ORiON, Vol. 14, No. 1/2, pp.57-66 *ISSN 0259-191-X*

USE OF DATA ENVELOPMENT ANALYSIS AND REGRESSION FOR ESTABLISHING MANPOWER REQUIREMENTS IN A BANK

L. PAUL FATTI

Department of Statistics & Actuarial Science
University of the Witwatersrand
Johannesburg
South Africa

and

KATHLEEN CLARKE

Standard Bank of South Africa, Ltd
Johannesburg
South Africa

ABSTRACT

We describe an approach towards forecasting the manpower requirements in each of the branches of a bank, based on regression models fitted to the sets of efficient branches. DEA is employed to identify the efficient branches within a category, using the numbers of employees in the different grades at each branch as input variables, and the average volumes of different types of work performed by them during a month as output variables. Forecasts of future volumes of work are obtained by fitting a model which takes into account branch and seasonal effects, as well as separate trend effects for each of the branches. The models have been tested on data from a large bank, with very encouraging results. The approach holds great promise for use towards a decision support system for managing the bank's total branch manpower requirements.

1. INTRODUCTION AND BACKGROUND

Manpower is one of the largest and most expensive resources utilised in a bank. It is thus of great importance for the bank to ensure efficient utilisation thereof and to plan its future manpower requirements effectively.

The study described here concerns a large bank with well over four hundred branches operating around South Africa and having considerable international operations. One of the major ongoing areas of concern in managing the bank's daily operations efficiently is to establish and maintain appropriate staffing levels within the different branches.

A number of factors are taken into account when establishing the numbers of people which the bank should employ in the various staffing grades at a branch. The overriding consideration is the volume of work which needs to be performed in the branch each day, but certain types of non-clerical grades need to be filled no matter what the workload of the branch. Work study methods are employed to establish standard times for performing certain types of work, such as processing a debit or credit entry, and these have been found to be of great value when trying to predict changes in staffing levels over the short term. However, when forecasting staffing levels over the longer term (twelve months or more), the joint effect on the staffing levels of the anticipated growth in all business, represented by the simultaneous changes in all volumes of work, has to be estimated.

In the light of this, when the Operations Research Department at the bank's head office was approached for help with this problem, it was decided to tackle it using regression analysis. The idea was to relate the actual volumes of work processed by the different branches during a fixed period of time, say a month, to the actual numbers employed in the different grades (or their total) in each branch over that time, using either univariate or multivariate multiple regression analysis. The regression model could then be used to forecast the manpower requirements of a branch on the basis of its forecast workload over the next planning period. This approach could then form the basis of a decision support system towards managing the manpower requirements in all the branches of the bank.

However, it soon became apparent that the fit of the regression model was just not good enough for it to be used as a reliable tool for manpower planning, as the variability of the actual staffing complements of branches around their values predicted by the model was just too great. After discussing this problem with the manpower planning department of the bank it became evident that the kinds of work performed by branches could be very different, depending on the types of market (such as commercial or personal) which they served. The branches were therefore grouped together into five homogeneous categories according to the specific kinds of work which they performed (for example, whether they performed their own back office functions or whether they were being supported by a centralised processing function) and separate models were fitted to the branches in each category.

While this improved matters, the variabilities about the individual regression models were still too great for them to be used for forecasting manpower requirements. Further discussions in the light of these findings revealed that the observed discrepancies from the models could be explained by the different efficiencies with which the different banks conducted their business. It was therefore decided to base the forecast manpower requirements of a branch on a regression model representing only the efficient branches in its category. The bank had a heuristic procedure, based on the ratio of the assessed workload (calculated from the volumes of work and standard times) to the available manhours, for identifying the branches it deemed to be efficient, but it was considered to be too subjective to be used for the present exercise. After some discussion it was decided to use Data Envelopment Analysis as an objective method for identifying the set of efficient branches in each category for fitting the regression models.

Data Envelopment Analysis (DEA) developed by Charnes, Cooper and Rhodes [1], [2] is an approach concerned with producing a measure of efficiency of an individual 'decision making unit' (DMU) with respect to the other units within the same group. The measure is based on the ratios of linear combinations of the inputs and outputs of the units, and Linear Programming is used to optimise the efficiency of a DMU with respect to the others. Good descriptions of DEA are given in Charnes, Cooper, Lewin and Seiford [3] and in Norman and Stoker [4].

The following sections detail how DEA was implemented to identify the efficient branches, how variables were selected to include in the DEA and in the subsequent regression analyses, and the forecasting method used for predicting future workloads of the branches. The paper finishes with a discussion of the results of the study and an evaluation of its effectiveness.

2. IDENTIFYING EFFICIENT BRANCHES USING DEA

The bank routinely collects information on the numbers of accounts of different types held at each of its branches, and on the numbers of each of a large number of kinds of transactions processed by them. These are referred to collectively as 'volumes of work' and are summarised on a monthly basis on the bank's computerised database. The first task of the Operations Research team was to identify a set of these volumes which would best describe the activity and growth of a branch in business terms. Discussions with the relevant parties as well as statistical (correlation and regression) analyses which were performed on the data, resulted in the identification of eight volumes of work which were considered to be the best measurements of the output of a branch. These volumes are listed in Table 1. The data for these volumes covered a period of sixteen months, and their monthly values, averaged over this period, were used as output variables in the DEA.

The average numbers of people employed within the various grades at a branch over the sixteen month period were used as input variables. While the bank, in practice, only needed forecasts for the total staffing requirements at each branch, it was felt that an efficient branch was one which utilised its various grades of staff efficiently. After extensive debate, certain grades, such as 'non-clerical', were excluded from the analysis, as they were considered to be part of the overhead 'fixed cost' of a branch and not related to assessed workload. The five employment grades used as input variables in the DEA are also listed in Table 1.

Table 2 gives the names of the five categories of branch and the numbers of branches in each, and separate DEA runs were performed for each of them. Since there were only eight outputs and five inputs to consider, compared to more than eighty branches (on average) within each category, it was decided to solve the dual formulation of the DEA linear programming model.

Use was made of the SAS/OR [5] software package to design a program for solving the linear programs iteratively over all the branches within a category. Each branch, in turn, was used in the objective function, and its efficiency was computed relative to all the branches within the category. Only those branches with an efficiency of 100%

were considered to be efficient. The numbers and corresponding percentages of efficient branches in each category are given in Table 2.

Table 1: Input and output variables used in the DEA

Input Variables (Employment Grades)
Branch manager
Branch Administrator
Clerical Grades
Checking Grades
Supervisory Grades
Output Variables (Volumes of Work)
(1) No. of cheque accounts with credit balances
(2) No. of cheque accounts with debit balances
(3) No. of savings accounts
(4) No. of cheque account paper-based debit entries
(5) No. of cheque account paper-based credit entries
(6) No. of savings account paper vouchers
(7) No. of teller transactions
(8) No. of foreign exchange transactions

Table 2: The five categories of branches

	Category of Branch	No. Branches	No. Efficient	% Efficient
a.	Self accounting mainframe	54	33	61%
b.	Hub accounting centre	69	49	71%
c.	Feeder	156	80	51%
d.	Self accounting micro	72	29	40%
e.	Frontline service	96	65	68%
	Total	447	256	57%

3. FITTING REGRESSION MODELS TO THE EFFICIENT BRANCHES

Once the set of efficient branches had been established for each category, they were used to fit regression models using the SAS/STEPWISE [5] procedure. An optimal set of predictors was selected from the eight volumes of work, with the total complement at each branch as the dependent variable. Because of the relatively high correlations which exist between some of the volumes, the final models always incorporated less than the full set.

Table 3 gives the volumes used and the squared multiple correlation coefficients for the regression models fitted to the efficient branches in each of the five categories.

Table 3: Predictors and Squared Multiple Correlation Coefficients for the Regression Models

Branch Category	Predictors	R^2
a.	(3) and (5)	94,2%
b.	(2), (6) and (8)	95,1%
c.	(1), (3), (5) and (8)	97,0%
d.	(1) and (7)	97,7%
e.	(3), (4) and (8)	93,5%

It is evident from this table that parsimonious models could be found for each of the five categories, and that these gave very good fits to the data from the efficient branches.

4. FORECASTING FUTURE GROWTH IN THE VOLUMES OF WORK

In order to use these regression models to forecast future manpower requirements, it was necessary to forecast the volumes of work, to be used as predictors in these models, for each branch over the planning period. While the bank has considerable experience in applying forecasting techniques, such as seasonal smoothing and ARIMA modelling, to the various banking variables for planning and budgeting purposes, these all require adequate historical data for them to be applied with any measure of confidence. Due to the very large amount of data which is generated by each branch every month, it is the bank's practice to keep detailed records for only a limited period of time and to store only summarised data on a longer term basis.

As a result there were only relatively short recorded histories of the volumes of work data available for each of the branches (usually the most recent sixteen monthly averages) from which to forecast future volumes. The OR team therefore developed the following model for forecasting the volume of work in a particular branch during a particular month, which incorporates seasonal factors, branch factors and monthly trends at at each branch:

$$x_{ij} = \mu + \alpha_i + \beta_j + \gamma_i\left(j - \frac{t+1}{2}\right) + \varepsilon_{ij}$$

where:

x_{ij}	=	volume of work in branch i during month j
i	=	1,...,b branches
j	=	1,...,t months
μ	=	overall mean volume of work
α_i	=	effect of branch i
β_j	=	(seasonal) effect of month j
γ_i	=	monthly trend for branch i
and ε_{ij}	=	random error term.

Least squares estimation yields the following parameter estimates:

$$\hat{\mu} = x_{..}$$

$$\hat{\alpha}_i = x_{i} - x_{..}$$

$$\hat{\beta}_j = x_{.j} - x_{..} - \hat{\gamma}_{.}\left(j - \frac{t+1}{2}\right)$$

$$\hat{\gamma}_i = \sum_{j=1}^{t}\left(j - \frac{t+1}{2}\right)\left(x_{ij} - \hat{\beta}_j\right) \Big/ \left(\sum_{j=1}^{t}\left(j - \frac{t+1}{2}\right)^2\right)$$

where $x_{i.}$ denotes the average volume of work at branch i over all months, etc. The last two equations are solved iteratively, starting with an initial value of $\hat{\gamma}_{.} = 0$ for the average monthly trend over all branches. For any month which appears twice over the period for which the model is fitted, the least squares squares estimate of its effect is the average value of the corresponding two $\hat{\beta}$'s.

This model was fitted separately for each of the volumes which was used as predictor in each of the five categories of branches. In all cases the least squares equations converged in a small number of iterations, yielding a set of models which required only a limited amount of historical data (twelve months' or more) from each branch for forecasting the volumes of work used as predictors in the manpower planning models. Note that while these forecasting models impose common seasonal factors over all branches in a category, they allow for different trends in each of the branches. Indeed, it was not uncommon to find cases where the trend in a particular volume of work was positive for some branches, indicating an increasing average activity in this area, while it was negative for other branches in the same category.

5. RESULTS AND CONCLUSIONS

In order to test the efficacy of using the proposed approach for forecasting the manpower requirements in the bank, the fitted regression models were applied to the averages of the actual volumes of work measured at each of its branches over the most recent six-month period. The results obtained by this method were then compared with those obtained by the bank using its previously mentioned heuristic method procedure for selecting efficient branches. Table 4 gives the total manpower forecasts for each of the categories of branches over the six-month period, which were obtained by:

1. the method described in this paper, and
2. using regression models fitted to the branches which were selected via the bank's heuristic procedure.

Table 4: Total Manpower Forecasts using the two different methods

Category	Method 1	Method 2
a.	1161	1318
b.	3059	3002
c.	3312	3703
d.	626	699
e.	3084	3426
Total	11242	12148

Table 5 gives the percentages of banks in which Method 1 gives lower, higher and the same manpower forecasts, respectively, as Method 2.

Table 5: Percentages of lower, higher and the same manpower forecasts

Category	% Lower	% Higher	% Equal
a.	72%	22%	6%
b.	54%	39%	7%
c.	78%	11%	11%
d.	63%	0%	37%
e.	61%	26%	13%

From Tables 4 and 5 it is clear that the method described in this paper produces manpower estimates that are generally lower than those produced by Method 2, and results in a saving of 7,5% in the total estimated manpower requirements. Since the only difference between these two methods is the way in which the efficient branches is selected, it follows that DEA is a stricter selection method than the heuristic one used by the bank.

From the bank's point of view, the fact that the proposed method provides lower estimated manpower requirements than its previous method yet, as is clear from Table 1, is based on a reasonably large subset of branches, is good news. It implies that it has an effective, defensible, yet stricter method of manpower budgeting than it has had up till now. Given the fact that manpower is such an expensive resource for the bank, the proposed method could potentially lead to significant savings by trimming numbers from the inefficient branches without affecting the efficient ones.

In view of the success of the trial, the bank has decided to implement the proposed approach for its next round of manpower forecasts. The method described in Section 4 is currently being used to forecast the volumes of work over the planning period, and these will then be used in the regression models to estimate the manpower requirements in each of the five categories of branches over the next year. Hopefully, the whole approach which has been described in this paper will in due course be incorporated into a decision support system for managing the bank's total branch manpower requirements.

REFERENCES

[1] Charnes, A., Cooper, W.W. and Rhodes, E. *Measuring the Efficiency of Decision Making Units,* European Journal of Operational Research, 2, 429-444 (1978).

[2] Charnes, A., Cooper, W.W. and Rhodes, E. *Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through*, Management Science, 27, 668-697 (1981).

[3] Charnes, A., Cooper, W.W., Lewin. A.Y. and Seiford, L.M. eds., *Data Envelopment Analysis - theory, methodology and applications*, Kluwer, New York (1993).

[4] Norman, M and Stoker, B. *Data Envelopment Analysis - the Assessment of Performance*, Wiley, London (1991).

[5] SAS Institute, Inc., Cary, North Carolina (1990).