td>
<td>0.09</td>
</tr>
<tr>
<td>RTS</td>
<td>1.76</td>
<td>1.51</td>
<td>1.23</td>
<td>1.38</td>
<td>1.44</td>
<td>1.54</td>
</tr>
</tbody>
</table>

Notes: RTS is returns to scale, which is defined as $\text{RTS}_0 = (1 - \phi_1)\sum_{e=1}^{E}\hat{e}^{-1}$, where $\phi_1$ is the coefficient of InAREA in equation 4, and E is elasticity (see equation 5). Values of RTS greater than 1 indicate increasing returns to scale, less than 1 indicate decreasing returns.
transactions, and the additional labor time needed to complete these transactions in this poor area.

Table 6 also shows that branches in all five market types operate at increasing returns to scale. However, the branches located in the financial district and in suburban centers have lower RTS. Since these branches tend to be the largest (see Table 1), this is reasonable: one would expect that they are most likely to be operating at a scale large enough to come close to exhausting potential economies. In sum, the estimates of elasticities and returns to scale for the various market types agree with what might be expected. H4 is supported.

CONCLUSIONS AND DIRECTIONS FOR FURTHER RESEARCH

In comparing retail outlets, managers must take into consideration the fact that consumers are active participants in the production of retail services. Therefore, any performance evaluation of multiple retail outlets must consider the potential differences in market conditions and consumer types faced by the different outlets, which might impact on their productivity. While the ideal solution to this problem would be to have data on numbers of different customer types patronizing each branch, and on numbers and characteristics of different types of transactions, these data are not likely to be readily available in many different retail settings. Given this lack of information on customers and transactions, we proposed a clusterwise regression procedure as a method of controlling for the impact of unmeasured customer characteristics on efficiency.

Our initial experience with this procedure was encouraging. We uncovered five distinct cost structures in our data set. We found that alternative methods based on a common production frontier for all branches did not adjust for differences in the structure of cost, implying that they could not be used for making efficiency comparisons. To validate our procedure, we compared it to an a priori classification of the branches by managers. We found that our method eliminated efficiency differences between the groups defined by the managers, while alternative procedures did not, and that our procedure provided a clustering of the banks which was related to that provided by the managers. Moreover, estimates of elasticities and returns to scale obtained from the clusterwise procedure had face validity when compared across the market types defined by the managers.

When we use the a priori classification of markets by the bank's management as a criterion to assess the validity of our results, we do not require that their classification be completely valid, without any error. Rather, the operational concept of validity is convergence: one would not expect two different methods for measuring the same thing to give comparable results unless they were in some way measuring the concept that they were designed to measure. However, one might argue that if the a priori classification forms directly comparable groups of branches, one could estimate translog cost functions for each of these groups using simple regression analysis, thus eliminating the need for the clusterwise approach. There are at least two reasons to justify the clusterwise approach. First, one might not necessarily have this type of a priori information, which is not required for the clusterwise approach. Second, some of the groups identified by the managers might be too small (12 branches in the Financial District, and 10 branches in Subur-
ban Centers, in our illustration), making it difficult (if not impossible) to estimate a translog cost function within each group.

To the extent that the clusterwise cost estimates reflect differences in the costs of serving different customer types, comparisons of efficiency relative to each cluster’s cost become more valid and equitable. However, since differences between cost functions for each cluster could also be partly due to differences in efficiency, the efficiency comparisons based on the clusterwise technique have to be used with some degree of judgment. Unfortunately, the only foolproof way to determine relative productivity would be to have the missing data on customer inputs.

While this paper has presented the clusterwise approach as a method for providing equitable comparisons of the efficiency of outlets, it has other potential applications in a retail setting. In particular, it might be used as a device for exploring the impact of market conditions on output in cases where these are not measured directly but thought to have a possible influence on productivity. In these cases, results from the clusterwise procedure might provide insights into which market factors may be affecting productivity, and would also determine whether further work can proceed on the assumption of a single cost function which can apply to all of a firm’s outlets. For example, if our study had been undertaken prior to the existence of the independent classification of outlets, the results would have indicated that the assumption of a single cost function is untenable, and the results could also have been used to infer that the location and economic status of market areas has an impact on the cost function.

The problems discussed in this paper, and the method which we propose for solving them, are not limited to banks, or even to evaluating different retail outlets. Virtually any attempt to compare the productivity of different units, such as sales territories, franchises, branch offices, entails similar problems which might be attacked with the procedures discussed in this paper.

While this study has focused on efficiency and productivity, future work should also examine effectiveness, i.e., the ability of the branch to achieve the firm’s goals. That is, we have studied how to determine differences between branches in the extent to which they minimize cost for a given level of output. There is also a need to examine whether they maintain the most desirable level of output. Finally, like most studies of retail productivity, this study was hampered by a lack of data on the customer’s input into the process. Firms need to be sensitive to the need to collect these data, and better methods for doing so need to be devised.

**APPENDIX A**

In this appendix, we first outline the general mathematical structure of the DEA procedure. We then indicate how the DEA application in this study was set up.

**General Structure of Data Envelopment Analysis**

Define the following notation:
Productivity Assessment

\[ y_{ib} = \text{units of output}_i \text{ produced by outlet (or branch) } b \]
\[ x_{jb} = \text{units of input}_j \text{ used by outlet (or branch) } b \]
\[ z_{kb} = \text{units of allocative input}_k, \text{ not under direct control by the store manager, used by outlet } b \]

Then, for the purposes of our study, it suffices to know that the technical efficiency of any retail outlet (designated by the subscript o) is obtained by solving the following linear programming problem:

\[
\min T_o
\]
\[ \text{s.t. } y_{io} \leq \sum_b \lambda_b y_{ib} \quad i = 1, \ldots, I \quad (A1a) \]
\[ T_o x_{jo} \geq \sum_b \lambda_b x_{jb} \quad k = 1, \ldots, J \quad (A1b) \]
\[ z_{ko} \geq \sum_b \lambda_b z_{kb} \quad t = 1, \ldots, K \quad (A1c) \]
\[ \lambda_b \geq 0 \quad b = 1, \ldots, B \quad (A1d) \]
\[ \sum_b \lambda_b = 1 \quad (A1e) \]

The outlets b for which \( \lambda_b > 0 \) define the efficiency frontier for the particular outlet under analysis (designated by the subscript o). Let \( T_o^*, s_{ib}^*, s_{jb}^* \), denote the optimal value of the objective function, and the slacks for the constraints in Equations 1a and 1b, respectively. The optimal \( T_o^* \) measures the efficiency of the outlet under study, relative to this efficiency frontier. The maximum possible value of \( T_o^* \) is one, which holds when the outlet is on the efficiency frontier; as the relative efficiency of the outlet declines \( T_o^* \) moves toward zero. The outlet under analysis can be made efficient either by:

1. Reducing its usage of inputs \( x_{jb} \) by \((1 - T_o^*)x_{jb} + s_{jb}^*\), while increasing the outputs by \( s_{ib}^* \); or
2. Increasing the outputs \( y_{ib} \) by \((1 - T_o^*)y_{ib} + s_{ib}^*\), while reducing the inputs by \( s_{jb}^* \).

The dual value associated with Constraint 1e indicates whether the outlet under analysis operates under increasing, constant or decreasing returns to scale, depending on whether this dual value is less, equal or greater than zero, respectively (Mahajan, 1991).

DEA Model Employed in This Study

The following DEA model was applied to each branch, using data aggregated over the 5 months:

\[
\min T_o - \varepsilon \left\{ s_+^{\text{Cash}} + s_+^{\text{Other}} + s_+^{\text{PayC}} + s_+^{\text{MREV}} + s_-^{\text{LABOR}} + s_-^{\text{AREA}} \right\} \quad (A1)
\]
The linear-programming model above was implemented within LINDO (Schrage, 1986), a widely used linear-program software package. We applied the model to the 188 branches, and identified only 19 of them as efficient (i.e., $T_o = 1$ and $s^+_{OTHER} = s^+_{PAYC} = s^+_{MREV} = s^+_{LABOR} = s^+_{AREA} = 0$). These 19 branches define all the facets of the efficiency frontier, and the efficiency of every other branch is defined relative to a convex combination of subsets of these efficient branches.

APPENDIX B

In this appendix, we present a brief description of the clusterwise regression procedure developed by Wedel and Steenkamp (1991), as it is applied to our problem. Further details about this methodology can be found in Wedel and Steenkamp (1991). To simplify the exposition, re-define Equations 2 and 3 as

$$ y_{bt} = \sum_g u_{bg} \left[ \sum_k b_{gk} x_{bkt} + e_{bgt} \right] $$

where:

$y_{bt}$ = dependent variable for branch $b$ at period $t$

$x_{bkt}$ = predictor $k$ for branch $b$ at period $t$

$u_{bg}$ = fuzzy weight allocating branch $b$ to group $g$

$e_{bgt}$ = regression residual

$\Sigma_g u_{bg} = 1$; $\Sigma_b u_{bg} > 0$

The objective is to minimize the weighted sum of squares

$$ SSE = \sum_g \sum_b \sum_t u_{bg} e_{bgt}^2 $$
subjected to the constraint $\Sigma g u_{bg} = 1$. The constant $\alpha > 1$ is a weight that affects the degree of overlap among the groups. By solving this optimization problem, Wedel and Steenkamp (1989) derive the fuzzy-clustering regression estimates:

$$b_g = (X'U_g^{\alpha}X)^{-1}X'U_g^{\alpha}y,$$

where $U_g^{\alpha} = \text{diag}(u_{bg}^\alpha)$, and

$$u_{bg} = \frac{1}{\sum_{g'} \left[ \sum_t e_{bg't}^2 \right]^{1/(\alpha-1)}}$$

**Acknowledgment:** The authors gratefully acknowledge the assistance of Ibere Arco e Flexa e Naor Leal in data collection and comments. They also thank the editor and three anonymous referees for their helpful comments on an earlier draft.

**NOTES**

1. Productivity refers to output from a given set of inputs, while efficiency refers to inputs needed to produce a given set of outputs. Since there is a duality between the two terms, they can generally be used interchangeably.
2. We are indebted to Prof. Michel Wedel for making the estimation algorithm available to us.
3. Since $\ln C = \ln w + \ln \text{LABOR}$, subtracting $\ln w$ from the left side of 2 leaves $\ln \text{LABOR}$. Because costs must be homogeneous of degree 1 in factor prices, constraints on the translog require that $\Sigma \beta_j = 1$, $\Sigma \beta_{jg} = 0$, $\Sigma \delta_j = 0$ (Caves, Christensen and Swanson, 1980). Thus, if there is only one factor, the coefficient of $\ln \text{w}$ on the right side of Equation 2 must = 1, and the coefficients of higher order terms involving $\ln \text{w}$ must be zero. Subtracting $\ln \text{w}$ from the right side of Equation 2 to balance the equation gives Equation 6.
4. The measure JRML is a weighted sum of squared errors which the clusterwise regression procedure seeks to minimize for a given number of clusters.
5. In order to reduce the risk of reaching a local optimum in the clusterwise regression procedure (Wedel and Steenkamp, 1991, p 388), five solutions were obtained from different random initial estimates. The solution reported is the one with the best fit.
6. Data on each branch in different time periods could be incorporated into the analysis, but would require the treatment of each branch-month combination as an independent decision making unit.

**REFERENCES**


Productivity Assessment


