Audits and logistic regression, deciding what really matters in service processes: a case study of a government funding agency for research grants

Samohyl, Robert

Universidade Federal de Santa Catarina, Departamento de Engenharia de Produção

August 2012

Audits and logistic regression, deciding what really matters in service processes. A case study of a government funding agency for research grants. [1]
(PRELIMINARY)

Robert Wayne Samohyl, PhD
(samohyl@deps.ufsc.br)

ABSTRACT

Governmental agencies, the back office of private firms and nongovernmental organizations experience bureaucratic processes that are often repetitive and out-of-date. These imperfections cause resource misuse and support activities that diminish to the value of the process. An important element of these bureaucratic processes is checking whether certain projects approved by the office have actually been successful in their proposed objectives. Banks and credit card companies must evaluate whether creditors have fulfilled their supposed financial worthiness, tax authorities need to classify sectors of the economy and types of tax payers for probable defaults, and research grants approved by government funding agencies should verify the use of public funds by grant recipients. In this study, logistic regression is used to estimate the probability of conformity of research grants to the financial obligations of the researcher analyzing the correlation between certain characteristics of the grant and the grant´s final status as approved or not. The logistic equation uncovers those characteristics that are most important in judging status, and supports the analysis of results as false positives and false negatives. A ROC curve is constructed which reveals not only an optimal cutoff separating conformity from nonconformity, but also discloses weak links in the chain of activities that could be easily corrected and consequently public resources preserved.

KEYWORDS: Logistic regression; ROC curve; probability; audits; government; research grants

1. INTRODUCTION[2]

This article deals with the problem of auditing formal office procedures that classify whether or not a given project that has already been completed by its authors should actually receive final approval from the office staff. If the project were classified by the staff as nonconforming, then it would be returned to its authors for reformulation and eventual reevaluation by the staff. On the other hand, those projects that are approved are passed on to finalization including in many cases final storage. Picture if you will the process of evaluating credit applications. Credit is extended to the client of a

[1] Aplicação de conceitos de controle estatístico e gestão da qualidade para a melhoria dos processos de prestação de contas de projetos de pesquisa, fatores determinantes da aprovação de prestação de contas de projetos de pesquisa. Termo de Outorga 7003/2011-3, coordinator Robert Wayne Samohyl, PhD, Industrial Engineering, Federal University of Santa Catarina, Brazil, Núcleo de Normalização e Qualimetria.

[2] Thanks to Armin Koenig for some very helpful comments.

financial institution only after the undertaking of a formalized review process of creditworthiness. After a certain period of time has elapsed, and sufficient history has been acquired on the client's activities as debtor to the institution, the client´s creditworthiness is investigated, comparing the degree of creditworthiness originally assigned to the client with the results of the historical analysis. Hopefully the original evaluation should compare successfully with the actual facts, if approval were given then the client's credit history should manifest conformity. Another example is the governmental process of reviewing tax returns.

In general terms, this kind of verification of a classification scheme usually depends upon the verification of several forms filled in with specific information and corroborated by relevant documentation. Information is checked and documents are authenticated. It is common practice to use a checklist of the most important items in the form and corroborating documents that guides the process indicating inadequacies, and if the project is rejected it is returned to its authors for corrections. A checklist may have tens or even hundreds of items. Naturally, if the checklist is the result of an historical evolution within the institution encompassing changes in management and technologies, through time it may have become repetitive and cumbersome, substantiating the need for an audit and analysis of procedures leading to improvements in the process, the reformulation of checklist items in light of their ability to classify conforming and nonconforming projects. The in-house investigation of internal processes is called an internal audit, usually applied sporadically with the intention of uncovering activities and other elements of key processes that no longer produce value for the institution, in the case studied here to verify the correspondence between the checklist and the true state of the project.

In this article we offer a procedure based on logistic regression that identifies the degree of correlation among items of the checklist and the consequent approval or rejection of the financial accounts a given project which in turn is revealed as actually conforming or nonconforming. Along the way, the use of the receiver operating characteristic (ROC) curve programmed in the R language (R Development Core Team (2012))[3] will be essential not only to evaluate the procedures under study but also to suggest changes for a simpler and more valuable process. ROC analysis has been advanced mostly in the medical science literature (R. Kumar and A. Indrayan, 2011), however with the development of several R packages (T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer 2005) which simplify applications, the use of the ROC methodology has been spreading into other areas (O. Komori, 2009).

Quality control guidelines are easily adapted to the question of internal audits. Office bureaucracy in both the private and public sectors is analogous to an industrial process, following a certain

[3] R package ROCR is especially important for the development of this article.
All calculations were done with the R language under the GNU license. Where it seemed appropriate, comments and R code are given.

number and kind of procedures defining activities and responsible personnel who receive inputs usually in the form of verbal and written communications that are processed and then delivered to the next step of the bureaucratic process (Fugee Tsung, Y. Li, and M. Jin, 2008) . Audits evaluate these procedures in order to eliminate unnecessary steps in the process or combine steps to economize on resources, consequently aproximating judgements from the staff to the reality under scrutiny (M. Cecchini, H. Aytug, G. J. Koehler, and P. Pathak, 2010).

In the case studied here, internal auditors in the public sector evaluate the financial side of research grants. They look to establish the conformity of the financial activity of the research project, checking for the proper use and transcription of grant finances. Based on a checklist of important items, the financial events of the project are either approved or rejected, and the conformity of project finances classified. It should be mentioned that the actual scientific results of the research project are evaluated in a separate audit, and consequently are not part of the discussion of this article.

2. DATA

The data base is composed of 540 research projects (observations) and the respective financial accounts from a funding agency on the State level which allocates funds to research projects. These research projects have already suffered preliminary internal audits and have been judged either as conforming or nonconforming by the internal auditors. Consequently the data base is constructed from a primary data source which is the original forms, bank statements and other documents submitted by the researchers at the end of their research activities. A summary of the data can be found in table 1.

Approvals: Financial accounts approved (s) or rejected (n). Primary response variable	n = non conformity, rejected	166
	c = conformity, successfully approved	367
	NA = not available; missing	7
Value (in local currency – Reals) first installment	Min.	9,90
	Max.	1,264,781.00
Total value of grant	Min.	9,90
	Max.	2,394,961.00
fisjur: Beneficiary is a private person or a legal entity	fis (private person)	513
	jur (legal entity)	27
Project budget is for expenses or equipment	Equipment	167
	Expenses	373
Project coordinator has a doctorate	n (no)	102
	s (yes)	401
	NA	37
Type of grant program	Universal	203
	over-the-counter	61
	events	50
	Agriculture	40
	Youth researchers	29
	others	154
	NA	3
Institution where project is located	UFSC	163
	private business	46
	UNIVALI	46
	UNISUL	29
	FURB	27
	other	194
	NA	35
City where project is located	Florianópolis	198
	Itajaí	37
	Joinville	36
	Blumenau	33
	Chapecó	30
	other	163
	NA	43
Item 1, Art. 37 paragraph – Financial accounts should be divided into categories of either expenses or equipment and should be separated into two distinct files and should receive different process numbers.	n = no, item did not pass	13
	s = item successfully passed	519
	NA	8
Item 2, Art. 37 I – Scientific technical report has been archived at the URL of the funding agency?	n	47
	s	169
	NA	324
Item 3, Art. 37 II – Balance sheet (official form TC 28) has been signed and archived at the official government URL.	n	59
	s	475
	NA	6
Item 4, Art. 37 III – All receipts, boarding passes and any other	n	23

fiscal documents arranged in chronological order.	s	232
	NA	285
Item 5, Art. 37 IV – All monthly bank statements containing all bank balances and financial movement, from the first deposit until the closing of the account.	n	39
	s	121
	NA	380
Item 6, Decreto 2.060/09,art.37, V – If necessary, uncashed checks (but written) separately listed and accounted for.	n	122
	s	397
	NA	21
Item 7, Art. 37 VI – grant funds not utilized, returned to the funding agency and appropriately verified.	n	6
	s	30
	NA	504
Item 8, Art. 37 VII – All bank deposits must be accounted for and all appropriate forms signed by all parties.	n	115
	s	57
	NA	368
Item 9, Art. 37 VIII – If necessary, verification that equipment has been donated to the research institution.	n	119
	s	30
	NA	391
Item 10, Law 4320/64,art.63. Verification of payments for services and consulting.	n	103
	s	396
	NA	41
Item 11, Decreto 2.060/09, Art. 42. In order to finalize the closing out process of the entire research grant, both financial accounts and the final scientific technical report must be appropriately approved.	n	120
	s	118
	NA	302
Item 12, Copy of contract with the funding agency and agreed work plan	n	119
	s	401
	NA	20
Item13, Verification of extraordinary daily expenses, if necessary.	n	10
	s	91
	NA	439

Chamada.Pública, a specífic grant program			number of projects
001/04			13
001/06			30
001/09			29
001/10			32
002/05			19
002/06			3
002/08			2
002/09			31
002/10			1
003/06			224
004/04			1
004/07			28
004/08			8
004/09			4
005/09			4

006/07			1
007/06			21
007/08			1
007/09			8
008/06			17
008/09			5
009/09			2
010/09			1
012/09			8
013/09			10

Table 1. Financial and demographic statistics.

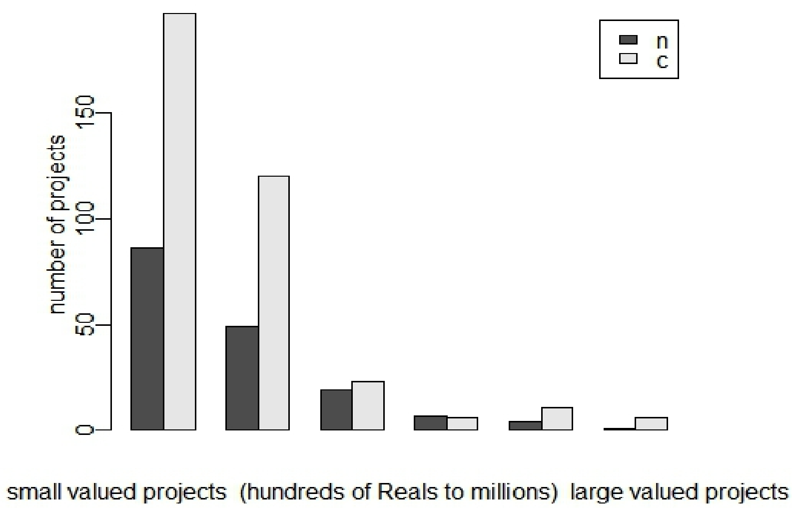

Figure 1. Conforming and nonconforming projects by total value of the grant.

Among the variables that describe some important aspects of the research project, financial characteristics are contained in the 13 items from the checklist of obligations which lists the relevant items necessary for closing out the financial accounts of the grant. The checklist is an internal document for the internal auditors and is not seen by the researcher. Figure 1 shows the number of projects that were classified as conforming (c) or nonconforming (n) by the office staff by value of the grant, the third entry in the list of data in table 1. An intuitive look at figure 1 shows the absence of a strong relationship between value of the grant and its final disposition as conforming or not. Regardless of the value of the grant, conforming projects hover around 70% of the total, easily seen in figure 1. This result is later supported by the logistic regression in section 4.

3. CONTINGENCY TABLE.

In the following tables the major problem of the process under study becomes apparent, that in most cases there is little apparent correlation between the verification of the financial accounts of the grants and the items of the checklist, suggesting that the checklist in its present form needs revision. Later on in this article we determine through logistic regression the most important items of the checklist as predictors of the probability of approval limited to a small number of checklist items not always with the desired causality direction, and almost all other variables are statistically insignificant.

In the contingency tables of table 2 we test the relationship between the conformity (c) or nonconformity (n) of the grant as related to the successful approval (s) or nonapproval (n) of a specific item from the checklist.[4] The column marked "n" define the number of nonconforming grants and "c" conforming grants. Rows define whether an individual item from the checklist was approved or not. Entries in the table show the degree of consistency between grant approval and relevance of a checklist item. In the northwest corner of the tables we count the number of times a specific item was not approved (n obs), however the respective financial accounts were judged as conforming (c). In principle but not always the case, a rejected item should lead to a nonconforming grant. For instance, verify the values in the last contingency table for item 12, where 53 grants while judged as conforming (c) did not pass approval for item 12 (n obs). Clearly the question is why was the grant judged as conforming given the rejection of item 12? Inconsistent results appear in all the items, which would lead us to believe that the checklist needs revision. The numbers in the northwest entry are consistent results showing the rejection (n obs) of the item and the nonconforming (n) nature of the grant. The numbers in italics are expected values based on the proportions of conforming grants and approvals of the checklist item. Essentially, if proportions were 50%, then each entry would be 25% of the total, assuming randomness and independence. The chi-squared for each table measures the randomness of the entries in the table by comparing expected frequencies to observed. There is strong evidence of randomness for items 3 with p value practically equal to 1,0, and somewhat weaker evidence for item 4. In other words, for these items the result of the classification procedures are no better than a random draw. All other items reject the randomness assumption. In fact, for the column of nonconforming grants, observed item rejections (n obs.) are more numerous than the expected value suggesting that office procedures are at least better than a random draw. The first contingency table relates conformity to the academic qualification of the coordinator of the research project, whether she has a doctorate. The observed count for nonconformity and no doctorate is 28 which is very close to the expected value of 32, the expected value of a random draw. In fact, all expected values are very close to the observed for the conformity/Doctorate contingency table which leads to the chi-squared test not rejecting the hypothesis of a random draw. The chi-squared test suggests that conformity is independent of the

[4] All 12 items from the checklist were not present in the table both to be concise and emphasize only items that were chosen as relevant by the logistic equation applied in subsequent sections. For more on contingency tables see chapter 2 of Agresti, A (2002).

doctorate. The logistic regression estimates will elaborate on this result.

Conformity				**Conformity**			
Doc	n	c	Total	**Item.5**	n	c	Total
n obs.	28	**74**	102	n obs.	41	**44**	85
n expected	*32*	***70***		n expected	*27*	***58***	
s obs.	125	267	392	s obs.	123	313	436
s expected	*121*	*271*		s expected	*137*	*299*	
Total	153	341	494	Total	164	357	521
	0.310	0.690			0.315	0.685	
Pearsons Chi-squared with Yates correction				Qui^2 = 12.31	g.l. = 1	**p = 0.0004**	
Qui^2 =0.55	g.l.= 1	**p = 0.45**					
Item.2	n	c	Total	**Item.6**	n	c	Total
n obs.	93	**91**	184	n obs.	48	**74**	122
n expected	*57*	***127***		n expected	*37*	***85***	
s obs.	70	270	340	s obs.	109	284	393
s expected	*106*	*234*		s expected	*120*	*273*	
Total	163	361	524	Total	157	358	515
	0.310	0.690			0.305	0.695	
Qui^2 = 48.6	g.l.= 1	**p = 0.000**		Qui^2 = 5.38	g.l. = 1	**p = 0.02**	
Item.3	n	c	Total	**Item.10**	n	c	Total
n obs.	19	**40**	59	n obs.	47	**55**	102
n expected	*19*	***41***		n expected	*32*	***70***	
s obs.	147	324	471	s obs.	110	283	393
s expected	*148*	*324*		s expected	*125*	*268*	
Total	166	364	530	Total	157	338	495
	0.313	0.687			0.317	0.683	
Qui^2=0.000	g.l.= 1	**p = 0.99**		Qui^2 = 11.41	g.l. = 1	**p = 0.0007**	
Item.4	n	c	Total	**Item.12**	n	c	Total
n obs.	15	**18**	33	n obs.	65	**53**	118
n expected	*10*	***23***		n expected	*36*	***82***	
s obs.	145	343	488	s obs.	93	305	398
s expected	*150*	*338*		s expected	*122*	*276*	
Total	160	361	521	Total	158	358	516
	0.307	0.693			0.306	0.694	
Qui^2 = 2.89	g.l.= 1	**p= 0.0887**		Qui^2 = 41.62	g.l. = 1	**p = 0.0000**	

Table 2. Contingency tables of the relation of conforming and nonconforming grants and some

specific items of the checklist.

4. THE LOGISTIC EQUATION AND ESTIMATION.

Among several possible alternatives for estimating the effect of a variable on the probability of the existence of another, logistic regression is one of the most widely used. Consider a variable Y which is essentially binomial having two possible states as, for example, conforming and nonconforming. At the close of a research project, the financial accounts of a research grant can be either approved as conforming or rejected as nonconforming by the funding agency. The probability of conformity is written as *P(c)*. In turn, the value of *P(c)* may depend on a host of variables that characterize the project such as the approval of the items in the checklist, or on predictor variables such as the monetary value of the grant or the institution where the project was undertaken. among many others. The predictor variables X_i may be quantitative or qualitative.[5] The logistic equation in simple linear form, where an interaction term[6] has been included $X_1 * X_2$, is

(1)
$$\ln \frac{P(c)}{1-P(c)} = a + b_1 X_1 + b_2 X_2 + b_{12} X_1 * X_2$$

The left hand expression is the logit also known as the log odds[7]. Theoretically the logit can vary between positive and negative infinity as *P(c)* varies from 1 to 0. When the logit is null, *P(c)* is 0.5. To better appreciate the nonlinear relationship between *P(c)* and the predictor variables, equation (1) can be rewritten as

(2)
$$P(c) = \frac{e^{a + b_1 X_1 + b_2 X_2 + b_{12} X_1 * X_2}}{1 + e^{a + b_1 X_1 + b_2 X_2 + b_{12} X_1 * X_2}}$$

The regression procedure will estimate the values of the coefficients, and then an estimate of *P(c)*, the response variable, for specific values of X_i. Most of the variables in the data base are binary. Furthermore, the logistic regression should evaluate interaction terms and other non linearities if necessary. Consequently, the principal objective of the analysis is to determine the influence of a predictor variable X_i on the value of *P(c)*. The estimated equation represents the value of the probability of project conformity when a specific items in the checklist has been approved or other characteristics are present.

The logistic equation was estimated using several different subsets of variables, and criteria for selecting the best subset were based on judgement of the individual significance of each coefficient

---

[5] Here we will not go into detail however several references on logistic regression are given here. The original methodology is from the 18th and 19th century (Cramer, 2002). The first modern application is Reed, L. J. and J. Berkson (1929)..A much respected text is Agresti(2002).

[6] Statistical results presented below have an interaction term.

[7] odds = *P(c)*/(1 – *P(c)*), consequently *P(c)* = odds/(1+odds).

and the AIC of the overall equation. The subsets included several interaction terms.[8] In the next table 3, the results from the logistic regression are reported for the coefficient estimates and other appropriate statistics. While some of the results are not significant at the traditional 5% level, these variables were not eliminated from the model for other reasons, such as the value of the AIC or due to the importance of the variable in the interaction terms. Surprisingly, even though there is a total of 13 items in the checklist, only six items are correlated with the approval or rejection of the grant ´s financial accounts. Moreover, two other variables are important, whether the coordinator of the project has a doctorate (Doc) and the specific details of each grant program (CP*number*). Model fit is adequate suggested by the values of the residual deviance (393) and degrees of freedom (415).

	With interactions				No interactions			
Coefficients:	Estimate	Std.Error	z value	Pr(>|z|)	Estimate	Std.Error	z value	Pr(>|z|)
Intercept	-0.358	1.350	-0.265	0.791	-2.600	0.619	-4.203	0.000
Doc	-3.387	1.454	-2.329	0.020	-0.426	0.318	-1.343	0.179
CP00106	1.822	0.947	1.924	0.054	1.908	0.934	2.044	0.041
CP00109	1.163	0.581	2.000	0.045	1.004	0.560	1.792	0.073
CP00110	2.534	0.823	3.079	0.002	2.135	0.791	2.700	0.007
CP00205	1.469	1.015	1.447	0.148	1.286	0.958	1.343	0.179
CP00407	1.763	0.693	2.543	0.011	1.592	0.683	2.332	0.020
CP01209	-1.921	0.842	-2.281	0.023	-1.411	0.762	-1.853	0.064
CPother	-1.597	0.957	-1.670	0.095	0.449	0.493	0.910	0.363
Item.12	3.229	0.957	3.372	0.001	1.373	0.302	4.542	0.000
Item.10	-2.682	1.143	-2.347	0.019	0.557	0.316	1.765	0.078
Item.2	2.288	0.661	3.460	0.001	1.063	0.262	4.056	0.000
Item.3	-1.953	1.041	-1.876	0.061	-0.216	0.402	-0.536	0.592
Item.5	1.778	0.378	4.704	0.000	1.437	0.344	4.176	0.000
Item.6	0.553	0.289	1.913	0.056	0.526	0.279	1.887	0.059
Doc:CPother	2.778	1.159	2.396	0.017				
Doc:Item.12	-1.979	1.008	-1.962	0.050				
Doc:Item.10	3.698	1.199	3.084	0.002				
Doc:Item.2	-1.454	0.718	-2.026	0.043				
Doc:Item.3	1.975	1.151	1.716	0.086				
Null deviance:		533.97	434 df		Null deviance:		533.97	434 df
Residual deviance:		393.57	415 df		Residual deviance:		417.09	420 df
AIC:	433.57				AIC:	447.09		

Table 3. Logistic regression results with and without interactions.

The two regressions of table 3 are actually very similar, even if not very apparent at first glance. The first equation, which includes interaction terms, has been chosen as the most representative and its results will be utilized in the analysis of the next sections. The second equation has no interaction terms but is constructed with the same factors. Compare the coefficient values between the two equations for the factors which do not enter the interaction terms, in the table lightly shaded to distinguish them. Results are very similar between the two estimated equations. Furthermore, since the first equation produces better results in terms of the residual deviance and the AIC, the importance of the interaction terms is corroborated. Here is the explicit equation, coefficients taken from table 3.

[8] This is standard procedure for variable selection, see Gelman et al(2004).

$$(3)\quad \ln\frac{P(c)}{1-P(c)} = -0.358 - 3.387Doc + 1.822CP00106 + 1.163CP00109 + 2.5337CP00110$$

$$+ 1.469CP00205 + 1.7631CP00407 - 1.9211CP01209 - 1.5973CPother + 3.2287Item12$$

$$- 2.6816Item10 + 2.2879Item2 - 1.9525Item3 + 1.7779Item5 + 0.5525Item6 + 2.7776Doc*CPother$$

$$- 1.9785Doc*Item12 + 3.6977Doc*Item10 - 1.4535Doc*Item2 + 1.9751Doc*Item.3$$

The estimated equation contains a total of 14 binary variables: 6 items from the checklist, 7 grant programs, and the doctorate. Exactly how many cases are represented by equation (3)? Since projects are exclusively in one unique program at a time, only one of the grant programs can be activated for a specific case, all others excluded. This means that the number of cases represented by equation (3) in number of possible combinations is 896 ($= 7*2^{14-7}$). However for the sake of simplicity some reduction is possible if we eliminate those programs that do not enter the interaction terms (all programs with the exception of Cpother). There are a total of 8 binary factors remaining, Item2, Item3, Item5, Item6, Item10, Item12, Doc, Cpother, and the number of combinations is 256 (2^8). Some of these cases, the best, the worst and the middle, are aligned in table 4 ordered by the probability of conformity *P(c)*. The last case has essentially no chance of being approved (0.001), while the chance of approval for the first case is practically certain (0.999). Another result that will merit more detail subsequently is that the absence of the doctorate (column Doc – n) is present in table 4 for both large probabilities and small. On the other hand, items 10 and 12 follow a consistent pattern in terms of best and worst probabilities.

In order to get a general idea as to the structure of the equation and the relationship between *P(c)* and the predictor factors, Box plots are constructed for these cases. In figures 2 and 3, we see that item 2, 5, 6 and 12 are the only factors where the probability of conforming increases with the successful approval of the item on the checklist. However, and counter-intuitively, items 3 and 10 when successfully approved diminish the probability of conformity. Note the deleterious effect of the doctorate, and the difference in variability. This result corroborates the result already seen in the preceding table that the absence (n) of the doctorate appears consistently in both highly conforming and nonconforming projects.

Because of the complex nonlinear associations in the estimated equation (3), a common characteristic in logistic equation estimation, in the next section differential effects are measured and the importance of each factor determined.

Item.2	Item.3	Item.5	Item.6	Item.10	Item.12	Doc	CPother	probs
s	n	s	s	n	s	n	FALSE	0.999
s	n	s	n	n	s	n	FALSE	0.999
s	n	s	s	n	s	n	TRUE	0.997
s	n	n	s	n	s	n	FALSE	0.997
s	s	s	s	n	s	n	FALSE	0.996
s	n	s	n	n	s	n	TRUE	0.995
n	n	s	s	n	s	n	FALSE	0.995
s	n	n	n	n	s	n	FALSE	0.994
s	s	s	n	n	s	n	FALSE	0.993
s	n	s	s	s	s	n	FALSE	0.992
s	n	n	s	n	n	n	TRUE	0.708
s	n	n	n	s	s	n	TRUE	0.707
n	s	s	s	s	s	s	FALSE	0.706
n	n	s	s	s	s	s	FALSE	0.701
n	s	s	s	s	n	s	TRUE	0.691
n	n	s	s	s	n	s	TRUE	0.686
n	n	n	s	s	s	n	FALSE	0.677
s	s	s	s	n	n	n	TRUE	0.671
s	s	s	n	s	s	n	TRUE	0.669
s	s	s	s	n	s	s	FALSE	0.667
s	n	s	s	n	s	s	FALSE	0.662
n	s	n	n	n	n	n	TRUE	0.020
n	n	n	s	s	n	n	TRUE	0.017
n	s	s	s	s	n	n	TRUE	0.014
s	s	n	n	s	n	n	TRUE	0.013
n	s	n	s	s	n	n	FALSE	0.012
n	n	n	n	s	n	n	TRUE	0.010
n	s	s	n	s	n	n	TRUE	0.008
n	s	n	n	s	n	n	FALSE	0.007
n	s	n	s	s	n	n	TRUE	0.002
n	s	n	n	s	n	n	TRUE	0.001

Table 4. Best, worst and middle for the probability of conformity for given values of the factors.

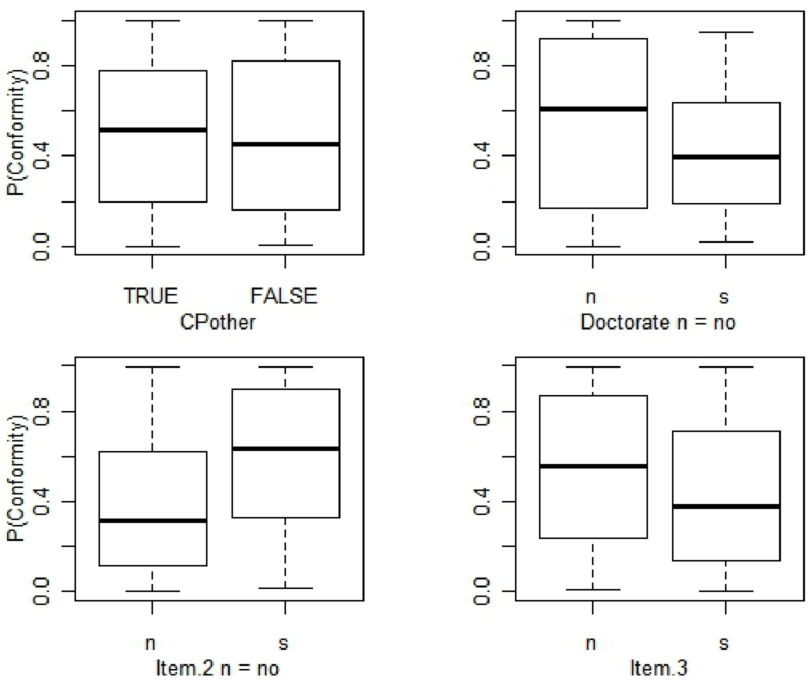

Figure 2. Box plots

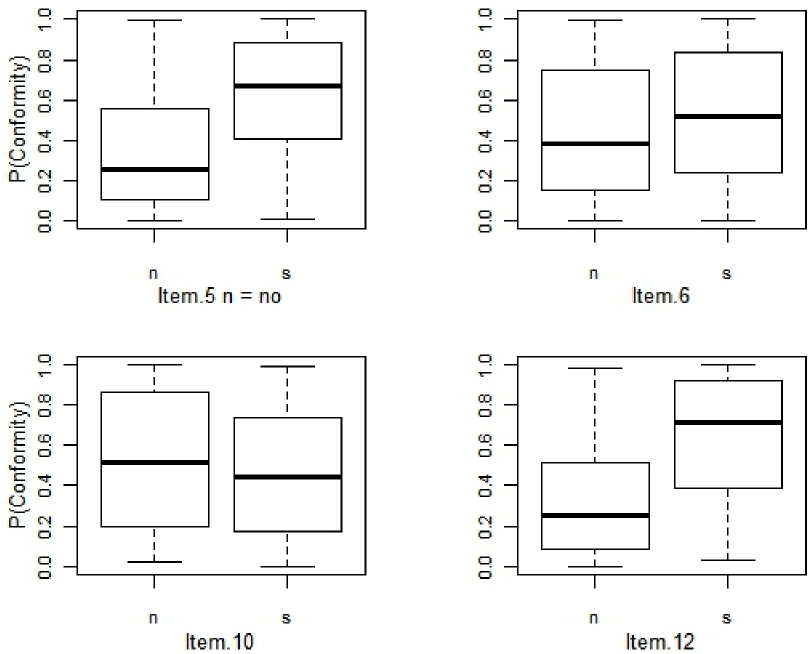

Figure 3. More Box plots

5. MEASURING DIFFERENTIAL EFFECTS OF THE PREDICTORS

Predictor factor influence can be represented by a finite difference equation. From equation (1)

$$(4) \qquad \Delta \text{logit} = \left[\ln \frac{P_1(c)}{1-P_1(c)} - \ln \frac{P_0(c)}{1-P_0(c)} \right] = b_1 \Delta X_1 + b_2 \Delta X_2 + b_{12} X_2 \Delta X_1 \; b_{12} X_1 \Delta X_2$$

where the value of $P_0(c)$ (= 0.686) is the prevalence of conforming projects in the sample from the original data. The expression ΔX_i as a binary factor can assume values of 0 or 1. Due to the nonlinear nature of the logistic regression, the impact of a predictor variable on the response – $\Delta P(c)$ – is not easily interpreted, however for those predictor variables that do not enter the interaction terms interpretations are more straightforward. Consider for instance Grant Program CP00106, which has a coefficient estimate of 1.822 indicating a positive influence on the logit (table 3). The regression coefficient can be transformed into a measure of impact on the probability of conformity of financial accounts $\Delta P(c)$ given Grant Program CP00106. From equation (2) and footnote 7,

$$(5) \qquad \Delta \text{logit} = \left[\ln \frac{P_1(c)}{1-P_1(c)} - \ln \frac{P_0(c)}{1-P_0(c)} \right] = 1.822 * \Delta \text{ CP00106}$$

$$\text{odds}_1/\text{odds}_0 = e^{1.822}; \qquad P_1(c) = 0.931$$

Substituting for $P_0(c) = 0.686$ (average of approvals in all grants) and Δ CP00106 = 1, the result is

$P_1(c) = 0.931$ and $P_1(c) - P_0(c) = 0.931 - 0.686 = 0.245$. Consequently, assuming that the other factors are fixed to yield the average of conformity (0.686, see the middle of table 4), the probability of conformity increases by about 25 percentage points for grant program CP00106. See the first line of table 5 at the end of this paper.

The calculation of $P_1(c)$ for a factor like Doc is more complicated because of the interaction terms. See equation 2 and imagine a partial derivative for $P(c)$ in terms of X_i. The impact of Doc on the probability of conformity besides being nonlinear is not simply a single value but rather depends on the presence of the interacting factors.

(6) $$\Delta \text{logit} = \left[\ln \frac{P_1(c)}{1-P_1(c)} - \ln \frac{P_0(c)}{1-P_0(c)} \right] =$$

$$(-3.387 + 2.7776\text{CPother} - 1.9785\text{Item12} + 3.6977\text{Item10} - 1.4535\text{Item2} + 1.9751\text{Item.3})\Delta\text{Doc}$$

$$= 1.631$$

$$\text{odds}_1/\text{odds}_0 = e^{1.631} \quad \rightarrow \quad P_1(c) = 0.918$$

Given the situation where all interaction terms are activated (Cpother = Item12 = Item10 = Item2 = Item3 = 1), $P_1(c) = 0.918$ and $\Delta P(c) = P_1(c) - P_0(c) = 0.232$. The presence of the doctorate in this case has a beneficial impact of about 23 percentage points on the conformity of the financial accounts. This result is on the fourth line from the bottom of table 5. On the other hand, if the situation is reflected by no interaction term activated, all equal to zero, then the presence of the doctorate causes a decline of 62 percentage points in the probability of conformity, as shown in table 5, three lines from the bottom. The last four numbers in column ΔP are especially noteworthy, giving the change in probability when the coordinator has a doctorate and all or some of the items of the checklist are verified. These last entries are especially disheartening considering that two configurations, when present together with the doctorate, diminishes the probability of conformity by more than 60 percentage points. Maximum negative impact is represented by a fall of about 68 percentage points, when the coordinator has a doctorate and items 12 and 2 are approved. This result is due to the estimation result that all relevant regression coefficients are negative summing to a strong negative impact. The conclusion here seems to be that doctorates make for poor coordinators. Another question to look into would be the large disparity in the impact of specific grant programs (see the first seven impact calculations in the last column of table 5). Some of the largest positive and negative impacts in the table are among these numbers. Comparing prerequisites among different grant programs should unveil some relevant characteristics.

6. STATISTICAL PROCESS CONTROL

Finally, another method for measuring the impact of the factors on *P(c)* is inspired from the Statistical Process Control literature and the construction of control charts (Shewhart, 1929). For instance, in the traditional control chart for defects, the c chart, and all control chart applications for that matter, the underlying process is assumed to be stable (stationary in the time series literature) with constant mean defects and variance per sampling unit. Since the process is assumed stable, periodic sample measurements within a control limit show that the process is momentarily acceptable, and for measurements outside the limit the process may have become unstable and an investigation is forthcoming. The placement of the limit in the control chart is a question of probability: neither should the limit be too far away from the process mean resulting in undisclosed process instability, however nor should it be too close to the mean producing an abundance of measurements outside the limit with no assignable cause (false alarms). We can offer operational concreteness to the theory of control limits by adjusting the control limit using the estimates from the logistic regression. Table 6 is organized to compare the probability of conformity between a base case where all items on the checklist are approved (first line of the table), with situations characterized by one or a few items rejected. The base diff column shows by how much *P(c)* declines when certain items are not approved. For instance, the difference between the first line and the second line where item3 is the only rejected item is a very small 0.001, demonstrating that item3 has little effect on the outcome. However, the seventh line in the table represents the rejection of item 5 and a decline in *P(c)* of 0.195, the strongest impact among the first lines of the table for individual item rejection. The first *P(c)* less than 0.5, which might be considered a benchmark value for judging nonconformity, comes from the joint rejection of item 5 and item 12. This result makes these two items a priority for the auditor: item 5 should be checked first and if rejected then the auditor should jump to item 12 and if also rejected then the financial accounts are judged nonconforming. This procedure would greatly reduce the workload of the auditor.

Item.2	Item.3	Item.5	Item.6	Item.10	Item.12	Doc	CPother	*P(c)*	base diff	items
s	s	s	s	s	s	s	TRUE	0.947	0	
s	n	s	s	s	s	s	TRUE	0.946	0.001	3
s	s	s	n	s	s	s	TRUE	0.912	0.035	6
n	s	s	s	s	s	s	TRUE	0.886	0.061	2
s	s	s	s	n	s	s	TRUE	0.867	0.08	10
s	s	s	s	s	n	s	TRUE	0.837	0.11	12
s	s	n	s	s	s	s	TRUE	0.752	0.195	5
s	n	n	s	s	s	s	TRUE	0.748	0.199	5 3
s	s	s	n	s	n	s	TRUE	0.748	0.2	12 6
n	s	s	s	s	n	s	TRUE	0.691	0.256	12 2
s	s	s	s	n	n	s	TRUE	0.651	0.296	12 10

s	s	n	n	s	s	s	TRUE	0.636	0.311	5 6
n	s	n	s	s	s	s	TRUE	0.569	0.378	5 2
s	s	n	s	n	s	s	TRUE	0.524	0.423	5 10
s	s	n	s	s	n	s	TRUE	0.465	0.482	5 12
s	n	n	s	s	n	s	TRUE	0.46	0.488	3 5 12
n	s	s	s	n	n	s	TRUE	0.447	0.5	2 10 12
n	s	n	n	s	s	s	TRUE	0.432	0.516	2 5 6
s	s	n	n	n	s	s	TRUE	0.388	0.56	5 6 10
s	s	n	n	s	n	s	TRUE	0.334	0.614	.. 5 12
n	s	n	s	n	s	s	TRUE	0.323	0.624	2 5 10
n	s	n	s	s	n	s	TRUE	0.274	0.673	.. 5 12
n	n	n	n	n	n	s	TRUE	0.071	0.876	.. 5 12

Table 6. Probability of conformity for some cases, control chart construction.

However, a question remains unanswered: What is the best **cutoff** probability *P(c)* that reduces the probability of error as judging *good* projects nonconforming and *bad* projects conforming. In the next section this question is analyzed in the light of the Receiver Operator Characteristic (ROC) curve, ubiquitous in the health sciences but rarely seen elsewhere.

7. THE RECEIVER OPERATOR CHARACTERISTIC (ROC) CURVE

Ziliak (2012) and Ord (2012) commenting on a special section of *The International Journal of Forecasting* with lead article by Soyer and Hogarth (2012) have recently called for more graphical analysis instead of the mechanical style of hypothesis testing common in Social Science research. They argue that graphical analysis may be easier to understand than a list of statistics and p values, and furthermore when well-elaborated a graphical analysis may even be more complete and more transparent. ROC analysis has been growing in popularity as one of many methods that respond to the challenges of graphical analysis. An interesting introduction can be found in Swets, *et al* (2000) and a pioneering application in crime forecasting can be found in Gorr, W. L. (no date) and Cohen, *et al* (2009).

In the health sciences a test which shows a positive result means that a medical condition or a specific substance has been indicated. A patient tests positive for cancer, an athlete positive for hormone abuse. One might say that a positive result is analogous to the rejection of the null hypothesis, the null being that the patient is healthy or the athlete is free of prohibited substances. The null hypothesis of this paper is that financial accounting is conforming. Within the population and consequently in the sample, sample size is the sum of positives (P) and negatives (N), and objects are classified as either positive or negative by the test. The positives and negatives in the sample will not coincide perfectly to the test results, some errors of classification will certainly appear. Therefore, the correct application of ROC analysis requires that the classification of individual elements be known, for instance which patients are really healthy or which athletes are free of drugs. The true classification is usually known after more tests are performed and conclusions are held with more certainty. The data in this paper were classified as financial accounts that are either conforming (c) or nonconforming (n). In ROC analysis there are at least two ways to measure the overall accuracy of the test. Tests can be graded on the basis of the true positive rate TPR = TP/P which measures the "sensitivity" (power) of the test, and answers the question of how many positives were correctly determined (TP) within the group of positives (P). Tests can also be graded on the true negative rate TNR = TN/N called a measure of "specificity".

Tests commonly result in numerical values that dictate whether the test object is positive or negative. Usually tests are set up so that large values of the numerical result indicate a positive and small measures indicate negative. Consequently, the choice of the cutoff value which separates positive from negative is an essential part for optimizing the accuracy of the test. If the cutoff value is too large then mistakes will occur that overemphasize negatives, false negatives (FN) will be in abundance, however, only a small number of false positives (FP) will be produced. On the other hand, if the cutoff value is too small, then positives will be produced in abundance and among them many false positives (FP), also known as false alarms and Type I error. However false negatives (FN), Type II error, will be diminished in number since there is an inverse relationship between FP and FN. Abdi (2007) refers to decision processes as liberal or conservative considering whether they allow for relatively large FP or relatively small, respectively. In many areas of Science, the cutoff value is fixed to produce a maximum of 5% for the false positive rate FPR = FP/N also known as the level of significance alpha (FPR = 1 – TNR), the area in the tail of the null distribution, representing a maximum cutoff value for not rejecting the null hypothesis. In other words, Economists tend to accept the liberal approach to decision making. In engineering and especially Statistical Process Control, the cutoff value is fixed to reflect a FPR of only 0.27% (approximately one quarter of one per cent), representing a conservative approach. The other side of the question is the false negative rate (FNR) also in some circumstances called beta (= 1 – TPR = 1 – "sensitivity"). McCloskey and Ziliak complain that Economists who ubiquitously use the 5% FPR, simply ignore the existence of the false negative rate (FNR) which at times for some applications may approach 100% but is not reported as relevant information in published works. ROC analysis allows for an appreciation of both Type I and Type II error, and through the use of a utility-disutility-cost function that weighs the relative importance of false negatives and false positives differently depending upon the situation under study and the preferences of the decision maker, a cutoff value can be chosen which minimizes a combination of the FNR and the FPR. In a later section, we will develop the function of weights and show that its structure also depends upon the prevalence of negatives and positives in the sample[9].

	REALITY WITHIN THE SAMPLE		
	TRULY POSITIVE (nonconforming)	TRULY NEGATIVE (conforming)	
TESTED AS POSITIVE	TP (hits) TP/P = sensitivity	FP (false alarm) FP/N = 1- specificity alpha	TOTAL TESTED POSITIVE
TESTED AS NEGATIVE	FN FN/P = 1- sensitivity beta	TN TN/N = specificity	TOTAL TESTED NEGATIVE
	TP+FN = P	FP+TN = N	N + P = TOTAL

Table 7. Tests and the truth: ROC concepts.

Diminishing the number of false positives and false negatives is the goal of quality audits, which simultaneously will maximize true positives and true negatives. In the context of quality audits, false positives (condemning good accounts) are considered relatively minor errors from the auditors point of view, because this kind of error can be fixed later at little expense to the auditor, whereas false negatives are much more prejudicial. Once an audit has OK´ed financial accounts that are in fact bad, there is no second chance to catch this mistake. Consequently, from the point of view of the auditor, the cost of the false negative is much greater than the cost of the false positive. However, even though the false positive is very costly to the researcher who will have to defend himself against the negative determination of the auditor, who determines the weights is the auditor not the researcher. A cutoff value is defined in an optimal sense to classify conforming and nonconforming financial accounts. The value of the cutoff will be chosen in order to minimize a combination of false positives and false negatives from the point of view of the auditor.

9 Based on risk analysis and utility functions (Metz,1978).

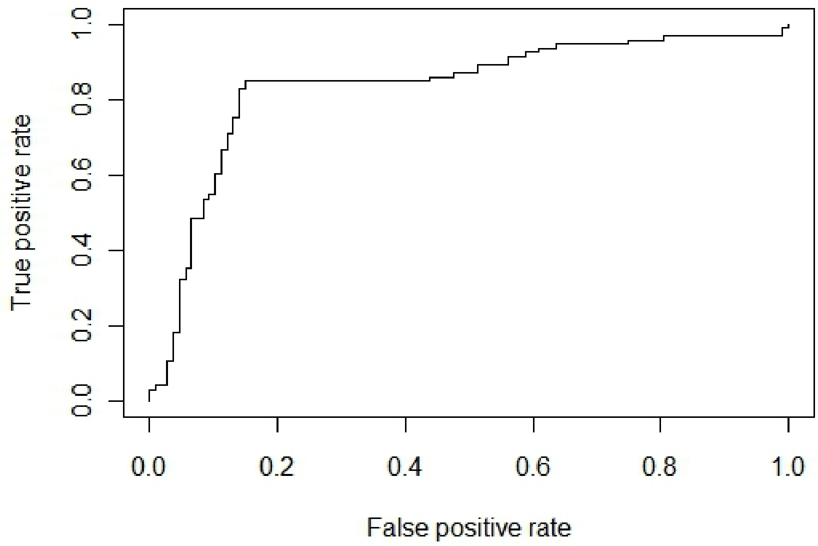

Figure 4. Example ROC curve from R Graph Gallery (2012), T. Sing, et al (2005)

The ROC curve is drawn as the relationship between FPR the false alarm rate and 1 – FNR the true hit rate. A stylized version is given in figure 4, with the horizontal axis representing the rate of false alarms (FPR) and the vertical axis representing the hit rate (1 – FNR = TPR). In the figure, starting at the origin, a decision maker would be interested in knowing that a small increase in the FPR would result in a corresponding but much greater TPR up to a value of FPR of approximately 0.2 and corresponding TPR of 0.8. The adjustment in TPR and FPR is made by selecting the appropriate cutoff value. These numbers are similar to the empirical results presented below and will be discussed in detail. In the next section, a method is presented for choosing an optimal pair of FPR and TPR by fixing the cutoff value based on a subjective evaluation of the perceived costs of FP (false alarms) and FN.

8. OPTIMIZING IN THE CHOICE OF FP AND FN.

By choosing the relative costs of FP and FN from the point of view of the relevant decision maker, a point considered as optimal can be chosen along the ROC curve. Cohen, et al (2009) use a utility function approach from decision theory that takes into account marginal utility and disutility from each result of table 7, but here we will apply a simpler cost equation, hopefully without loss of generality. Define a total cost function, linear for simplicity, to be minimized, based on FP and FN.

$$CT = FP + R * FN$$

where R is the perceived cost of FN, and the cost of FP is unitary. The application in this paper uses

R > 1 since the auditor will weigh FN with much more onus than FP, as suggested in the previous section.[10] Rearranging terms to explicitly show the proportion (*p*) of positives (*P*) in the total and redefining cost as unitary (*C*),

(7) $$C = FPR*(1-p) + R*FNR*p$$

Furthermore, rearranging in terms of TPR,

(8) $$TPR = \frac{(Rp-C)}{Rp} + \frac{FPR*(1-p)}{Rp}$$

This equation, often called the cost constraint, if placed in figure 4 would be a straight line with intercept at $\frac{(Rp-C)}{Rp}$ and slope equal to $\frac{(1-p)}{Rp}$. Along a given cost constraint costs are constant. Costs are minimized and consequently TPR and FPR are at optimal values when the ROC curve is just tangent to equation (8). At the optimal point, the slope of TPR = $\frac{(1-p)}{Rp}$ is equal to the slope of the ROC curve. This means that as *p* approaches 1.0 and as *R* the perceived cost of FN increases, optimal values of TPR and FPR increase (FNR decreases).

In the classification problem studied here, the major determinant for the values of FPR and FNR is the cutoff value, when too large false negatives will appear in excess and, to the contrary, when too small false positives will be exaggerated. In order to choose the correct value of the cutoff, the elements of the cost function will be assigned values either from sample estimates or from the preferences of the decision making auditors. In the next section the empirical ROC curve will be constructed.

9. ESTIMATES OF THE ROC CURVE.

Throughout this section, statistical results and analysis depend upon the R language (R Development Core Team (2012)) and the ROCR package by Sing, *et al* (2005). The ROC curve quantifies the tradeoff between false alarms FPR and hits TPR. As explained above the ROC curve shows how many false alarms must be tolerated in order to reach a certain level of test reliability to recognize the presence of the positive state. In the context of this article, the object to be tested or classified is the financial report associated with a research grant. The test classifies the report as conforming or nonconforming. The fundamentals of the test is the estimated logistic regression. The result of the test is the probability of conformity *P(c)*. In order to be aligned with the statistical literature on ROC curves, the result of the test will be defined as *P(n)* the probability of non-conformance, in other words the probability of the positive state traditionally identified with a

[10] For an interesting example in the context of natural disasters of an attempt to quantify costs of FP (announce an evacuation that is not necessary, false alarm) and FN (no evacuation is announced but disaster strikes) see Regnier (2008).

certain sickness, the presence of a certain substance, or a problem on the assembly line, or as presented here the non-conforming financial report. It would seem intuitively justifiable to use *P(n)* = 0.50 as the cutoff between conforming and nonconforming financial reports as was the case in table 6. This case is illustrated below in figure 5.

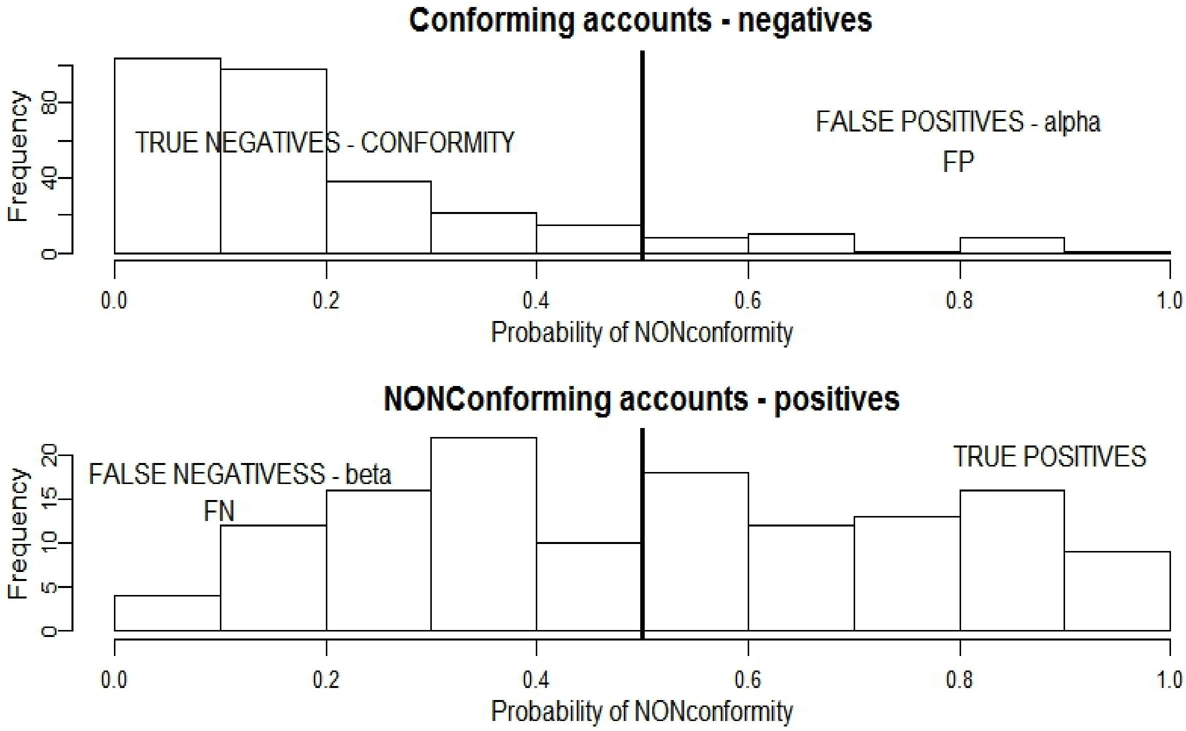

Figure 5. Separate histograms for the probability of nonconformity of positives and negatives, cutoff at 0.5.

Two separate histograms have been constructed to illustrate the probability of non-conformance for conforming accounts in the upper part of the figure and for nonconforming in the lower. The histogram on conforming accounts shows a reduced number of false positives, characterized by large nonconforming probability indicating accounts rejection even though the accounts had been originally approved as conforming by the staff, the false positive (FP). This small number of FP´s seems counterproductive since FP´s cause little cost to the staff as explained above. One might consider diminishing the cutoff value to less than 0.5 allowing an increase in relatively cheap FP´s and consequently diminishing the corresponding FN´s in the lower histogram of nonconforming accounts. Diminishing the number of FN´s seems the correct approach given their prejudicial status from the point of view of the staff.

Another way of seeing the same result is by constructing box plots for the same set of data used in the histograms.

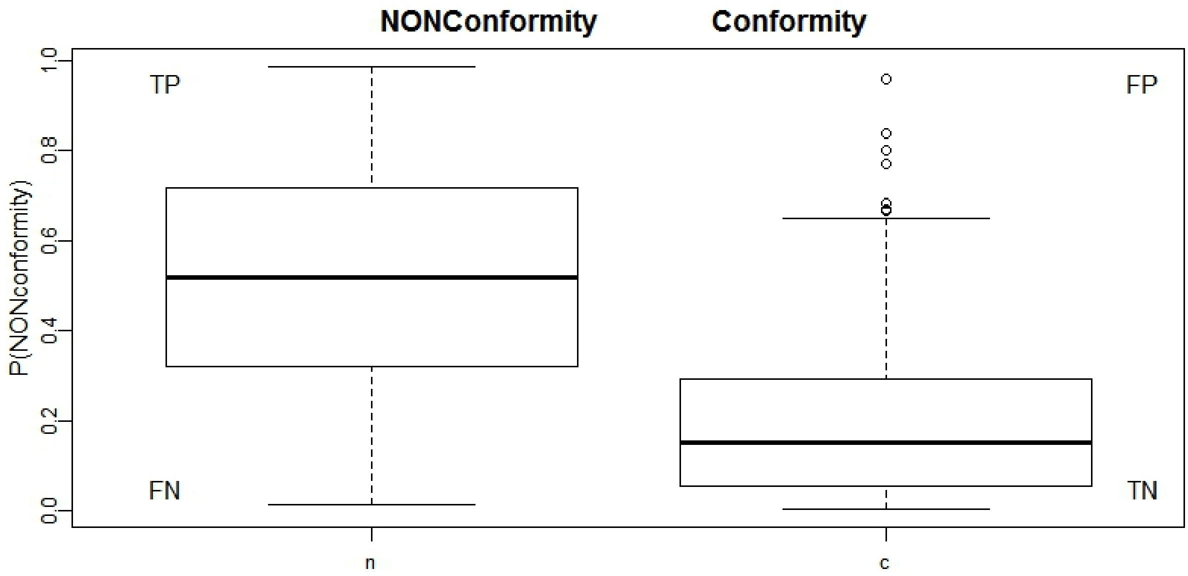

Figure 6. Box plots.

As was apparent from the histograms, the frequencies of conforming accounts is very skewed whereas the frequencies of nonconforming is symmetric. Once again, diminishing the cutoff based on the probability of non-conformance will increase FP at relatively low cost but decrease FN whose cost is burdensome.

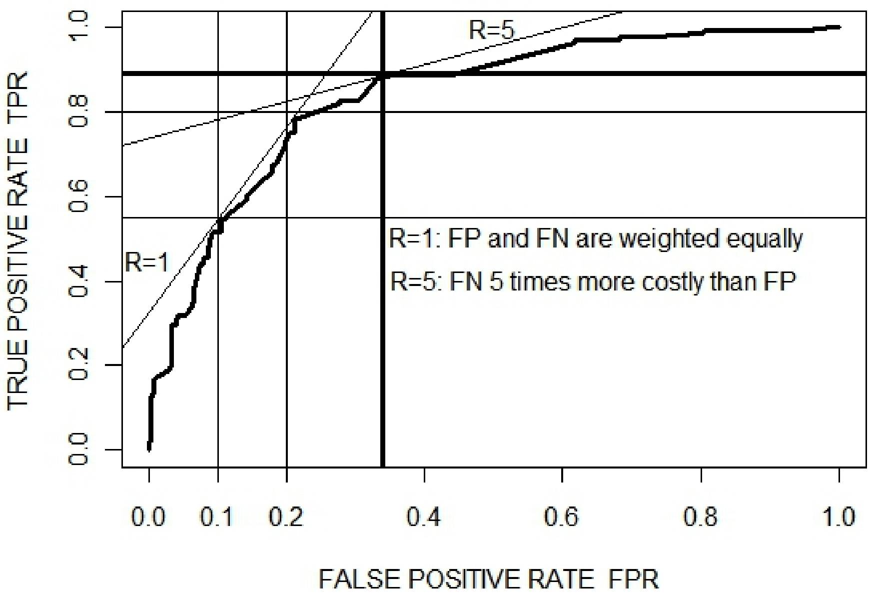

Figure 7. ROC curve.

The ROC curve presented in figure 7 offers a better view of the trade off between FP and FN. Curvature is relatively pronounced which implies that the relationship between the estimates of the logistic regression and the categorization of the data is not random but correlated. When tests are perfect, the ROC curve is pushed into the northwest corner (FPR = FNR = 0), when tests are worthless, no better than a random draw, the ROC curve becomes a straight line with origin at zero and slope equal to one.[11] For low values of FPR and TPR, allowing for a small increment in FPR would result in a relatively large increase in TPR, a situation that the decision maker may accept enthusiastically. However, as FPR increases in value, the return to TPR declines accordingly. Hence, the optimal combination of FPR and TPR is a point on the ROC curve that satisfies minimum cost as perceived by the decision maker. In figure 7, two cost constraints have been superimposed (equation 8). The steeper constraint labeled R = 1 represents weights that are equal between FP and FN, no preference is manifested favoring either one of the false results. Since costs are constant along the constraint, the optimal combination is not uniquely defined in this case, FPR is approximately 0.1 or 0.2. The flatter curve represents R = 5 which is the realistic perception of the staff, FN is 5 times more important (more costly) than FP (equation 8). Logically, this case should produce an FPR larger and an FNR smaller. The exact values as shown in figure 8 are in fact 0.34 and 0.89 for the FPR and the TPR, respectively. FNR (= 1 – TPR) is reduced to 0.11.

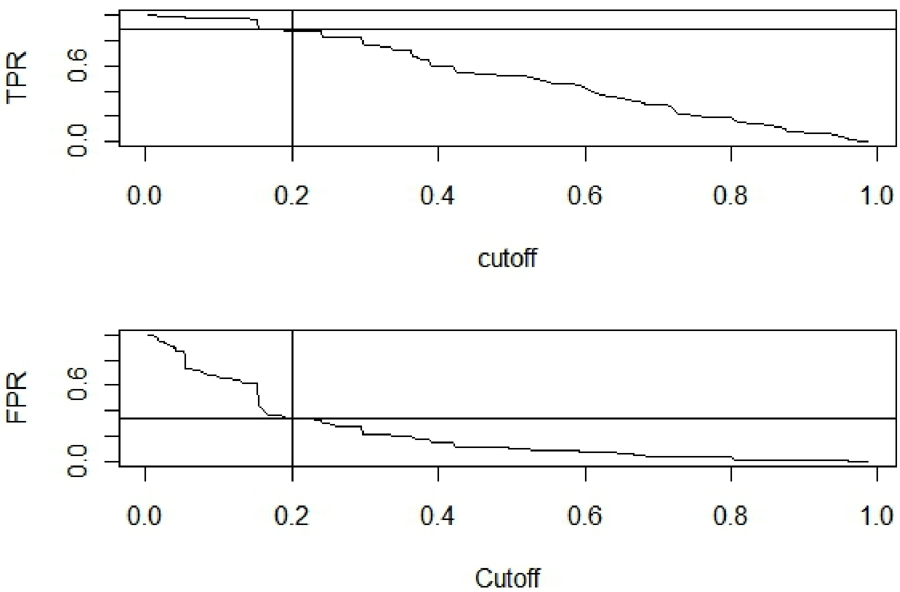

Figure 8. False positive rate and true positive rate as functions of the cutoff.

As explained above at the end of section 7, the optimal combination of FPR and TPR correspond to the same cutoff, and from figure 8, where horizontal lines have been drawn at the values of the optimal combination, visual inspection reveals that the cutoff should be 0.20. Consequently, when *P(n)* calculated from the logistic regression is less than 0.20, *P(c)* greater than 0.80, then financial accounts should be judged as conforming, otherwise they are non-conforming. The value of the cutoff may seem very small, however this is the value that best reflects the preferences of the decision makers. The optimal cutoff is quite different from P(.) = 0.50, suggested earlier as the

11 There is much relevant literature on the statistical analysis of ROC curves not elaborated in this article but an excellent reference is Kumar, R. and A. Indrayan (2011).

natural choice. Returning to table 6, readjusting the cutoff to 0.20 would mean that if any one of items 2, 10, 12, or 5 were rejected then a judgement of non-conformance would be forthcoming.

9. CONCLUSIONS

Our analysis has disclosed several avenues of actions to make the process of verifying financial accounts both faster and more accurate. First of all, the discrepancy between true positives and negatives, and the approval or not of specific items in the checklist may have its origin in operational definitions, analogous to the calibration of measurement instruments on the assembly line. With a renovated and updated checklist, eliminating some items and including new ones, procedures can be streamlined and made more reliable. Prime targets for this updating are those items which demonstrate no relationship with the classification of the financial accounts or possess a counter-intuitive signal.

The operational order of the checklist items should reflect the impact they have on the probability of non-conformance. As shown in section 6, items have different impact weights and large weights should correspond to priority items at the top of the list. This procedure would eliminate the necessity of reviewing all checklist items.

The project coordinator with a doctorate seems to present problems for conformance. There is a definite tendency in the results which connect the doctorate to non-conformance. Why this occurs may be for a psychologist to decide, but the fact is verified.

10. REFERENCES

Abdi, H. (2007). Signal Detection Theory, *Encyclopedia of Measurement and Statistics*, 8, pp. 313–324.

Agresti A (2002). Categorical Data Analysis. JohnWiley & Sons, Hoboken, New Jersey, 2nd edition.

Cecchini, M., H. Aytug, G. J. Koehler, and P. Pathak, (2010). Detecting Management Fraud in Public Companies, *Management Science*, vol. 56, no. 7, pp. 1146-1160, May

Cohen, J., S. Garman, and W. Gorr (2009), Empirical calibration of time series monitoring methods using receiver operating characteristic curves, *International Journal of Forecasting*, vol. 25, no. 3, pp. 484–497.

Cramer,J. S. (2003) The origins and development of the logit model, *Bliss*, August. pp. 1–19

Deming, W. Edwards (1990). *Qualidade: A revolução da Administração*. Rio de Janeiro, Editora Marques-Saraiva.

Dionne, G., F. Giuliano, and P. Picard (2008), Optimal Auditing with Scoring: Theory and Application to Insurance Fraud, *Management Science*, vol. 55, no. 1, pp. 58-70.

Dodge, H. F. (1928) A method of rating a manufactured product. *Bell System Technical*

Journal, 7, 350–368.

Fugee, Tsung, Y. Li, and M. Jin, (2008) Statistical process control for multistage manufacturing and service operations : a review and some extensions, *Int. J. Services Operations and Informatics*.

Gelman, Andrew, James S. Liebman, Valerie West, and Alexander Kiss (2004) A Broken System: The Persistent Patterns of Reversals of Death Sentences in the United States, *Journal of Empirical Legal Studies,* Volume 1, Issue 2, 209–261, July

George, M. L (2003). *Lean Six Sigma for Service.* New York: McGraw-Hill

Gorr, Wilpen L. (no date) Forecasting Exceptional Demand Based on Receiver Operating Characteristics ROC. working paper, pp. 1–15.

Hastie, T., R. Tibshirani, AND J. Friedman, (2008) *The Elements of Statistical Learning.* Stanford, California

Hawkins, D.M. E Olwell, D.H. (1998). *Cumulative sum charts and charting for quality improvement.* Sringer, New York.

Komori, O. (2009) A boosting method for maximization of the area under the ROC curve, *Annals of the Institute of Statistical Mathematics*, vol. 63, no. 5, pp. 961-979, Oct.

Kumar, R. and A. Indrayan, (2011) Receiver operating characteristic (ROC) curve for medical researchers., *Indian Pediatrics*, vol. 48, no. 4, pp. 277–287.

Nembhard, D. A. and Harriet Black, A demerits control chart for autocorrelated measurements, Quality Engineering, 13(2), 179-190 (2000-01)

Ord, J. Keith The illusion of predictability: (2012). A call to action. *International Journal of Forecasting,* Volume 28, Issue 3, July–September, Pages 717-718

Paladini, E. P. (2000). *Gestão da Qualidade.* São Paulo: Átlas.

R Development Core Team (2012) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http://www.r-project.org/.

R Graph Gallery (2012), http://addictedtor.free.fr/graphiques/RGraphGallery.php?graph=99, visited 8-28-12.

Reed, L. J. and J. Berkson (1929). The application of the logistic function to experimental data. Journal of Physical Chemistry 33, 760-779.

Regnier, E. Public Evacuation Decisions and Hurricane Track Uncertainty, *Management Science*, vol. 54, no. 1, pp. 16–28, Jan. 2008.

Samohyl, R.W. (2009). *Controle Estatístico de Qualidade.* Rio de Janeiro: Elsevier.

Shewhart, W. (1931). *Economic control of quality of manufactured product.* New York: D. Van Nostrand Company.

Sing, Tobias, Oliver Sander, Niko Beerenwinkel, Thomas Lengauer (2005). ROCR: visualizing classifier performance in R. *Bioinformatics* 21(20): 3940-3941

Sing, Tobias, Oliver Sander, Niko Beerenwinkel and Thomas Lengauer. (2009). Package ROCR: Visualizing the performance of scoring classifiers. Version dated 08-Dec-2009. Available at: http://cran.r project.org/web/packages/ROCR/.

Soyer, Emre, Robin M. Hogarth (2012). The illusion of predictability: How regression statistics mislead experts *International Journal of Forecasting,* Volume 28, Issue 3, July–September,*Pages 695-711*

Ziliak, Stephen T. Visualizing uncertainty: On Soyer's and Hogarth's The illusion of predictability: How regression statistics mislead experts, *International Journal of Forecasting,* Volume 28, Issue 3, July–September Pages 712–714

Doc	CP00106	CP00109	CP00110	CP00205	CP00407	CP01209	CPother	Item.12	Item.10	Item.2	Item.3	Item.5	Item.6	Doc:CPother	Doc:Item.12	Doc:Item.10	Doc:Item.2	Doc:Item.3		ΔP=P1-Po	Δlogit
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CP00106	0.245	1.822
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CP00109	0.189	1.163
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CP00110	0.279	2.534
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CP00205	0.219	1.469
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	CP00407	0.241	1.763
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	CP01209	-0.444	-1.921
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	CPother	-0.379	-1.597
0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	CPother	0.191	1.181
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	Item.5	0.242	1.778
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	Item.6	0.105	0.553
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	Item.12	0.198	1.250
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	Item.12	0.296	3.229
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	Item.10	0.172	1.016
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	Item.10	-0.556	-2.682
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	Item.2	0.148	0.834
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Item.2	0.270	2.288
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	Item.3	0.005	0.022
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	Item.3	-0.449	-1.953
1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	Doc	0.232	1.631
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Doc	-0.617	-3.387
1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	Doc	0.294	3.089
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	Doc	-0.684	-6.820

Table 5. Impact on the probability of conformity (column ΔP) given certain characteristics of the project.