

## Repositório ISCTE-IUL

---

Deposited in *Repositório ISCTE-IUL*:

2021-12-09

Deposited version:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Laureano, R. M. S., Trindade, G. & Laureano, L. M. S. (2021). Predictive models for managing financial incentives oriented to companies: Application to Portugal 2020. In Rocha, A., Gonçalves, R., Penalvo, F. G., & Martins, J. (Ed.), 2021 16th Iberian Conference on Information Systems and Technologies (CISTI). Chaves: IEEE.

Further information on publisher's website:

[10.23919/CISTI52073.2021.9476655](https://doi.org/10.23919/CISTI52073.2021.9476655)

Publisher's copyright statement:

This is the peer reviewed version of the following article: Laureano, R. M. S., Trindade, G. & Laureano, L. M. S. (2021). Predictive models for managing financial incentives oriented to companies: Application to Portugal 2020. In Rocha, A., Gonçalves, R., Penalvo, F. G., & Martins, J. (Ed.), 2021 16th Iberian Conference on Information Systems and Technologies (CISTI). Chaves: IEEE., which has been published in final form at <https://dx.doi.org/10.23919/CISTI52073.2021.9476655>. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

---

### Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

---

# Predictive models for managing financial incentives oriented to companies: Application to Portugal 2020

Raul M. S. Laureano

Instituto Universitário de Lisboa  
ISCTE-IUL, BRU, ISTAR  
Av. das Forças Armadas, Lisboa  
1649-026 Lisboa, Portugal  
[raul.laureano@iscte-iul.pt](mailto:raul.laureano@iscte-iul.pt)

Graça Trindade

Instituto Universitário de Lisboa  
ISCTE-IUL, BRU  
Av. das Forças Armadas, Lisboa  
1649-026 Lisboa, Portugal  
[graca.trindade@iscte-iul.pt](mailto:graca.trindade@iscte-iul.pt)

Luis M. S. Laureano

Instituto Universitário de Lisboa  
ISCTE-IUL, BRU  
Av. das Forças Armadas, Lisboa  
1649-026 Lisboa, Portugal  
[luis.laureano@iscte-iul.pt](mailto:luis.laureano@iscte-iul.pt)

**Abstract** — Since Portugal joined the European Union (EU) that it has been receiving incentives/funds to reduce disparities with other EU countries. Despite this goal, disparities between European regions still exist and the impact of such funds is questionable. What if it is possible to predict the success of such incentives when the funds are awarded to the beneficiaries? Using data from the database of The Agency for Competitiveness and Innovation (IAPMEI), for the programs National Strategic Reference Framework (QREN), and Portugal 2020, two predictive models are developed to estimate the number of applications to be received and the schedule of expected payments to beneficiaries for a four-month period. The results allow for a better prediction, on one hand, of the resources to be allocated to the evaluation process of the applications, and on the other hand, of the financial execution plan for the upcoming period, in order to prepare the financial execution.

**Keywords** – financial incentives; forecast.

## I. INTRODUCTION

The need for financing has always been a reality. To facilitate this demand, financial incentive tenders were created for companies, to grant financing without having to spend too much time and resources in a vast search for investment possibilities. From the investor's perspective, the creation of a bidding process for financial incentives ensures a better choice in the company where the investment will be made. Thus, the investor has an easier time evaluating each of the options, since these, through a tender, are calculated in comparison with each other, thus reducing their investment risk.

One of the investment programs created for this purpose was the “Portugal 2020”, which resulted from a partnership agreement between Portugal and the European Commission, covering the five European Structural and Investment Funds: European Regional Development Fund (ERDF), European Social Fund (ESF), Cohesion Fund (CF), European Agricultural Fund for Rural Development (EAFRD) and European Maritime Affairs and Fisheries Fund (EMFF). The program establishes economic, social, environmental, and territorial development policies, to promote growth and employment in Portugal, between 2014 and 2020.

The need to manage the portfolio of projects received and to forecast their success, supports the development of predictive models to ensure that these needs are efficiently fulfilled. If

Portugal does not execute the full amount allocated to it, it will have to return the unused funds to the EU. This is an issue that also needs to be considered.

The existence of forecasts allows to better program and plan the evaluation process of applications and allows to better estimate the financial execution and the expected load of the projects in the portfolio payment of applications, making the whole process more efficient using fewer resources and performed in less time.

The model for predicting demand in contests of Portugal 2020 will estimate the number of applications that will enter a contest for the chosen period, up to a maximum of one year, based on the historical patterns candidates evidenced in previous contests. Additionally, the model for estimating the amounts of payment requests (execution) will be based on the status of each project in execution, in view of the project portfolio held by the tender management company, and an execution plan to be verified in different time periods of no more than one year. Therefore, it will be possible to predict and prepare the expectations for the realization of the contest and the support to the companies in question.

The development of these models contributes to the literature on management of financial incentives and will serve as a groundwork to the development of similar models applicable to other calls and using different databases.

The remainder of the paper is organized as follows. The second section reviews the theoretical framework, where the topics considered relevant to the object of this study are identified and the main concepts addressed. In the third section the fundamentals for the execution of the work are discussed, which include data collection and treatment processes as well as the methods used. The fourth section presents the analyses and reports the results obtained and the difficulties experienced in the specification of the forecasting models. Finally, in the last section, the conclusions of this study and its limitations are presented, and possible steps for future research are suggested.

## II. LITERATURE REVIEW

### A. Evolution of the portuguese economy

The recession experienced in recent years in Portugal does not represent the future of economic activity in Portugal [1], as shown in Fig. 1.

However, the new crisis that all of us are living due to COVID-19 is having an extremely negative impact on the Gross Domestic Product (GDP) growth, namely in 2020 and probably in 2021, despite favorable perspectives as the vaccination process covers the worldwide population by the end of 2021.

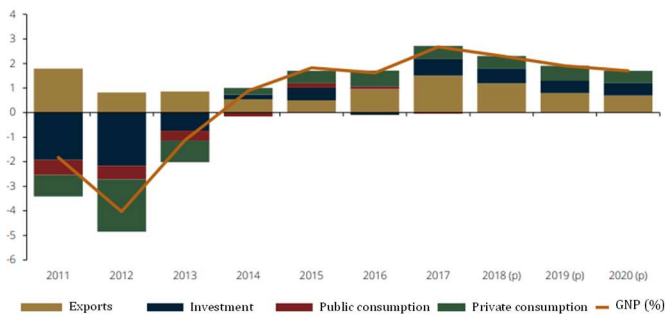


Figure 1: Net contributions to real GDP in Portugal  
Source: Banco de Portugal [1].

As shown in Fig. 2, the recession felt in 2014 had a negative impact on the increase in the number of insolvencies, despite the creation of new companies, which maintained a positive growth [2].

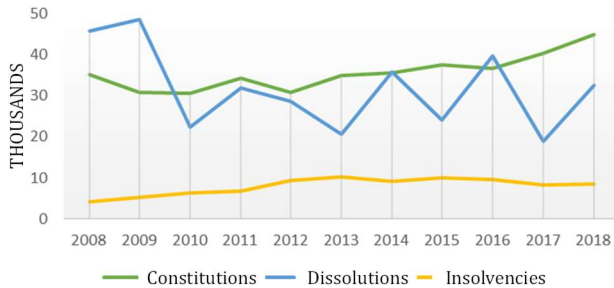


Figure 2: Constitution of legal persons and similar entities in Portugal  
Source: RACIUS [2].

The new reality is extremely negative, since many companies are closed and face bankruptcy, especially those related with tourism. However, the Portuguese Government is supporting the companies to minimize the insolvencies (e.g. layoff program and contribution to rentals) while the EU approved a huge financial incentive to be applied not only in helping companies, but also in public investments.

### B. Financial Needs of Companies in Portugal

If financing was essential for companies to be able to expand any part of their activity, nowadays, because of COVID-19, it is even more essential. However, organizations do not have the capacity, usually, to do this only with the internal resources they generate. Thus, external financing sources are needed. Also, the financing needs vary during the various growth phases of each company.

Despite the benefits for financing solutions such as self-financing, with a reduction in the dependence on external capital, companies end up choosing, mainly, bank credit as a source of financing [3]. However, there are risks associated with this type of financing since some companies are not able

to provide the required guarantees by the banks, ending up, not advancing with their investment projects [11][4].

### C. EU support programs in Portugal

One of the programs created by the Portuguese public entity for the development and support to micro, small and medium-sized companies (SMEs), was the Portugal 2020 program, which results from a partnership agreement between Portugal and the European Commission. This program brings together the activities of the five aforementioned European Structural and Investment Funds: i) CF - focused in reducing economic and social disparities and promoting sustainable development [5]; ii) ERDF - aimed at strengthening economic and social cohesion in the EU by bridging the imbalances between regions [6]; iii) EAFRD - whose goals are to foster the competitiveness of the agricultural sector, ensure the sustainable management of natural resources and climate action and achieve balanced territorial development of rural economies and communities, including the creation and maintenance of employment [7]; iv) ESF - whose objectives are to promote employment and support labor mobility, promote social inclusion and combat poverty, invest in education, skills and lifelong learning and improve institutional capacity and efficiency of public administration [8]; and v) EMFF whose objectives are to help fishermen in the transition to sustainable fishing, help coastal communities to diversify their economies, finance projects to create jobs and improve the quality of life European coasts and facilitate access to finance [9].

The Agency for Competitiveness and Innovation (IAPMEI) is responsible for supporting SMEs in the commercial, industrial, services and construction sectors, promoting competitiveness, business growth and supporting the internationalization of companies in these sectors in Portugal.

IAPMEI is present in a regional network of business support centers nationwide, in 12 cities, where, together with its website, it identifies products and services related to its responsibilities, such as financial incentive systems. These incentive and financing systems describe incentives and other financing solutions with direct or indirect intervention by IAPMEI, being available to companies for the development of their strategies.

There are currently three main incentive systems, corresponding to three domains of business development (Business Innovation and Entrepreneurship, Qualification and Internationalization of SMEs and Research and Technological Development), which encompass several subsystems, aimed at enhancing the development of national companies during the various phases of their life cycle and in their areas of competitiveness considered fundamental to operate in global markets [10].

The financial incentives currently available can take two forms: Non-Refundable Incentive (non-repayable financial support, subject to the objectives defined in the contract) and Refundable Incentive (interest-free loan, under repayment conditions defined in the contract).

#### D. Application and selection process

The Portugal 2020 Partnership Agreement is applied through IAPMEI as Operational Programs in addition to the European Territorial Cooperation Programs, in which Portugal will participate alongside other member states. These programs are managed by the various delegations of Managing Authorities for the operational programs (Norte, Centro, Lisbon, Alentejo, Algarve and Competitiveness and Internationalization Operational Program (POCI/Compete)) [11].

Applications to the programs are made through calls for tenders by opening notices which define the requirements and opening and closing calls for the financial support measures. For the application and selection process of the various programs, the registered candidates are defined through the competitions presented, and information about the programs that are running and the results of the programs already running can be found on the individual websites of each competition or on the website managed by IAPMEI [12]. The application is made at the 2020 counter until the deadline for the competition. After the end of the tender period, the evaluation of the project is carried out based on the terms of the notice of each tender. Then, the opinions are sent by IAPMEI to the managing authorities of the operational programs that then rank and select the candidates based on the score obtained during the evaluation process and until the budget limit determined by the competition is reached.

### III. METHODOLOGY

The methodology adopted is the CRoss Industry Standard Process for Data Mining (CRISP-DM), that implies a strict connection between the business and the data analyst, with the project being successful, when the model proves to be useful for the business [13].

Knowing the problem and the goals of the managing entity, namely the forecast of applications and the estimation of the financial execution, that is, the payments to be made to projects already approved and the expected load of payment requests for projects in the portfolio, the data understanding and data preparation are the next two CRISP-DM phases.

According with the business manager, it is necessary to consider the information available of the previous and current national programs, respectively, QREN (National Strategic Reference Framework) and Portugal 2020. Each program has its own system (Incentive System) for the Financial Incentive Program, and each system includes one or more measures (Incentive System Measures). Finally, each measure is presented in competition through a notice, in which, in some cases, it still has several phases.

#### A. Data

The data were provided by the managing entity of the financial program of Portugal 2020, IAPMEI, and reflect the raw information of the projects executed or in execution present or not in the portfolio that were approved in the contests associated with the Portugal 2020 programs and the QREN.

The data provided were already divided into two datasets. The set of data referring to the applications (dataset A) made available consists of 37461 records with 87 variables, which needed to be processed to fit the objective of explaining the forecast of the number of applications per measure. The dataset of payments (dataset B) request is comprised of 52916 payment records and 30 variables, which needed to be processed to fit the purpose of explaining the amount.

The chosen variables to be analyzed, with the monitoring and approval of the business specialist, are presented in Table I. After the quality analysis (e.g., errors, missing-values and outliers), at the suggestion of the business specialist, records that presented an application status in its life cycle were filtered as "Request for Assistance", thus resulting in a final total of 37188 records.

The variables analyzed and included in the dataset B, with the monitoring, supervision, and approval of the business are presented in Table II. For these variables, their metadata was then organized and analyzed again in terms of quality. During this evaluation, at the suggestion of the business specialist, payments whose payment decision direction was not "Favorable" were filtered from payment requests, resulting in a total of 45514 payment records with a "Favorable" decision. Some of these payments still did not have a payment date and were excluded, resulting in a final total of 33266 records of payments made.

TABLE I. DESCRIPTION OF THE VARIABLES IN THE DATASET A

<i>Variables</i>	<i>Description</i>	<i>Classification</i>
Number of applications	Identification of the application	Nominal qualitative
Program	Financial incentive framework program	Qualitative dichotomous nominal
NIF	Promoter tax identification number	Nominal qualitative
System	Financial incentive program system	Nominal qualitative
Measure	Measure of the incentive system	Nominal qualitative
Notice	Notice of tender for the incentive scheme measure	Nominal qualitative
Phase / Contest	Phase of the competition to which the candidate submitted	Ordinal qualitative
Status	Status of the application in its life cycle	Nominal qualitative
Application date	Promotor application date	Date
NUTS II	Project location of the candidate company in territorial unit 2	Nominal qualitative
Total approved incentive	Total incentive approved	Continuous quantitative

In the data preparation phase and to meet the mentioned objective, the variable "Four-month period" was created based on the variable "Payment date". It was then decided to proceed with the evaluation of the number of times that each call had opened to check if there would be feasibility for forecasting the number of applications per call.

TABLE II. DESCRIPTION OF THE VARIABLES IN THE DATASET B

Variables	Description	Classification
Program	Financial incentive framework program	Qualitative dichotomous nominal
Application	Identification of the application	Nominal qualitative
System	Incentive system of the financial incentive program	Nominal qualitative
Measure	Measure of the incentive system	Nominal qualitative
Notice	Tender notice of the incentive scheme measure	Nominal qualitative
PO	Operational program name	Nominal qualitative
Contracted incentive	Incentive contracted with the beneficiary company	Continuous quantitative
Incentive to pay	Incentive to pay the beneficiary company	Continuous quantitative
Payment date	Incentive payment date	Date

### B. Modelling

After the descriptive statistics, the beginning of the analysis proceeds with the selection of data modeling techniques, Pearson coefficients, the multiple regression models, and autoregressive models with the goal of forecasting future data, applying the “Bottom-Up” methodology to prepare the dataset. After choosing a model that passes the validity and quality tests applied, this must be reviewed by the business specialist interested in the model to ensure that the model meets the expectations of the business for which it is developed.

## IV. RESULTS

The forecast of the total number of applications for each measure of each program, taking into account the confidence margin, with a 95% accuracy rate, where the forecast will be made for the next several competitions, divided individually, according to what was requested by the business representative.

The forecast of the expected execution value for the project portfolio for a period of less than one year whose objective is a 95% hit rate, taking into account the confidence margin, where the forecast will be made for several subsequent time periods divided in four-month terms, as requested by the business representative.

### A. Forecast of the total number of applications

In order to predict the number of applications per four-month period, the records obtained from the applications were grouped, by analyzing the values of the variables and the proximity to the final variable, following the principle of Occam's razor [14]. It started with the simplest model and then other independent variables were added to the model to increase its explanation power while trying to keep it as simple as possible.

A first model was created (Model 1). In order to understand the relationship between the number of applications and the number of candidates, it was included the periods year, month, and four-month period and it was estimated the correlation coefficient between the sum of notices and the number of candidates for each of temporal period.

It was estimated a weak/moderate linear relationship between the variables in the sample ( $Pearson(27) = 0.427$ ),

but, in the population, the variables are also correlated ( $p - value = 0.026$ ).

In Fig. 3 it is possible to visualize the existence of two outliers (cases 17 and 14 in the dataset) that were eliminated from the estimations.

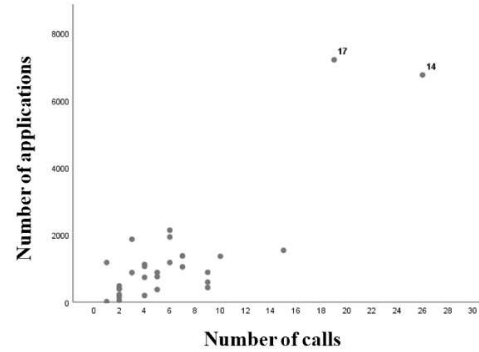


Figure 3 – Graphical representation between the number of applications and the number of calls per four-month period

Then, the assumptions of the simple regression model were validated in Annex. The estimated model is the following:

$$\hat{Y}_{i1} = 479.372 + 76.089 X_{i1} \quad \text{Eq. (1)}$$

(0.019)      (0.026) *p - values*

$R^2 = 0.182$

where the Number of applications (per 4-month period)=  $Y_i$  and the Number of calls (4month period)=  $X_i$

For one increase in the sum of notices per four-month period, it is estimated that the number of candidates increases about 76,1 applications, in average.

A second model (Model 2) was estimated. The variable time was added to the specification in the previous model following the forecasts described in [15]. It is believed that there is a causal relationship with previous values, in the sense that a candidate does not apply again with the same project in the next opening period, knowing that it has already been approved. As such, a new autoregressive independent variable [AR (1)] with the immediately previous period was added to the first model. The validation of the assumptions of the model are presented in Annex.

$$\hat{Y}_{i2} = 603.793 + 78.298X_{i1} - 70.176 AR(1) \quad \text{Eq. (2)}$$

(0.092)      (0.027)      (0.612) *p - values*

$R^2 = 0.182 \quad VIF = 1.003 \quad DW = 2.106$

The variable time is not significant in explaining variations in the number of applications (per four-month period).

Finally, a third model (Model 3) was created. It is also possible that there is a seasonality relationship with the immediately preceding period (the first four-month period of the year and the first four-month period of the following year). To test this possibility, another independent variable was added, the [SAR (1)], which led to the following model:

$$\hat{Y}_{i3} = 714.917 + 76.081 \hat{X}_{i1} - 96.435 AR(1) - 7.592 SAR(1) \quad \text{Eq. (3)}$$

(0.255)      (0.052)      (0.564)      (0.612) *p-values*

$\bar{R}^2 = 0.074 \quad VIF = 1.014; 1.158; 1.171 \quad DW = 2.106$

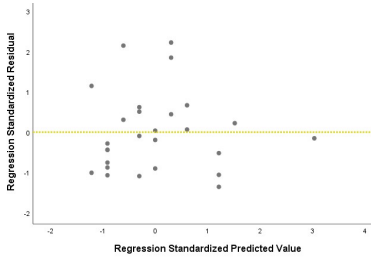


## VI. ANNEX

### A. Forecast of the total number of applications

#### First model (Model 1) assumptions:

- (1) Linearity of the relationship: it was already validated with the Pearson correlation.
- (2) Homoscedasticity or constant random error variance



The circles are randomly dispersed towards zero, thus the assumption is validated.

- (3) Normality of the random errors

#### Tests of Normality

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual	.103	27	.200*	.919	27	.058

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

For a level of significance of 0.05, this assumption is validated.

- (4) Independence between random errors for different periods

#### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.427 <sup>a</sup>	.182	.150	540.915	2.135

a. Predictors: (Constant), N\_Calls

b. Dependent Variable: N\_Applications

RA:  $[d_U; 4 - d_U] \approx [1,45; 2,55]$

DW = 2,135  $\in [1,45; 2,55]$ . Therefore, there is not an autocorrelation problem of the random errors.

#### Second model (Model 2) assumptions:

- (1) Independence between random errors for different periods

#### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.445 <sup>a</sup>	.198	.129	556.222	2.106

a. Predictors: (Constant), Time\_1lag, N\_Calls

b. Dependent Variable: N\_Applications

- (2) Independence between independent variables

The Variance Inflation Factor (VIF) greater than 5, indicates the presence of the problema of multicollinearty [16]. Thus, this assumption is validated.

- (3) There is not a problem with the autocorrelation of the random errors since the Durbin-Watson test is close to 2.

### B. Forecast the four-month execution of the projects

#### First model (Model 4) assumptions:

- (1) Linearity of the relationship: it was already validated with the Pearson correlation.

- (2) Independence between random errors for different periods

#### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.996 <sup>a</sup>	.993	.992	14773190.5391	1.534

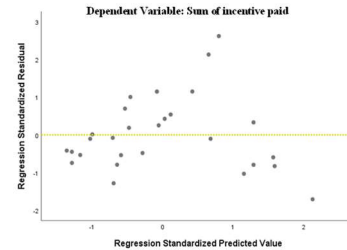
a. Predictors: (Constant), Sum of approved incentive

b. Dependent Variable: Sum of incentive paid.

Acceptance area:  $[d_U; 4 - d_U] \approx [1.49; 2.51]$

DW = 1,534  $\in [1.49; 2.51]$ , meaning that this assumption is validated.

- (3) Homoscedasticity or constant random error variance



Assumption is validated.

- (4) Normality of the error terms

#### Tests of Normality

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual	.130	28	.200*	.945	28	.149

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

As  $SW(28) = 0,945; p - value = 0,149$ , it is concluded that the distribution of the random errors follow a normal distribution.

#### Second model (Model 5) assumptions, adding the variable time in the specification model:

- (1) Linearity of the relationship.
- (2) Independence between random errors for different periods

#### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.998 <sup>a</sup>	.995	.995	12169728.4771	2.041

a. Predictors: (Constant), Time, Sum of approved incentives

b. Dependent Variable: Sum of incentives paid.

DW = 2.041  $\approx 2$ , meaning that this assumption is validated.

- (3) Independence between independent variables

Model		Unstandardized Coefficients			Collinearity Statistics		
		B	Std. Error	t	Sig.	Tolerance	VIF
1	(Constant)	-2847843	4329223	-.658	.517		
	Sum of approved incentive	.489	.021	23.220	.000	.138	7.248
	Time	2180088	597469.	3.649	.001	.138	7.248

This assumption is not validate since the two independent variables are correlated ( $VIF = 7.248$ ). Therefore, the estimations of the coefficients in Model 4 for the second goal are good enough.