Paper 9280-2016

An Application of the DEA Optimization Methodology to Optimize the Collection Direction Capacity of a Bank

Miguel Angel Díaz Rodríguez, Oscar Javier Cortes Arrigui, Scotiabank, Bogotá, Colpatria

ABSTRACT

In the Collection Direction of a well-recognized Colombian financial institution, there was no methodology that provided an optimal number of collection agents, to improve the collection task and make possible that more customers be compromised with their minimum monthly payment of their debt. The objective of this paper is to apply the Data Envelopment Analysis (DEA) Optimization Methodology, to determine the optimal number of agents to maximize the monthly collection in the bank. We show that the results can have a positive impact in the credit portfolio behavior and reduce the collection management cost. DEA optimization methodology has been successfully used in various fields to solve multi-criteria optimization problems, but it is not commonly used in the financial sector mostly because this methodology requires specialized software, such as SAS® Enterprise Guide®. In this paper, we present the PROC OPTMODEL and we show how to formulate the optimization problem, program the SAS® Code, and how to process adequately the available data.

INTRODUCTION

In a bank the credit life cycle is composed by four main stages: credit marketing, acquisition and underwriting, account management and recovery and collections. Debt collection is the final stage and is where creditors try to recover the debts owed to them. It is the primary mechanism of contract enforcement in consumer credit markets and affects millions of consumers. Creditors can try to collect on their own, or they may outsource collections to third-party firms. (Fedaseyeu & Huntb, 2014)

The collection process begins when the payment of one single customer has not arrived on due date, then the customer will be classified in the bucket or group of debtors where days past due (dpd) are between 1dpd to 30dpd. A third party firm will contact the customer to remind him/her to pay, and the customer can or cannot agree to pay. If the reminder is ignored and the customer does not pay, the customer's dpd will advance until he/she can be classified on another bucket. The contact process will be repeated until the customer pays or until he/she is classified on the highest bucket of dpd and will be declared on bankruptcy. Buckets are between 1dpd to 30dpd, 31dpd to 60dpd, 61dpd to 90dpd and bankruptcy.

The collection process is done according to these buckets and each collection agent will be assigned to a particular dpd bucket to work on each month. There is no discrimination between products and there are strategies that determine how often the customer should be contacted on the month according to a statistical model that predicts how likely the customer is to pay.

In this paper we use Data Envelopment Analysis optimization methodology to determine the best profiles of the collection agents and find the optimal number of agents by profile required to maximize monthly collection. In the first place we provide information about the call center and the information sources, in second place we introduce DEA methodology, in third place we formulate the optimization problem, describe de PROC OPTMODEL in SAS and finally asses the results and conclude.

COLLECTIONS

Debt collection firms play a fundamental role in contributing on the stability of the credit portfolio. Through the recovery of outstanding debts, debt collection firms helps businesses to remain solvent, prevents passing the costs of both monetary loss and risk on to consumers and give latitude to the management to decide if it is possible to continue extending credit to consumers. (Adams, 2016) Risk appetite and growth also depend directly on the recovery levels for each product.

Now-days collections are done by third-party firms, more efficient on negotiating with delinquent clients and experts on designing flexible payment plans for outstanding debts. Nevertheless there are improvement opportunities on each of these debt collection firms, who have limited budget and resources to achieve their main goal, normalize delinquent clients and recover outstanding.

The collections call center we focus our analysis on, is placed in Bogotá. Manage the delinquent accounts of a well-recognized Colombian bank with two million customers and who outsourced their collections eleven years ago to get better results. Nevertheless the current capacity of the place is for a maximum of four hundred people and nowadays three hundred ninety four people work in there, then they are working at almost their maximum capacity. This entity is willing for somebody to help them with anything that allows them to improve their efficiency in collections, which is measured mostly in number of normalized delinquent clients and recovered outstanding.

Although average monthly recovery depends on the product and dpd bucket, these call center recovers around 30 million dollars each month. Even if 30 million dollars sounds great, their expenses are as important as their recovery levels. Variable costs, fixed costs and investment, represent a big challenge for this firm to run efficiently.

The investment is already done, fixed costs will be stable over time, and both of them only depend of the capacity of the call center. But variable costs represent an important saving opportunity for the firm if collections are done efficiently. Variables include the number of calls, the time in each call and the number of agents doing the calls. Then the problem is not only to reduce the number of people who work there. The real challenge is to find those excellent agents profiles who with the minimum number of calls and time in each call, will recover the maximum outstanding.

Table 1 presents hypothetical monthly variable costs which vary depending on the intensity of each agent and are the ones that we should reduce by training for improved efficiency on the call center.

Monthly Variable Costs		
Maximum capacity	400	agents
# Monthly Calls per agent	150	calls
Av time in each Call	2.0	min
Call cost per minute	0.10	USD
Agent salary	258	USD
TOTAL Variable Costs	103255	USD

Table 1 Monthly hypothetical variable costs of the call center.

Table 2 present hypothetical monthly fixed costs which depend on the number of agents working in the call center. If the number of agents is reduced that wouldn’t represent important savings, then the cost optimization problem should focus on variable costs.

Monthly Fixed Costs		
Maximum capacity	400	agents
Energy per agent	6	USD
Water per agent	10	USD
Clean per agent	5	USD
TOTAL Fixed Costs	8226	USD

Table 2 Monthly hypothetical fixed costs of the call center.

Table 3 present hypothetical monthly investment costs which depend on the number of agents working in the call center. Those are assumed to be done once each year, but the costs are divided into twelve months for comparing purposes. Neither those are our focus.

Investment / 12 months		
Maximum capacity	400	agents
Computer cost	27	USD
Telephone cost	5	USD
Software licence cost	20	USD
TOTAL Investment	20968	USD

Table 3 Calculated monthly hypothetical investment costs for the call center

Some other additional information is that labor time on the call center is composed by the five days of the week plus two Saturdays on the month. Each day is divided in two journeys of eight hour each, morning and afternoon. The collections are done in two different physical places.

BACKGROUND

DATA ENVELOPMENT ANALYSIS

According to (Charnes et alter, 1997) Data Envelopment Analysis was developed originally by Rhodes (in his doctoral dissertation directed by W.W. Cooper) to assess the efficiency of “Follow -Through”, an academic program for public schools in United States. As the objective function of the optimization problem was fractional, Charnes and Cooper provide linear transformation methods for the conversion of generic optimization programs with fractional objective functions into linear and resoluble programs. (Fuentes, 2011).

DEA is a method for measuring efficiency of Decision Making Units (DMUs), which are referred to be units of organizations such as universities, factories and banks, which perform the same function. The methodology use linear programming techniques to “envelop” observed input – output vectors as tightly as possible. (Boussofiane, Dyson, & Thanassouli, 1991) One main advantage of DEA is that it allows several inputs and outputs to be considered at the same time. In this case, efficiency is measured in terms of inputs or outputs along a ray from the origin. (Emrouznejad, 2000)

The program seeks to optimize the measured efficiency of each analyzed DMU to create an efficiency frontier based on pareto. (Charnes et alter, 1997) First the empirical production frontier is constructed and then the efficiency of each DMU that didn’t belong to the efficiency frontier is assessed. A DMU is efficient and belong to the production frontier when it produces more of some output and no less than other outputs without consuming more inputs (output oriented). Or when using less of some input and no more of other inputs produces the same outputs (input oriented). (Fuentes, 2011).

Some advantages of DEA are: it use multiple inputs and outputs, it is useful when prices of factors are unknown, because it generate its own weights, it offer detailed information of each DMU to identify the efficient units, and it don’t demand knowledge of the production function, just of the inputs and outputs.

OPTIMIZATION PROBLEM

Now that the general concepts have been explained, in this stage we will focus on the formulation of the optimization problem.

Initially Charnes et alter formulate the following optimization problem:

$$Max. h_o = \frac{\sum_{r=1}^{s} U_r * Y_{ro}}{\sum_{i=1}^{m} V_i * X_{io}}$$

$$S.A.$$

$$\frac{\sum_{r=1}^{s} U_r * Y_{rj}}{\sum_{i=1}^{m} V_i * X_{ij}} \leq 1 \quad \forall j : 1 \dots n$$

$$U_r, V_i \geq \varepsilon > 0 \ \forall r, i$$

Where:

h_o= Objective function. Measure for efficiency.

X_{ij}=Input i of the DMU j

Y_{rj}=Output r of the DMU j

U_r, V_i=Weights of inputs and outputs respectively (problem solutions)

The objective function seeks to optimize the efficiency of each DMU and requires that each ratio is less than the unit (1), the nearest the ratio is to one the more efficient would be the DMU. The ratio of the objective function gives the proportion of number of outputs generated by each input of the DMUs. The program will find the values for U and V weights that make that h_o reaches it maximum with the condition that each ratio must be less than one. This gives autonomy to the optimization program to assign the weights for each DMU that is better for the solution, without taking into consideration the researcher judgment, so it will be more mathematical than judgmental. (Fuentes, 2011).

The problem of that model was that it wasn't linear, then they transform it into a linear model as follows:

$$Max. h_o = \sum_{r=1}^{s} U_r * Y_{ro}$$

$$S.A.$$

$$\sum_{i=1}^{m} V_i * X_{io} = 1$$

$$\sum_{r=1}^{s} U_r * Y_{rj} - \sum_{i=1}^{m} V_i * X_{ij} \leq 0 \quad \forall j : 1 \dots n$$

$$U_r, V_i \geq \varepsilon > 0 \ \forall r, i$$

Nevertheless the dual output oriented model was preferred over the primal, because it offers a better interpretation of the efficiency and comparison group is easier to identify. (Fuentes, 2011) It is formulated as follows:

$$Max. \varphi_o + \varepsilon(\sum_{i=1}^{m} S_{i-} + \sum_{r=1}^{s} S_{r+})$$

$$S.A.$$

$$\sum_{j=1}^{n} \lambda_j * X_{ij} + S_{i-} = X_{io} \quad \forall i : 1 \dots m$$

$$\sum_{j=1}^{n} \lambda_j * Y_{rj} + S_{r+} = \varphi_o * Y_{ro} \quad \forall j : 1 \dots s$$

$$\lambda_j, S_{i-}, S_{r+} \geq 0$$

Where:

λ_j=Weight of each DMU in each peer group of the DMU_o

X_i=Input i

Y_r=Output r

φ_o=Parameter that measure the efficiency of the evaluated unit

ε=Real positive small number

S_{i-} S_{r+} =Slack variables for inputs and outputs respectively. Transform inequality constraints in equality constraints, and when including slacks on the objective function will prevent dual maximums.

Finally, this output oriented model can be formulated without slack variables and inequality constraints as a model that compares DMU_o with each DMU that produces the same as or more than DMU_o using less or the same than DMU_o.

$$Max. \varphi_o$$

$$S.A.$$

$$\sum_{j=1}^{n} \lambda_j * X_{ij} \leq X_{io} \quad \forall\, i : 1 \dots m$$

$$\sum_{j=1}^{n} \lambda_j * Y_{rj} \geq \varphi_o * Y_{ro} \quad \forall\, j : 1 \dots s$$

$$\lambda_j \geq 0$$

Where:

λ_j=Weight of each DMU in each peer group of the DMU_o

X_i=Input i

Y_r=Output r

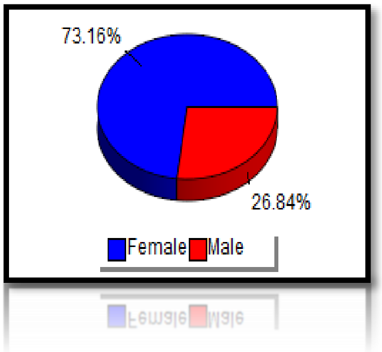
φ_o=Parameter that measure the efficiency of the evaluated unit

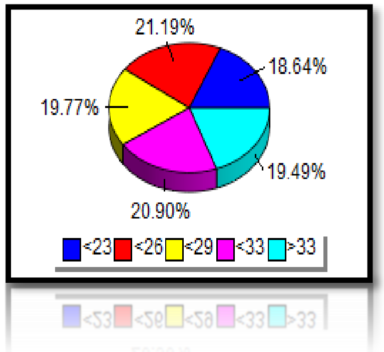
DATA

Three main sources of information were considered on the project. The first one with demographic information of agents (gender, education, age), the second one had the daily performance indicators for each agent (recovered money, daily number of calls, among others) and the third one had the daily work status of the agents (in license, no longer working, on vacations, among others). The process done with each data base was as follows:

1. Understand each data set.
2. Upload each data set to SAS.
3. Rename and recodify variables content to avoid nuisance factors.
4. Group metrics by date and agent.
5. Exclude missing values and target the population.
6. Merge the sources
7. Group by profile.

While not the only variables that could be considered with regard to efficiency in collection, demographic variables will be considered in this example. DMUs were created based on demographic variables such as gender, age, marital status, education and number of children, each DMU will be the state of each variable then for gender assessment of efficiency the DMUs will be Males and Females. The population participation for each DMU to be assessed, depending on the chosen demographic variable is illustrated on figures 1 to 6.





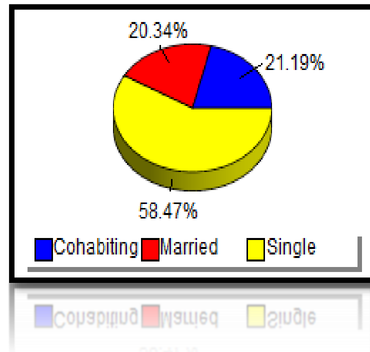
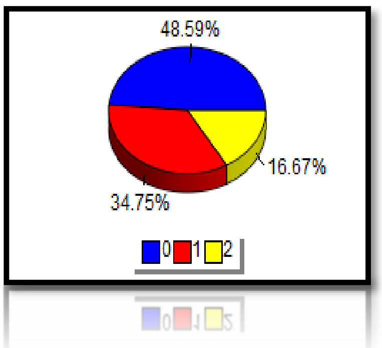


Figure 1. Agents by gender **Figure 2. Agents by age rank.** **Figure 3. Agents by marital status**



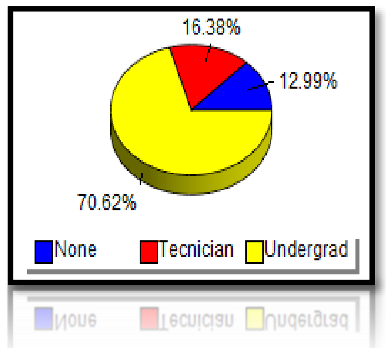
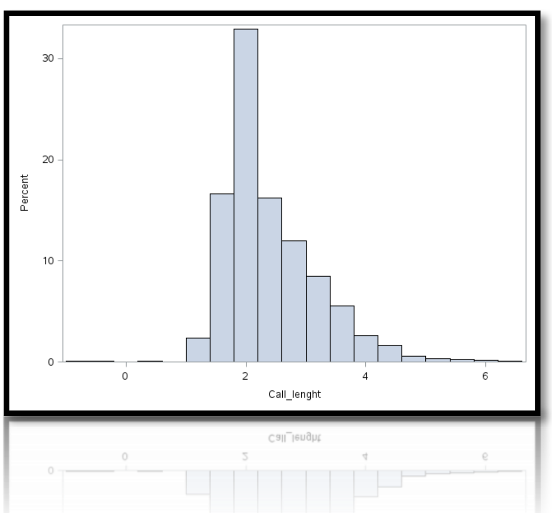


Figure 4. Agents by studies. **Figure 5. Agents by number of kids**

A large set of daily performance indicators for each agent were analyzed with the business. Then the chosen input and output variables considered the main goals of the call center at 30 dpd bucket and the technical process. The chosen variables distributions are described below.



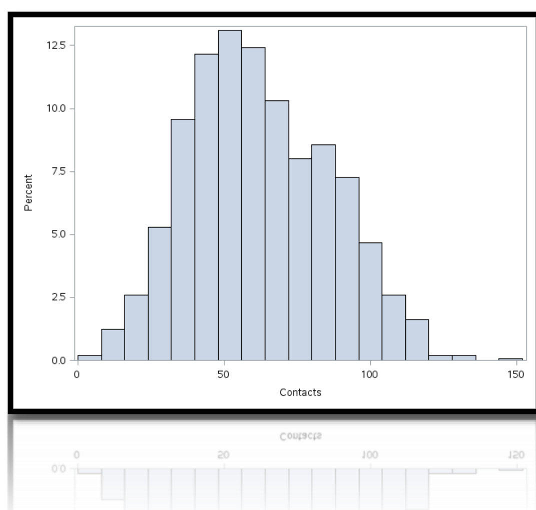
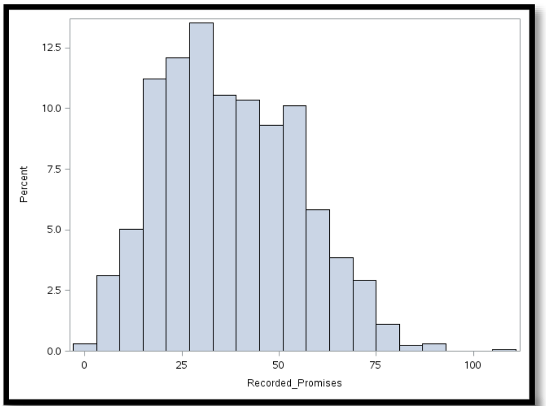


Figure 1 Call length distribution **Figure 2 Number of contacts distribution**



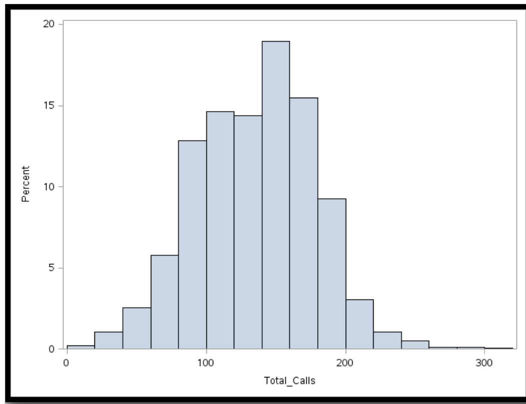
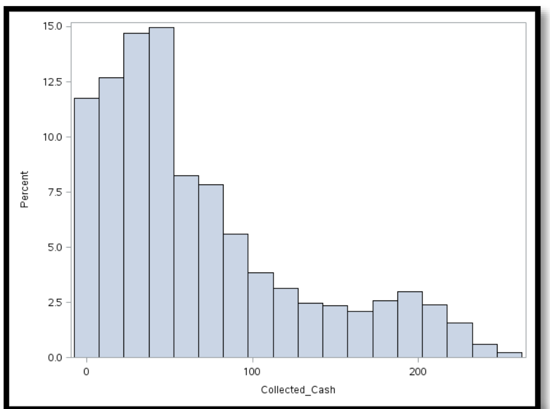


Figure 3 Number of recorded promises distribution. **Figure 4 Total number of calls distribution**



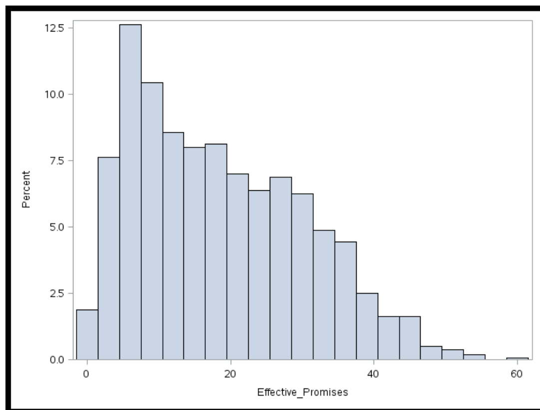


Figure 5 Collected cash distribution **Figure 6 Number of effective promises**

It is highly recommended to be careful defining inputs and outputs because results depends on a well definition of the variables. The preparation of the data sets, which include excluding missing values, recoding variables merging among others to then group by DMU's has to be done carefully. Take in account GIGO principle, if garbage is introduced to the model garbage will be obtained as result.

PROC OPTMODEL

The OPTMODEL Procedure was fundamental to solve the optimization program and the coding steps are described below.

1. Declare the sets.
2. Declare the variables, parameters and constraints.
3. Read data from the SAS data sets.
4. Formulate the optimization problem objective function and the constraints.
5. Execute the solve statement.
6. Print the variables to analyze the results.

A macro variable was created to assess the DMUs depending on the chosen demographic variable that describe agents.

```
%let DMU=age;/*age.gender.age_r.marital.edu.kids.*/

data inputs;
set  data;
keep &DMU
i_sum_Call_lenght
i_sum_Contacts
i_sum_Recorded_Promises
i_sum_Total_Calls;
run;
data outputs ;
set  data;
keep &DMU
o_sum_Collected_Cash
o_sum_Effective_Promises;
run;

data var_inputs;
input variable $32.;
datalines;
i_sum_Call_lenght
i_sum_Contacts
i_sum_Recorded_Promises
i_sum_Total_Calls
;run;
data var_outputs;
input variable $32.;
datalines;
o_sum_Collected_Cash
o_sum_Effective_Promises
;run;

proc optmodel ;
/* declare sets */
  set <str> data;
  set <str> inputs;
  set <str> outputs;
/* declare variables, parameters and constraints */
  str DMU;
  num X {data,inputs};
  num Y {data,outputs};
  var fi >= 0;
  var lambda {data}>= 0;
  num ho {data};
/* read data from SAS data sets */
  read data work.var_inputs  into inputs = [variable];
  read data work.var_outputs into outputs = [variable];
  read data work.inputs  into data = [&DMU] {i in inputs}  < X [&DMU,i] = col(i)>;
  read data work.outputs into data = [&DMU] {j in outputs} < Y [&DMU,j] = col(j)>;
/* formulate f.o. and constraints */
   max fo = fi;
   con input   {i in inputs}: sum{k in data} lambda[k] * X[k,i] <= X[DMU,i];
   con output {r in outputs}: sum{k in data} lambda[k] * Y[k,r] >= fi * Y[DMU,r];
/* execute */
  do DMU = data;
   solve;   ho [DMU] = 1 / fi.sol;
  end;
/* print variables to evaluate */
 print X Y lambda fi DMU ho;
quit;
```

RESULTS

The results of the optimization program gave us the most efficient profiles depending on each evaluated DMU. Presented on Figures 13 to 18 the bar charts below represent the objective function h_o, used to asses efficiency. If there is any value for h_o lower than one is possible to say that the DMU is not efficient compared to the other from the same group.

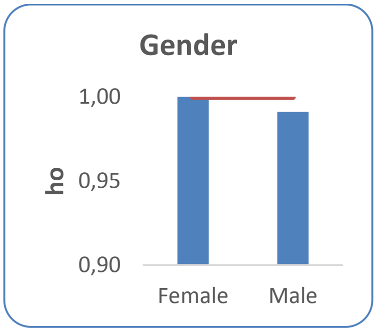


Figure 7 ho for Gender

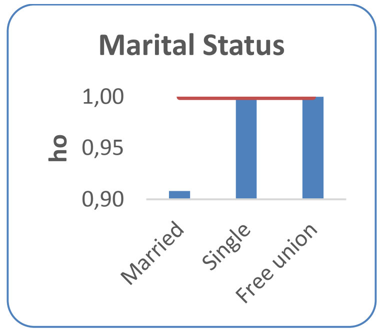


Figure 8 ho for marital status

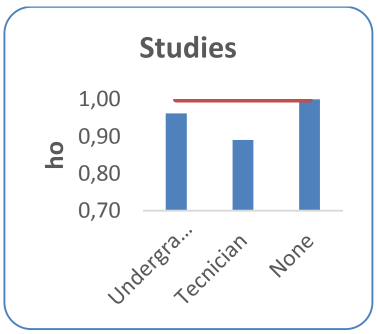


Figure 15 ho for studies

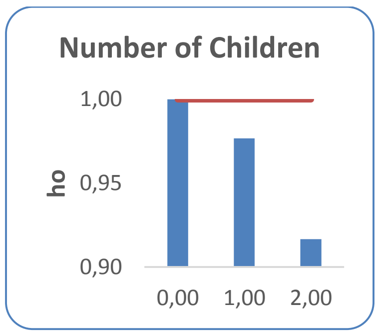


Figure 96 ho for n of kids

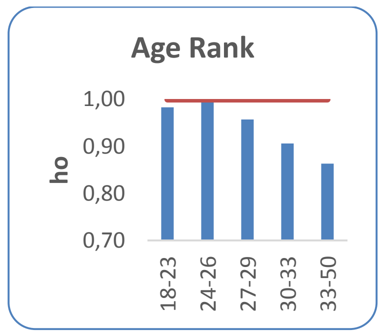


Figure 107 ho for age rank

The most efficient agent profiles suggested by DEA for gender were the Females, for marital status the singles, for studies the ones who didn't specify (none), for number of children cero, for age rank from 24 to 26 years old.

The SAS OPTMODEL Procedure output (in Display 1 below), print by default a Solution Summary specifying: the programming solver, the algorithm for solving the problem, the name of the objective function, the solution status, the objective value, the first, dual and bound infeasibilities, the number of iterations, the presolve time and the solution time. For this case we didn't have to specify the techniques to be used on the code because by default the solver is Linear Programing (LP), the algorithm Dual Simplex.

Using the demographic characteristics and efficient estimates from the example, we simulate new recovery levels and found that recovery will increase by 4,000 USD in 12 months in 4 DPD groups, which represents a recovery of 200,000 USD yearly if underperforming agents are given additional training. Savings represent 15USD by month and 175USD yearly. See Tables 4 and 5, also Figures 18 and 19.

Monthly Variable Costs		
Maximum capacity	400	agents
# Monthly Calls per agent	150	calls
Av time in each Call	1.5	min
Call cost per minute	0.10	USD
Agent salary	258	USD
TOTAL Variable Costs	103248	USD

Table 4. Monthly variable cost with best profiles

Monthly Variable Costs		
Before	After	Delta
$ 103,255	$ 103,248	-0.01%

Table 5. Monthly variable cost savings delta

Salto de página

The OPTMODEL Procedure

Solution Summary	
Solver	LP
Algorithm	Dual Simplex
Objective Function	fo
Solution Status	Optimal
Objective Value	1.0108124789
Primal Infeasibility	1.818989E-12
Dual Infeasibility	0
Bound Infeasibility	0
Iterations	7
Presolve Time	0.00
Solution Time	0.00

[1]	[2]	X	Y
F	i_sum_Call_lenght	2137.07	
F	i_sum_Contacts	60154.00	
F	i_sum_Recorded_Promises	35514.00	
F	i_sum_Total_Calls	130729.00	
F	o_sum_Collected_Cash		66900
F	o_sum_Effective_Promises		17499
M	i_sum_Call_lenght	832.88	
M	i_sum_Contacts	21888.00	
M	i_sum_Recorded_Promises	13119.00	
M	i_sum_Total_Calls	47204.00	
M	o_sum_Collected_Cash		23565
M	o_sum_Effective_Promises		6251

[1]	lambda
F	0.36108
M	0.00000

fi	DMU
1.0108	M

[1]	ho
F	1.0000
M	0.9893

Salto de página

Display 1. OPTMODEL Procedure results interface of SAS Enterprise Guide 6.1

The historical portfolio recovery levels by collection strategy, high risk, mid risk, low risk (defined by a logit model) are as shown in figures 19 and 20. We expect that with this project the recovery levels increase at least in 1 basic points for each month.

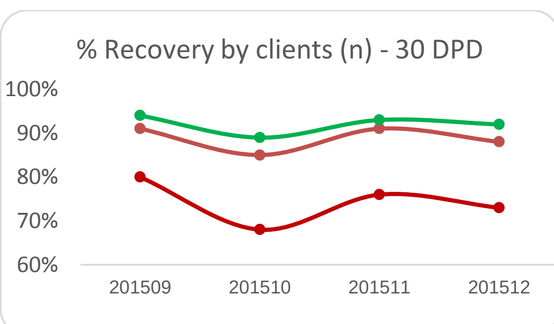


Figure 118 Recovery levels by balance

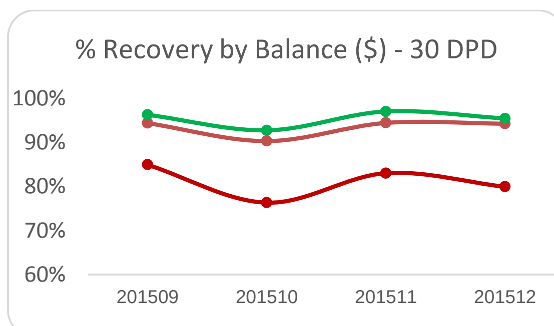


Figure 19 Recovery level by clients

CONCLUSION

Data Envelopment Analysis optimization technique, considering the dual output oriented model formulated without slack variables and inequality constraints, is a useful optimization methodology to asses efficiency of DMU's such as agent profiles in call centers.

Increase the recovery levels is possible with the same capacity but efficient people. Using Data Envelopment Analysis is revolutionary and useful and may provide future direction for training towards better collection efficiency.

The OPTMODEL Procedure was fundamental to solve the optimization program as the results of the optimization program gave us the most efficient profiles depending on each evaluated DMU.

Results depends on a well definition of the input and output variables. Preparation of the data sets (excluding missing values, recoding variables, merging sources) has to be done carefully. Chosen input and output variables should be consider the main goals of the business.

ACKNOWLEDGMENTS

To colleagues and managers for the sponsoring and ideas provided to develop this work.

RECOMMENDED READING

- *Building and Solving Optimization Models with SAS/OR*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Miguel Angel Díaz Rodríguez
Multibanca Colpatria, Scotiabank Group
Universidad de los Andes, Bogotá
+573152666888
ma.diaz216@uniandes.edu.co
https://co.linkedin.com/in/migueldr

Oscar Javier Cortes Arrigui
Multibanca Colpatria, Scotiabank Group
+573043611731
ojcortesa@unal.edu.co
https://co.linkedin.com/in/ojcortesa

REFERENCES

Adams, J. (2016). The Role of Third-Party Debt Collection in the U.S. Economy. *ACA International* , 6.

Boussofiane, A., Dyson, R., & Thanassouli, E. (1991). Applied data envelopment analysis. *European Journal of Operational Research*, vol. 52, pp. 1-15.

Charnes , A., Cooper, W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: an application of data envelopment analysis to Program Follow Through. *Management Science*, 668-97.

Charnes et alter. (1997). Data Envelopment Analysis: Theory, Methodology and Applications,. *New York, Kluwer Academic Publishers, Second edition*.

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 16.

Emrouznejad, A. (2000). An Extension to SAS/OR for Decision System Support. *Proceedings of 25th Sas International Conference*, 6.

Fedaseyeu, V., & Huntb, R. (2014). The Economics of Debt Collection: Enforcement of Consumer Credit Contracts. *Working Papers*, 46.

Fuentes, R. (2011, 11 28). Tecnicas econométricas en detalle. El análisis envolvente de datos. Alicante, Alicante, España. Retrieved from Universitat d'Alacant.

Lancheros, J., & Nieto, C. (2015). An Application of the DEA Optimization Methodology to Make More Effective and Efficient Collection Calls. *SAS Global Forum 2015*, 5.