COMPARING TRADE-OFF ADJUSTMENTS IN CREDIT RISK ANALYSIS OF MORTGAGE LOANS USING AHP, DELPHI AND MACBETH

Fernando A. F. FERREIRA [a,b,*], **Sérgio P. SANTOS** [c,d]

[a] *ISCTE Business School, BRU-IUL, University Institute of Lisbon, Avenida das Forças Armadas, 1649-026 Lisbon, Portugal*
[b] *Fogelman College of Business and Economics, University of Memphis, Memphis, TN 38152-3120, USA*
[c] *Faculty of Economics, University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal*
[d] *CEFAGE, University of Évora, Portugal*

ABSTRACT. Due to the severe restrictions on access to credit resulting from the current economic climate, credit risk analysis of mortgage loans has been considered paramount for banking institutions and is currently accompanied by higher credit underwriting standards. In this paper, we present an empirical comparison of three decision support tools (i.e. Analytic Hierarchy Process (AHP), Delphi, and Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH)) in the specific context of trade-off readjustments in credit risk analysis of mortgage loans. We conducted a panel study with credit analysts and focused on five lines of comparison: ease of use; time-consumption; ease of applicability; accuracy; and overall evaluation. Results indicate that Delphi surpasses AHP and MACBETH in terms of ease of use, time-consumption and ease of applicability. As for accuracy, the differences obtained between AHP and MACBETH are not significant, and both methods perform better than Delphi. Most of the decision makers considered AHP the "overall best" approach.

KEYWORDS: Comparison of decision support methods; Mortgage lending; Risk evaluation; Trade-offs

1. INTRODUCTION

The relationship between financial and real estate markets and mortgage lending has been particularly highlighted after the most recent United States' subprime (cf. Kowalski, Shachmurove 2011; Puri et al. 2011; Yeager 2011; Beltratti, Stulz 2012). In particular, researchers have emphasized how important mortgage lending decisions are to revert the dramatic changes and declining values of real assets in many markets. As a consequence of poor lending decisions in the past, banks have been requiring higher credit underwriting standards, which impose heavier restrictions on access to credit, namely in terms of mortgage lending (Barth, Hollans 2012).

As expressed by Lopez and Saidenberg (2000: 152), banks have improved their credit-scoring risk systems over the years with the aim of *"better quantifying the financial risks they face and assigning the necessary economic capital"* (for further details

see e.g. Pau, Tambo 1990; Altman, Saunders 1997; Jacobson, Roszbach 2003; Yurdakul, İç 2004; Chambers et al. 2009; Xu, Zhang 2009; Yu et al. 2009; Twala 2010; Wang et al. 2011; Blackburn, Vermilyea 2012; Leow, Mues 2012). Nonetheless, many defend that the progress achieved does not mean that the current approaches to measure credit risk are without limitations (cf. Lopez, Saidenberg 2000; Doumpos, Zopounidis 2001, 2011; Doumpos et al. 2002; Twala 2010). Specifically, in line with Ferreira et al. (2011a), there is a lack of transparency in the way trade-offs among evaluation criteria are defined. Starting from this observation, and considering that elicitation methodologies have over the years successfully handled trade-offs among criteria (cf. Saaty 1980; Belton, Stewart 2002; Bana e Costa et al. 2005; Saaty 2008; Ferreira 2013), there is considerable scope to explore the applicability of these tools also in terms of credit risk evaluation of mortgage loans – a topic that, following Mari and Renò (2005: 92), has been *"quite neglected in the financial literature"*.

* Corresponding author. E-mail: fernando.alberto.ferreira@iscte.pt; fernando.ferreira@memphis.edu

Methodological comparisons among multiple criteria approaches have been widely reported in the literature (see, among others, Olson *et al.* 1995; Boucher *et al.* 1997; Zanakis *et al.* 1998; Scholl *et al.* 2005; Perini *et al.* 2009; Zhou, Ang 2009). However, the use of these approaches in the evaluation of credit risk has been very scarce and, as far as we are aware, no study compares their effectiveness in determining the relative importance of the different criteria used in risk analysis of mortgage loans. By comparing three decision support methods to define the relative importance of a pre-established set of mortgage loan risk evaluation criteria in use by one of the largest banks operating in Portugal, this study aims to contribute to this under-researched area. The methods compared are the Analytic Hierarchy Process (AHP), the Delphi method, and the Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH). A panel study with credit analysts was conducted and the performance of each method assessed against five major lines of comparison: ease of use; time-consumption; ease of applicability; accuracy; and overall evaluation.

The remainder of the paper is organized as follows. The next section underlines the economic relevance of mortgage lending. Section 3 briefly describes the institutional and managerial context of the study, while section 4 briefly introduces the three methods used. Section 5 explains their application in terms of trade-offs readjustments and presents the results of their comparison. Section 6 concludes the paper.

2. MORTGAGE LENDING ECONOMIC RELEVANCE

One of the top reasons supporting the economic relevance of mortgage lending is the fact that whilst home buying is usually seen as the major investment for most households, these do not frequently have the capital available to outright purchase a house. As such, bank mortgage loans are usually seen as the easiest solution and, from this perspective, it seems obvious that mortgage lending has stimulated the economy over the years.

Lima *et al.* (1995) and Mari and Renò (2005) present several interlinked reasons supporting the relevance of mortgage lending in economic terms. According to the authors, mortgage lending not only allows the expansion of financial services but also promotes the development of the housing construction industry and associated business activities. As a result of the financial (and thus economic) ex-

pansion, job creation and household consumption are stimulated, increasing money circulation and contributing to the country's Gross Domestic Product (GDP). As emphasized by Mari and Renò (2005: 83), *"the market for mortgage loans is of primary importance in any developed country"*. It is worth noting, however, that due to their inherent decreasing liquidity and limited access to debt markets, banks have imposed heavier underwriting requirements on access to mortgage credit (e.g. reductions of the *loan-to-value (LTV)* and *rate of effort*), and this has resulted in lower credit risk analysis outcomes, higher spreads and additional costs (often unaffordable to the borrower).

3. THE CURRENT CREDIT-SCORING SYSTEM AND ITS LIMITATIONS

The risk evaluation of mortgage loans in Portugal tends to be standardized and based on a ten-point scale. If a credit application scores between "1" and "5", the lending decision is usually *favorable*; scores above "5" support credit refusal. The final score of each credit-risk evaluation is a composite result, also known as overall score. This overall score is the result of aggregated partial scores, which, in turn, are associated to pre-established weighted criteria.

Although the types of mortgage loans available in the Portuguese banks usually depend on the market conjuncture, the criteria on which mortgage loan risk assessment relies are usually the same. Table 1 identifies the criteria commonly adopted by the five most representative banks operating in Portugal and reveals the weights used by one of them (whose name has been kept confidential upon request).

Table 1. Criteria and respective weights

Criteria	Weights %
Profession	12.00
Employment situation	8.00
Marital status	3.75
Age	2.50
Household	3.75
Cross-selling	10.00
Deposit portfolio	2.50
Average balances	2.50
Rate of effort	15.00
Responsibilities in BP	15.00
LTV	15.00
Existence of guarantors	5.00
Economic situation of the country	2.50
Political situation of the country	2.50

Source: Administrative Information.

As shown in Table 1, *LTV*, *rate of effort* and *responsibilities in BP* – Banco de Portugal – (Portuguese Central Bank after translation) are the three most weighted criteria in the current system for mortgage loan risk evaluation. Although easily understandable, the current mortgage credit-scoring system is not without its drawbacks. Specifically, the following limitations have been identified: a) the system is unable to consider behavioral criteria; b) considering that the trade-offs among criteria are previously and administratively established, there is an apparent lack of transparency and rationality, which prevents a proper understanding of the final decisions; and c) considering the existence of geographic idiosyncrasies, and that the trade-offs are the same for all bank branches, there is a possibility of inappropriate evaluations of mortgage loan applications.

In the literature, there are several decision support methods which have proved very valuable in addressing these problems in other contexts, and which might therefore also be very valuable to mortgage loan risk assessments. In particular, it is our opinion that the AHP, Delphi and MACBETH methods can provide a better basis for discussion and justification of mortgage lending decisions, leading to a more transparent evaluation mechanism. In order to explore how each of these methods can help banking institutions address the limitations previously identified, we carried out a comparative analysis involving bank directors and credit experts from the five most representative banks operating in Portugal.

It is important to emphasize, however, that the issue of which method to select or which methods to compare in order to identify the most appropriate ones is a challenging one. Considering that each decision support method has strengths and weaknesses, the choice of methods is usually determined by the decision analyst and is dependent on the decision context. In the particular case of this research, four major factors impacted on our decision on which methods to compare. Firstly, the AHP, Delphi and MACBETH methods have been recognized in the literature for being simple and facilitating trade-off calculations. Therefore, it was felt that any of these methods had potential to assist mortgage loan risk assessments. Secondly, the authors of the manuscript had previous experience in the use of these three methods to improve understanding and to assist decision making across several organizational contexts. Familiarity with the methods was considered an important factor to ensure a proper implementation and comparison.

Thirdly, the experts in risk analysis of mortgage loans who participated in the study had very stringent time constraints, preventing the inclusion of other methods in the comparison as further comparisons would have increased considerably the duration of the experiment. Finally, to the best of our knowledge, no previous research had compared the performance of the AHP, Delphi and MACBETH in determining the relative importance of the credit rating criteria used in risk scoring systems.

4. BRIEF METHODOLOGICAL BACKGROUND

In this section, we present a brief overview of the three decision support methods under analysis. The methods will be presented according to the order they were applied during our comparative experiment: Delphi, AHP and MACBETH.

4.1. The Delphi technique

The Delphi technique was developed in the 1950s by Norman Dalkey, Olaf Helmer and respective collaborators at the *RAND Corporation* (cf. Dalkey, Helmer 1963), and its procedure begins with an individual survey, which should be completed by individuals considered experts on the topic under discussion.

As outlined by Ferreira and Monteiro Barata (2011: 246), the operational structure of the method is based on *"a well-established sequence of successive individual questions supplemented with information and advice, which permits correcting the first stages of the process. […] it is a tool, which, under certain parameters, enables consensus. […] and is based on the rational principle that 'n' human minds are better than one when confronting the lack of precise knowledge about a certain subject"* (for further discussion, see also Linstone, Turoff 1975). The basic principles of the method are: *anonymity*, *controlled feedback* and *statistical treatment of the responses*.

4.2. Basics of the AHP approach

The AHP was developed by Thomas Saaty in the mid-1970s, with the purpose of overcoming the cognitive limitations of decision makers (cf. Saaty 1980). The implementation of the AHP begins with the identification of the criteria to be used in the evaluation of alternatives (also known in the literature as actions). These criteria are then organized in a hierarchical structure, which includes three key elements: criteria, subcriteria and alterna-

tives. Once this hierarchy has been defined, the relative importance of each of its elements has to be determined through a process of pairwise comparisons. The answers provided by the decision makers to the pairwise comparisons at each level of the hierarchy are then synthesized into square matrices. In each matrix, the number in row i and column j provides the relative importance (or priority) of a certain criterion C_i over another criterion C_j, as can be observed in the matrix form (1) presented below:

$$A = \begin{bmatrix} a_{ij} \end{bmatrix} = \begin{bmatrix} 1 & a_{12} & a_{13} & ... & a_{1j} \\ 1/a_{12} & 1 & a_{23} & ... & a_{2j} \\ 1/a_{13} & 1/a_{23} & 1 & ... & a_{3j} \\ ... & ... & ... & 1 & ... \\ 1/a_{1j} & 1/a_{2j} & 1/a_{3j} & ... & 1 \end{bmatrix}.$$

(1)

The conversion of pairwise comparisons into numerical values is based on a one-dimensional scale from 1 to 9, also known as 'Saaty's fundamental scale'. The next step of the process consists in calculating the weights w for each criterion in the different hierarchical levels and in relation to the alternatives considered. This is done by applying a mathematical technique known as eigenvector to the matrices of comparison. The mathematical expression (2) allows to estimate the eigenvector matrix.

$$W_i = \left(\prod_{i=1}^{n} a_{ij} \right)^{1/n}.$$

(2)

Laininen and Hämäläinen (2003) defend that the results obtained from the application of this mathematical expression must be standardized. Specifically, as shown in expression (3), where T is the normalized eigenvector, the standardization procedure consists in obtaining the proportion of each element relative to the sum.

$$T = \left| W_1 / \sum W_i ... W_n / \sum W_i \right|.$$

(3)

This procedure allows for the eigenvector ordering of priorities, and should be repeated until the normalized result of the last operation gets very close to the result of the preceding operation (*e.g.* irrelevant differences after the third decimal place). The eigenvector provides a hierarchy (also known as priority order) for the criteria involved and, to assess whether the data are logically related, the solution should be tested in terms of consistency by calculating the eigenvalue. Saaty (2008) proposes the following sequence of procedural steps to test the consistency of the solution:

– Calculate the eigenvalue (λ_{max}) in accordance with the expression (4), where w is obtained by summing the columns of the matrix of comparisons:

$$\lambda_{max} = T.w.$$

(4)

– Compute the consistency index (CI) using equation (5), where n represents the order of the matrix:

$$CI = \frac{\lambda_{max} - n}{n - 1}.$$

(5)

– Estimate the consistency ratio (CR) through equation (6), where RI is a random consistency index and depends on the order of the respective matrix:

$$CR = CI / RI.$$

(6)

The literature considers CR acceptable if its value is below 0.10. Contrarily, CR values above 0.10 require a revision of the matrix of comparisons (cf. Saaty 1980; Yurdakul, İç 2004; Xu, Zhang 2009; Perez-Gladish, M'Zali 2010). Following the procedure used to obtain the relative importance of the criteria, the levels of preference of the alternatives are determined by comparing pairs of alternatives in each criterion. Considering the levels of relative preference, and based on the additive model presented by equation (7), an overall assessment for each alternative considered can be made explicit:

$$V(a) = \sum_{j=1}^{n} p_j v_j(a) \ with \ \sum_{j=1}^{n} p_j = 1 \ and$$

$$0 < p_j < 1 \ (j = 1, ..., n).$$

(7)

Specifically, *V(a)* corresponds to the overall value of alternative *a*; *pj* is the weight of criterion *j* and *vj* is the level of preference of the respective alternative in criterion *j*.

4.3. The MACBETH approach

The MACBETH approach was developed during the 1990s by Carlos Bana e Costa and Jean Claude Vansnick (cf. Bana e Costa, Vansnick 1994). As in the AHP, the MACBETH process also involves the drafting of judgments between pairs of actions. The novelty lies, however, in the use of a semantic scale composed of categories of difference of attractiveness. Technically, let $X = \{a, b,..., n\}$ denote a finite set of n alternatives of choice. The technical procedure consists in associating each element of X to a value x (based on a value function $v(.)$: $X \rightarrow R$) such that differences $v(a) - v(b)$ (with a considered strictly more attractive than b (i.e. $a \ P \ b$))

are as compatible as possible with the preferences directly expressed by the decision makers. In this sense, for all pairs of actions (a, b) allocated to the same category C of semantic differences of attractiveness, the differences $v(a) - v(b)$ will belong to the same interval. In addition, the association of asymmetric partitions of the ray of positive real numbers to partition classes of ordered pairs (a, b) (with $a\,P\,b$) occurs whereas two contiguous ranges correspond to two consecutive categories of differences of attractiveness (Bana e Costa *et al.* 2008). This is done based on a value function v and function thresholds s_k as presented in formulation (8), where $P^{(k)}$ stands for a value preference that is stronger the greater the k:

$$a\,P^{(k)}\,b: \quad s_k < v(a) - v(b) < s_{k+1}. \tag{8}$$

The thresholds s_k are positive real constants and one of the objectives of the MACBETH approach consists in allocating the difference of attractiveness between each pair of actions $(a, b) \in X$ to one of the following categories of semantic differences of attractiveness: $C_0 = $ *Null* (or indifference (i.e. $a\,I\,b$)); $C_1 = $ *Very weak*; $C_2 = $ *Weak*; $C_3 = $ *Moderate*; $C_4 = $ *Strong*; $C_5 = $ *Very strong*; and $C_6 = $ *Extreme* (Bana e Costa *et al.* 2005). Additionally, formulations (9) and (10) should be analyzed for consistency purposes (Junior 2008; Ferreira *et al.* 2012):

$$\forall a,b \in X : v(a) > v(b) \Leftrightarrow aPb \tag{9}$$

$$\forall k,k^* \in \{1,2,3,4,5,6\}, \forall a,b,c,d \in X \text{ with } (a,b) \in C_k$$
$$\text{and } (c,d) \in C_{k^*} : k \geq k^* + 1 \Rightarrow v(a) - v(b) \geq v(c) - v(d) \tag{10}$$

Formulation (9) represents the logical assumption that if action a is strictly more attractive than action b (i.e. $a\,P\,b$), then the value of a must be greater than the value of b (i.e. $v(a) > v(b)$), allowing the association of numbers to both actions. Naturally, if action a is considered as attractive as action b (*i.e.* $a\,I\,b$), meaning that no difference between them is detected, then $v(a) = v(b)$ and the pair $(a, b) \in C_0$. In accordance with Bana e Costa *et al.* (2008: 28), formulation (10) stands for the guiding principle *"that all of the differences allocated to one semantic preference difference category are strictly larger than those allocated to a lower category"*. With consistent value preferences, linear programming is applied using formulation (11) (Junior 2008), which minimizes $v(n)$ and generates an initial scale for discussion:

$$Min\,v(n)$$
$$S.T.: \forall a,b \in X : aPb \Rightarrow v(a) \geq v(b) + 1$$
$$\quad \forall a,b \in X : aIb \Rightarrow v(a) = v(b)$$
$$\quad \forall (a,b),(c,d) \in X,$$
if the difference of attractiveness between
a and b is bigger than between c and d, then:
$$v(a) - v(b) \geq v(c) - v(d) + 1 + \delta(a,b,c,d)$$
$$v(a^-) = 0$$
where:
n is an element of X so that
$$\forall a,b,c,... \in X : n(P \cup I)a,b,c,...$$
a^- is an element of X so that
$$\forall a,b,c,... \in X : a,b,c,...(P \cup I)a^-$$
$\delta(a,b,c,d)$ is the minimal number of categories
of difference of attractiveness between
the difference of attractiveness
between a and b and the difference
of attractiveness between c and d.
$$\tag{11}$$

Technically, it should be pointed out that n is the most attractive alternative of X (or at least as attractive as the others) (i.e. $n\,(P \cup I)\,a, b, c,...$), and the minimization of its value guarantees the minimal length of the initial scale. Also, a^- is the least attractive alternative of X (or at least as attractive as the others) (i.e. $a, b, c,...\,(P \cup I)\,a^-$), and the respective value should be considered the "zero" of the scale (Bana e Costa *et al.* 2008). Through the additive model presented by equation (7), an overall value for each alternative can be estimated.

5. EXPERIMENTAL RESULTS

Our experiment was conducted after the intervention of the European Central Bank, European Union and International Monetary Fund (i.e. the so called "Troika") in Portugal, which required credit risk analysis of mortgage loans to be more cautious and forced Portuguese banks to impose higher credit underwriting standards.

Considering that AHP, Delphi and MACBETH have been characterized as very effective in transparently handling trade-offs among evaluation criteria (cf. Belton, Stewart 2002; Saaty 2008; Ferreira, Monteiro Barata 2011; Bana e Costa *et al.* 2012; Ferreira *et al.* 2012, 2013), our experiment consisted in applying and examining the potential value of each of these three methods in terms of trade-offs readjustment in a given structure

of pre-established risk evaluation criteria (see Appendix 1). In this sense, different aspects will be addressed in the following subsections, namely: actors involved; problem definition; technical procedures undertaken; analyses carried out; comparison and discussion of results and, finally, recommendations. For illustrative purposes, these aspects will be discussed using the credit scoring system of one of the five largest banks operating in Portugal as reference.

5.1. Actors involved

Decision makers are expected to play an active role when dealing with group decision support systems (cf. Ferreira *et al.* 2011b). In order to compare the performance of each of the three decision support systems under analysis, a panel of five experts in risk analysis of mortgage loans was formed and a group meeting was organized to discuss the problem and collect the views of the experts on the different systems. The group meeting was conducted by a team of two facilitators (i.e. researchers, scientists), assisted by a communication technician, who was responsible for registering the outcomes of the session.

5.2. Problem definition

As already outlined, our experiment aimed to apply the AHP, Delphi and MACBETH methods to a structure of administratively pre-established risk evaluation criteria and compare their performance. This allowed us to test which one of these three methods offers the greatest potential to provide decision makers with the most effective and fairer mortgage risk evaluation system.

In practical terms, our research problem consisted in eliciting value judgments from a panel of mortgage credit risk analysts and, based on the outcomes, readjust trade-offs among the criteria considered (see again Table 1). It is important to clarify that it was not an objective of the intervention to introduce changes in the current mortgage loan risk evaluation system in terms of criteria selection and/or operational structure (see Appendix 1). This structure is administratively established by the bank and we were not asked (or allowed) to change it.

5.3. Evaluation criteria and impact levels

The criteria used by the banking institution under analysis to assess the risk of mortgage loans and the way these criteria are structured into a credit scoring system are illustrated in Figure 1.
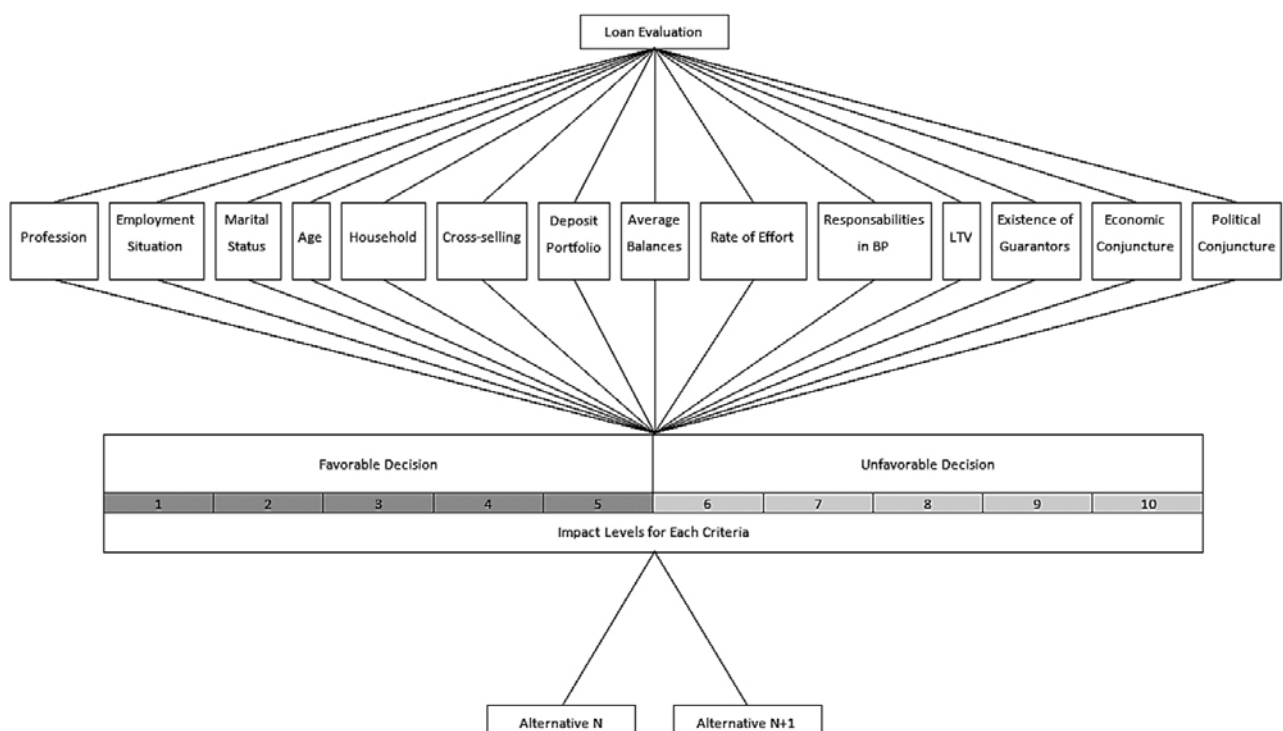
Fig. 1. Structure of evaluation criteria and respective impact levels (Ferreira *et al.* 2014a)

Complementing the analysis of Figure 1, it is important to highlight the existence of two decision levels. The first decision level is based on a discrete scale from "1" to "10" and projects the partial evaluation of the credit applications in each criterion involved. This means that there are ten administratively defined levels of partial evaluation, and credit scores between "1" and "5" support partial credit approvals. On the other hand, the credit should be refused if the credit scores are between "6" and "10" (cf. Appendix 1). The second decision level considers the weights of the criteria and is based on a continuous scale from "1" to "10", allowing the projection of an overall score for each mortgage loan application. It is precisely at this second decision level that our experiment is focused. In particular, we aim to explore and compare the applicability of the AHP, Delphi and MACBETH methods in terms of trade-off readjustments in the context of credit scoring. Figure 1 was shown to the experts during the group session. This allowed us to: (1) explain the research problem; (2) encourage discussion among panelists; and (3) establish the basis for a proper elicitation of value preferences.

5.4. Judgments, comparisons and trade-off procedures

In this subsection, we individually present the way each method was applied. Because the panel members did not know each other before the group meeting, we decided to apply the Delphi technique first. This allowed us to preserve the anonymity of the participants and avoid group influences.

5.4.1. Application of the Delphi technique

At the beginning of the session, explanations about the Delphi technique were given to the panel members. In particular, the individual survey (see Appendix 2) was meticulously explained to each of the experts and they were then invited to express their perceptions in terms of trade-offs among criteria. Once collected, the individual responses of the decision makers were inserted into an Excel file and a second survey (see Appendix 3), contemplating the statistics of the first round, was given to the panelists. The panel members where then invited to compare the results obtained, rate their answers again and, if necessary, re-rank the answers given in the first round. After collecting and analyzing the individual responses of the second round, it was possible to note that the changes from the first round to the second were not significant. Table 2 presents the statistical outputs obtained after the second round.

The questionnaire took approximately 30 minutes to fill in and was followed by 10 minutes of discussion in order to collect the views of the participants regarding the method. According to the panel members, the Delphi technique is easy to understand and reveals great potential in practical terms (further discussion and comparison of results are presented in subsection 5.7).

5.4.2. Application of an AHP-based process

Following the example of the Delphi session, the testing of the AHP-based process was also preceded by an explanation about the AHP method and

Table 2. Statistics of the second round

Variables	N	Mean %	Median %	Standard deviation %	Coefficient of variation %
Profession	5	5.2	5	1.7888	0.344
Employment situation	5	8.3	7	7.1203	0.858
Marital status	5	1.2	1	0.8366	0.697
Age	5	2.5	2	1	0.400
Household	5	7.1	4	7.3348	1.033
Cross-selling	5	4.6	3	3.7815	0.822
Deposit portfolio	5	5.4	5	4.1593	0.770
Average balance	5	10.5	10	3.0822	0.296
Rate of effort	5	14.8	15	4.8166	0.325
Responsibilities in BP	5	11.4	15	5.6833	0.499
LTV	5	11.6	15	4.7749	0.412
Existence of guarantors	5	8.1	10	2.6077	0.322
Economic situation of the country	5	5.6	7	3.9115	0.698
Political situation of the country	5	3.7	4	2.4899	0.673

'Saaty's fundamental scale'. The decision makers were then asked to pairwise compare and order the criteria by decreasing overall importance. This procedural step lasted approximately one hour and was the starting point for the projection of priorities. Figure 2 presents the questionnaire of pairwise comparisons, where the priorities were projected. This process lasted another hour and was conducted using the *Super Decisions* software (http://www.creativedecisions.net/). The outputs of our AHP-based application are shown in Table 3.

Once the trade-offs had been obtained through the AHP-based process, they were provided to the

panel members for discussion (details and comparison of results are presented in subsection 5.7). From a technical perspective, it should be pointed out that the results obtained present a good consistency ratio of 0.026412 (i.e. below 0.10) (cf. Saaty 1980; Yurdakul, İç 2004; Xu, Zhang 2009; Perez-Gladish, M'Zali 2010).

5.4.3. Application of the MACBETH approach

The testing of the MACBETH approach was also preceded by an explanation of this technique. In particular, special emphasis was given to the categories of semantic differences of attractiveness introduced by Bana e Costa and Vansnick (1994). The process started with the ranking of the evaluation criteria (also known as Fundamental Points of Views (FPVs) in the MACBETH literature) by decreasing overall attractiveness. Once this step was concluded, the panel members were invited to pairwise compare the criteria and project their value judgments using the categories previously presented. The process was conducted using the M-MACBETH software (http://www.m-macbeth.com/), and the matrix of comparisons took approximately sixty minutes to complete (Fig. 3). Table 4 summarizes the results obtained, which were provided to the decision makers for further discussion.

Fig. 2. Pairwise comparisons [Super Decisions software] (Ferreira *et al.* 2014a)

Table 3. Criteria ranking and "AHP" trade-offs

Criteria	Nomenclature	Ranking	(New) Weights %
Profession	CRT01	7	5.79
Employment situation	CRT02	12	1.74
Marital status	CRT03	13	1.38
Age	CRT04	11	2.10
Household	CRT05	10	2.80
Cross-selling	CRT06	8	4.66
Deposit port-folio	CRT07	6	7.44
Average bal-ances	CRT08	5	8.68
Rate of effort	CRT09	2	15.97
Responsibili-ties in BP	CRT10	1	19.30
LTV	CRT11	3	14.04
Existence of guarantors	CRT12	4	11.59
Economic situation of the country	CRT13	9	3.37
Political situation of the country	CRT14	14	1.14

Source: Adapted from Ferreira *et al.* (2014a).

Table 4. Criteria ranking and "MACBETH" trade-offs

Criteria	Nomenclature	Ranking	(New) Weights %
Profession	FPV01	7	7.78
Employment situation	FPV02	12	3.39
Marital status	FPV03	13	2.44
Age	FPV04	11	5.08
Household	FPV05	10	5.75
Cross-selling	FPV06	8	7.11
Deposit port-folio	FPV07	6	8.46
Average bal-ances	FPV08	5	9.07
Rate of effort	FPV09	2	11.50
Responsibilities in BP	FPV10	1	11.84
LTV	FPV11	3	11.17
Existence of guarantors	FPV12	4	9.64
Economic situation of the country	FPV13	9	6.43
Political situation of the country	FPV14	14	0.34

Source: Adapted from Ferreira *et al.* (2014b).

Weighting (Lending Decision)

	[FPV10]	[FPV09]	[FPV11]	[FPV12]	[FPV08]	[FPV07]	[FPV01]	[FPV06]	[FPV13]	[FPV05]	[FVP04]	[FPV02]	[FPV03]	[FPV14]	[All Inf.]	Current scale
[FPV10]	no	very weak	weak	weak	weak	weak	moderate	moderate	moderate	moderate	moderate	strong	vstrg-extr	vstrg-extr	extreme	positive	11.84
[FPV09]		no	very weak	vweak-weak	weak	weak-mod	weak-mod	weak-mod	weak-mod	weak-mod	moderate	v. strong	v. strong	extreme	positive	11.50
[FPV11]			no	vweak-weak	weak	weak-mod	weak-mod	weak-mod	weak-mod	weak-mod	moderate	v. strong	v. strong	extreme	positive	11.17
[FPV12]				no	vweak-weak	weak	weak-mod	weak-mod	weak-mod	weak-mod	moderate	mod-strg	mod-strg	vstrg-extr	positive	9.64
[FPV08]					no	vweak-weak	weak	weak	weak	weak-mod	moderate	moderate	mod-strg	v. strong	positive	9.07
[FPV07]						no	weak	weak	weak	weak-mod	moderate	moderate	mod-strg	v. strong	positive	8.46
[FPV01]							no	weak	weak	weak-mod	weak-mod	weak-mod	mod-strg	strong	positive	7.78
[FPV06]								no	weak	weak	weak-mod	weak-mod	mod-strg	strong	positive	7.11
[FPV13]									no	weak	weak	weak	moderate	moderate	positive	6.43
[FPV05]										no	weak	weak	weak-mod	moderate	positive	5.75
[FVP04]											no	weak	weak-mod	moderate	positive	5.08
[FPV02]												no	weak	weak	positive	3.39
[FPV03]													no	weak	positive	2.44
[FPV14]														no	positive	0.34
[All Inf.]															no	0.00

Consistent judgements

Fig. 3. Matrix of comparisons and proposed trade-offs [M-MACBETH software]
(Ferreira *et al.* 2014b)

Once the "new" trade-offs were obtained using the three different elicitation techniques, we tested the results by assessing the risk of two real mortgage loan applications, as discussed in the next subsection.

5.5. Evaluating mortgage loan credit risk

To test the trade-offs obtained from the application of the three methods in analysis, information on mortgage credit applications (identified as "Betas" from now on) was requested from the bank. Table 5 presents the information that was kindly provided under conditions of strict anonymity and refers to two mortgage loan applications. Whilst we would have liked to have data for a larger sample of applications, the information provided was an administrative decision over which we had no control.

Based on the information presented in Table 5, we were then able to determine the partial scores obtained by the two credit applications regarding each evaluation criterion and, by combining these scores with the respective weights, calculate the overall risk score for each application. Criteria weights and rankings, as well as the partial and overall scores obtained during the experiment for the two Betas, are presented in Table 6.

Although Delta 1 and Delta 2 are two real mortgage loan applications, they are not representative of the profiles of all applications made to the bank under analysis. However, the results displayed in Table 6 are very useful for illustrative purposes.

Table 5. Data from two anonymous mortgage credit applications

	Beta 1	Beta 2
Profession	Customer with a profession in a stable industry	Customer with a profession in a stable industry
Employment status	Customer with three years of permanent employment	Customer with sixteen years of permanent employment
Marital status	Single	Married
Age	25–35	36–50
Household	1 Element	4 Elements
Cross-selling	Customer who holds a single bank account and households his/her salary	Customer who holds several financial products and households his/her salary in the institution
Average balance	[0–1.000€]	>2.500€
Deposit portfolio	[2.500€-25.000€[	[25.000–50.000€]
Rate of effort	34.60%	41.40%
Responsibilities in BP	Without other responsibilities	Without other responsibilities
LTV	77.78%	53.16%
Existence of guarantors	No	No
Economic situation of the country	Stable with difficulties	Stable with difficulties
Political situation of the country	Stable	Stable

Source: Administrative information.

Table 6. Summary of the experiment's outcomes

Beta 1

	Partial score a)	Current weight %	Delphi W(%)	Delphi Ranking	AHP W(%)	AHP Ranking	MACBETH W(%)	MACBETH Ranking
Profession	3	12.00	5.2↓	10	5.79↓	7	7.78↓	7
Employment status	2	8.00	8.3↑	5	1.74↑	12	3.39↑	12
Marital status	1	3.75	1.2↓	14	1.38↓	13	2.44↓	13
Age	2	2.50	2.5●	13	2.10↓	11	5.08↑	11
Household	1	3.75	7.1↑	7	2.80↓	10	5.75↑	10
Cross-selling	4	10.00	4.6↓	11	4.66↓	8	7.11↓	8
Deposit portfolio	4	2.50	5.4↑	9	7.44↑	6	8.46↑	6
Average balance	3	2.50	10.5↑	4	8.68↑	5	9.07↑	5
Rate of effort	2	15.00	14.8↓	1	15.97↑	2	11.50↓	2
Responsibilities in BP	1	15.00	11.4↓	3	19.30↑	1	11.84↓	1
LTV	4	15.00	11.6↓	2	14.04↓	3	11.17↓	3
Existence of guarantors	6	5.00	8.1↑	6	11.59↑	4	9.64↑	4
Economic situation of the country	4	2.50	5.6↑	8	3.37↑	9	6.43↑	9
Political situation of the country	3	2.50	3.7↑	12	1.14↓	14	0.34↓	14
Overall Score (*CR = 0.026412)	--	2.745	2.865	--	2.975*	--	3.021	--

Beta 2

	Partial score a)	Current weight %	Delphi W(%)	Delphi Ranking	AHP W(%)	AHP Ranking	MACBETH W(%)	MACBETH Ranking
Profession	3	12.00	5.2↓	10	5.79↓	7	7.78↓	7
Employment status	1	8.00	8.3↑	5	1.74↑	12	3.39↓	12
Marital status	1	3.75	1.2↓	14	1.38↓	13	2.44↓	13
Age	4	2.50	2.5●	13	2.10↓	11	5.08↑	11
Household	4	3.75	7.1↑	7	2.80↓	10	5.75↑	10
Cross-selling	2	10.00	4.6↓	11	4.66↓	8	7.11↓	8
Deposit portfolio	3	2.50	5.4↑	9	7.44↑	6	8.46↑	6
Average balance	1	2.50	10.5↑	4	8.68↑	5	9.07↑	5
Rate of effort	2	15.00	14.8↓	1	15.97↑	2	11.50↓	2
Responsibilities in BP	1	15.00	11.4↓	3	19.30↑	1	11.84↓	1
LTV	3	15.00	11.6↓	2	14.04↓	3	11.17↓	3
Existence of guarantors	7	5.00	8.1↑	6	11.59↑	4	9.64↓	4
Economic situation of the country	4	2.50	5.6↑	8	3.37↑	9	6.43↑	9
Political situation of the country	3	2.50	3.7↑	12	1.14↓	14	0.34↓	14
Overall Score (*CR = 0.026412)	--	2.453	2.654	--	2.718*	--	2.837	--

a) Partial evaluation on the first decision level (see Appendix 1);
↑ Positive weight variation when compared to the current system;
● Null weight variation when compared to the current system;
↓ Negative weight variation when compared to the current system.

As shown in Table 6, the overall scores obtained by using each of the methods analyzed (including the existing evaluation system) suggest the approval of the credit applications by the bank. Although the overall scores are all below 5, recommending approval of the applications, the results obtained with the Delphi, AHP and MACBETH methods are more conservative than those obtained with the actual system, signaling a higher risk of the applications. The results in Table 6 also indicate that the use of these elicitation techniques introduced several changes to the pre-existing weights. Whilst the weight of some of the evaluation criteria increased (e.g. *deposit portfolio*; *average balance*), others decreased significantly (e.g. *profession*; *cross-selling*). The next step consisted in comparing the methods and discussing the outcomes of the experiment and the accuracy of the results.

5.6. The pre-questionnaire

Following the example of Perini *et al.* (2009), we decided to apply a pre-questionnaire before any comparison between methodologies. Among other things, this procedure was important to know whether the decision makers understood both the objectives of the group meeting and the operative mechanisms of the AHP, Delphi and MACBETH methods. The questions were easy to understand and the answers were given on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Table 7 presents the results of the pre-questionnaire.

In broad terms, and following Table 7, the panel members reported that they understood the Delphi, AHP and MACBETH methods (median of 5), the differences among methods (median of 4) and the objectives of the group meeting (median of 5). Most important, perhaps, is the fact that these statistical results confirm that the credit

analysts understood what they were supposed to do in terms of tasks required (median of 5). The next subsection presents a comparative analysis among methods.

5.7. Comparative analysis

The literature on decision support is fertile in presenting strengths and weaknesses of the three methods in analysis. Regarding the Delphi technique, it has been characterized as a tool that allows consensus based on controlled feedback of the answers and reflection on earlier judgments. The technique has been applied in the treatment of several different themes and, because it is anonymous, it allows participants to provide answers without the influence of the organization hierarchy. On the negative side, because the method does not require the physical presence of the agents involved in the process, the questions may be misunderstood by the respondents, jeopardizing the results. Additionally, because the technique is based on rounds and the survey is always the same, respondents tend to drop out of the process (for further details on the strengths and weaknesses of the Delphi technique, see e.g. Dalkey, Helmer 1963; Dalkey 1969; Hsu, Sandford 2007).

As for the AHP method, one of its most significant strengths results from requiring decision makers to compare pairs of alternatives, measuring the degree of inconsistency present in pairwise judgments and ensuring that only justifiable orders are used. Despite its intrinsic simplicity and ability to assess quantitative and qualitative factors, different types of criticism have been discussed in the literature. For example, the method has been criticized not only for the possibility of exhibiting rank reversal (Belton, Gear 1983; Boucher *et al.* 1997), but also for the use of the eigenvalue procedure to derive priorities (Bana e Costa, Vansnick 2008) (for further discussion on the strengths and

Table 7. Statistics of the pre-questionnaire

Questions – values for the answers range from 1 to 5	N	Mean	Median	Standard deviation
I understand what the objectives of this group meeting are	5	5.0	5	0.000
The questions and tasks asked to me are clear	5	4.6	5	0.548
I understand how to use Delphi and its practical utility	5	4.6	5	0.548
I understand how to use AHP and its practical utility	5	5.0	5	0.000
I understand how to use MACBETH and its practical utility	5	5.0	5	0.000
The differences among methods are clear to me	5	4.2	4	0.837
I have already used these (or similar) methods before this group meeting	5	3.0	3	1.581

limitations of the AHP method, see e.g. Olson *et al.* 1995; Yurdakul, İç 2004; Alonso, Lamata 2006; Saaty 2008; Perez-Gladish, M'Zali 2010).

As in the AHP case, the MACBETH approach is also based on pairwise comparisons, which, according to Dyer and Forman (1992), are easy to make, justify and agree on. As an interactive decision support technique, MACBETH has been characterized as following a constructivist approach, being easy to understand and solidly supported on mathematical background. Still, as in any other pairwise comparison-based approach, MACBETH requires a thorough and prior preparation of the information, namely in terms of mutual preferential independence tests, which depend on the number of criteria involved and number of pairwise comparisons to be made. In particular, *"the technique requires an enormous willingness on the part of decision makers, and a high dedication on the part of the facilitator […] the matrices completion can become a demanding task for the actors involved in the process and, as such, difficulties in gathering data may arise"* (Ferreira 2013: 443) (for further details on the strengths and limitations of the MACBETH technique, see e.g. Bana e Costa, Vansnick 1994; Belton, Stewart 2002; Ferreira *et al.* 2011a).

Whilst different types of strengths and limitations can be pointed to the three methods in analysis, it should be noted that, in accordance with Weber and Borcherding (1993), there is no such thing as a superior elicitation methodology and method choice should always be dependent on the decision context. This remark seems to be further supported by Ananda and Herath (2009) and Zhou and Ang (2009), who argue that it is very difficult to prove that one method or technique is superior to another in supporting the decision making process. From this premise, and because the AHP, Delphi and MACBETH methods have been recognized in the literature for being simple and facilitating trade-off calculations, in the last part of the meeting a comparative survey was applied to get the perspective of the decision makers regarding the methods and the accuracy of the results. It should be recalled that we focused on five lines of comparison, namely: ease of use; time-consumption; ease of applicability; accuracy; and overall evaluation. As in the pre-questionnaire, a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) was used. Table 8 presents the results obtained.

As shown in Table 8, the results indicate that in the particular context under analysis Delphi exceeds AHP and MACBETH in terms of ease of use, time-consumption and ease of applicability. In terms of the accuracy of the results, the differences obtained between AHP and MACBETH

Table 8. Comparison of methods

Questions – Values for the answers range from 1 to 5	N	Mean	Median	Standard deviation
The method is easy to understand and use:				
Delphi	5	4.0	4	0.707
AHP	5	3.8	4	0.837
MACBETH	5	3.6	4	0.548
The method is not time-consuming:				
Delphi	5	4.2	4	0.447
AHP	5	3.4	3	0.548
MACBETH	5	3.4	3	0.548
The method provides accurate results:				
Delphi	5	3.8	4	0.837
AHP	5	4.0	4	0.707
MACBETH	5	4.0	4	0.707
The method has great practical utility:				
Delphi	5	4.4	4	0.548
AHP	5	3.8	4	0.837
MACBETH	5	3.8	4	0.447
Overall, the method is good:				
Delphi	5	4.0	4	1.000
AHP	5	4.2	4	0.447
MACBETH	5	4.0	4	0.707

are not significant (cf. Table 6), and both methods perform better than Delphi. The majority of the decision makers considered AHP the "overall best" approach.

In the last thirty minutes of the meeting, the panel members were asked to analyze and discuss the outcomes of the session. Even considering that there is no superior elicitation method and that choice should depend on the decision context, as previously discussed, there was a generalized consensus that *"the adjustments made represent more caution in terms of credit approval, corroborating the Basel guidelines"* (in the decision makers' own words). Furthermore, it was also agreed that the use of the Delphi, AHP or MACBETH methods has the potential to make the scoring procedures more defensible and more adaptable to the idiosyncratic characteristics of specific bank branches, as they easily allow for the inclusion of the value systems of the decision makers of those branches into the scoring process.

6. DISCUSSION AND CONCLUSIONS

Mortgage loans are among the most highly sought financial products worldwide. However, considering the severe restrictions on access to credit resulting from the current economic crisis, mortgage loan risk evaluation has been considered paramount for banking institutions and is now accompanied by higher credit underwriting standards.

As pointed out before, banks have improved their credit-scoring risk systems over the years. Still, it is also acknowledged that the progress achieved does not mean that the current approaches are without limitations. In particular, following Ferreira *et al.* (2011a), there is a lack of transparency in the way trade-offs among evaluation criteria are defined. Starting from this observation, we applied three decision support methods (i.e. AHP, Delphi and MACBETH) to a hierarchical structure of pre-established risk evaluation criteria, which is currently in use by one of the largest banks operating in Portugal, and compared their performance from the point of view of five credit analysts. Five major lines of comparison were considered, namely: ease of use; time-consumption; ease of applicability; accuracy; and overall evaluation. Whilst we found no major differences on the performances of the three methods, our results indicate that Delphi surpasses AHP and MACBETH in terms of ease of use, time-consumption and ease of applicability. In terms of the accuracy of the results, the differences obtained between AHP and

MACBETH are not significant (cf. Table 6), and both methods perform better than Delphi. Most of the decision makers considered AHP the "overall best" approach. Caution needs to be taken, however, in interpreting our findings as this was an exploratory study; the number of participants in the experiment was small; and the application of the three decision support methods was sequential. The sequential use of the methods means that when the credit analysts used the MACBETH approach they were already familiar with the pairwise comparison procedure. Interestingly, however, this does not seem to have had a significant impact on the perceptions of the credit analysts as they still considered the AHP easier to understand and use than the MACBETH. Furthermore, one needs to bear in mind that the performance of elicitation techniques is always dependent on the actors involved and of the context under analysis (cf. Weber, Borcherding 1993; Ananda, Herath 2009; Zhou, Ang 2009). Whilst we did not find significant differences between the performances of the three methods, which prevents us from proposing one in detriment of the other two, our results show that any of the methods considered has the potential to add value to the existing credit scoring systems. In particular, the results suggest that these methods have the potential to provide decision makers with a more discerning framework and assist them in making better informed decisions. There was a generalized consensus amongst the participants that *"the adjustments made represent more caution in terms of credit approval, corroborating the Basel guidelines"* (in the panelists' own words). Following this, and considering the limitations of our experiment, we recommend for future work: a) different panel studies with a higher number of participants and within other banking institutions; b) the development of similar experiments but involving also the comparison of other multiple criteria decision methods (for a review, see Zavadskas, Turskis 2011); c) surveys of comparisons among different methods; and d) sensitivity and robustness analyses in order to explore which method provides more robust and reliable risk assessments. Improvements and updates will strengthen the results and comparisons presented in this study.

REFERENCES

Alonso, J.; Lamata, M. 2006. Consistency in the analytic hierarchy process: a new approach, *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* 14(4): 445–459. http://dx.doi.org/10.1142/S0218488506004114

Altman, E.; Saunders, A. 1997. Credit risk measurement: developments over the last 20 years, *Journal of Banking & Finance* 21(11/12): 1721–1742. http://dx.doi.org/10.1016/S0378-4266(97)00036-8

Ananda, J.; Herath, G. 2009. A critical review of multicriteria decision making methods with special reference to forest management and planning, *Ecological Economics* 68(10): 2535–2548. http://dx.doi.org/10.1016/j.ecolecon.2009.05.010

Bana e Costa, C.; Vansnick, J. 1994. MACBETH: an interactive path towards the construction of cardinal value functions, *International Transactions in Operational Research* 1(4): 489–500. http://dx.doi.org/10.1016/0969-6016(94)90010-8

Bana e Costa, C.; Vansnick, J. 2008. A critical analysis of the eigenvalue method used to derive priorities in the AHP, *European Journal of Operational Research* 187(3): 1422–1428. http://dx.doi.org/10.1016/j.ejor.2006.09.022

Bana e Costa, C.; De Corte, J.; Vansnick, J. 2005. On the mathematical foundations of MACBETH, in Figueira, J.; Greco, S.; Ehrgott, M. (Eds.). *Multiple criteria decision analysis: the state of the art surveys*. New York: Springer, 409–442. http://dx.doi.org/10.1007/0-387-23081-5_10

Bana e Costa, C.; De Corte, J.; Vansnick, J. 2012. MACBETH, *International Journal of Information Technology and Decision Making* 11(2): 359–387. http://dx.doi.org/10.1142/S0219622012400068

Bana e Costa, C.; Lourenço, J.; Chagas, M.; Bana e Costa, J. 2008. Development of reusable bid evaluation models for the Portuguese electric transmission company, *Decision Analysis* 5(1): 22–42. http://dx.doi.org/10.1287/deca.1080.0104

Barth, J.; Hollans, H. 2012. Housing finance: deposit guarantees, in *International encyclopedia of housing and home*. Amsterdam: Elsevier Ltd. 450–455. http://dx.doi.org/10.1016/B978-0-08-047163-1.00221-6

Belton, V.; Gear, A. 1983. On a shortcoming of Saaty's method of analytic hierarchies, *Omega* 11(3): 228–230. http://dx.doi.org/10.1016/0305-0483(83)90047-6

Belton, V.; Stewart, T. 2002. *Multiple criteria decision analysis: an integrated approach*. Dordrecht: Kluwer Academic Publishers. http://dx.doi.org/10.1007/978-1-4615-1495-4

Beltratti, A.; Stulz, R. 2012. The credit crisis around the globe: why did some banks perform better?, *Journal of Financial Economics* 105(1): 1–17. http://dx.doi.org/10.1016/j.jfineco.2011.12.005

Blackburn, M.; Vermilyea, T. 2012. The prevalence and impact of misstated incomes on mortgage loan applications, *Journal of Housing Economics* 21(2): 151–168. http://dx.doi.org/10.1016/j.jhe.2012.04.003

Boucher, T.; Gogus, O.; Wicks, E. 1997. A comparison between two multiattribute decision methodologies used in capital investment decision analysis, *The Engineering Economist* 42(39): 179–201. http://dx.doi.org/10.1080/00137919708903178

Chambers, M.; Garriga, C.; Schlagenhauf, D. 2009. The loan structure and housing tenure decisions in an equilibrium model of mortgage choice, *Review of Economic Dynamics* 12(3): 444–468. http://dx.doi.org/10.1016/j.red.2009.01.003

Dalkey, N. 1969. *The Delphi method: an experimental study of group opinion* [online]. United States Air Force Project Rand, Rand Santa Monica Company. Available at: http://www.rand.org/content/dam/rand/pubs/research_memoranda/2005/RM5888.pdf [accessed June 2013]

Dalkey, N.; Helmer, O. 1963. An experimental application of the Delphi method to the use of experts, *Management Science* 9(3): 458–467. http://dx.doi.org/10.1287/mnsc.9.3.458

Doumpos, M.; Zopounidis, C. 2001. Assessing financial risks using a multicriteria sorting procedure: the case of country risk assessment, *Omega* 29(1): 97–109. http://dx.doi.org/10.1016/S0305-0483(00)00028-1

Doumpos, M.; Zopounidis, C. 2011. A multicriteria outranking modeling approach for credit rating, *Decision Sciences* 42(3): 721–742. http://dx.doi.org/10.1111/j.1540-5915.2011.00328.x

Doumpos, M.; Kosmidou, K.; Baourakis, G.; Zopounidis, C. 2002. Credit risk assessment using a multicriteria hierarchical discrimination approach: a comparative analysis, *European Journal of Operational Research* 138(2): 392–412. http://dx.doi.org/10.1016/S0377-2217(01)00254-5

Dyer, R.; Forman, E. 1992. Group decision support with the analytic hierarchy process, *Decision Support Systems* 8(2): 99–124. http://dx.doi.org/10.1016/0167-9236(92)90003-8

Ferreira, F. 2013. Measuring trade-offs among criteria in a balanced scorecard framework: possible contributions from the multiple criteria decision analysis research field, *Journal of Business Economics and Management* 14(3): 433–447. http://dx.doi.org/10.3846/16111699.2011.631744

Ferreira, F.; Monteiro Barata, J. 2011. A snapshot of the Portuguese e-banking activity: insights and a conceptual framework to allocate strategic hindrances, *International Journal of Electronic Business* 9(3): 238–254. http://dx.doi.org/10.1504/IJEB.2011.042544

Ferreira, F.; Santos, S.; Dias, V. 2014a. An AHP-based approach to credit risk evaluation of mortgage loans, *International Journal of Strategic Property Management* 18(1): 38–55. http://dx.doi.org/10.3846/1648715X.2013.863812

Ferreira, F.; Santos, S.; Rodrigues, P. 2011a. Adding value to bank branch performance evaluation using

cognitive maps and MCDA: a case study, *Journal of the Operational Research Society* 62(7): 1320–1333. http://dx.doi.org/10.1057/jors.2010.111

Ferreira, F.; Santos, S.; Rodrigues, P. 2011b. From traditional operational research to multiple criteria decision analysis: basic ideas on an evolving field, *Problems and Perspectives in Management* 9(3): 114–121.

Ferreira, F.; Santos, S.; Marques, C.; Ferreira, J. 2014b. Assessing credit risk of mortgage lending using MACBETH: a methodological framework, *Management Decision* 52(2): 182–206. http://dx.doi.org/10.1108/MD-01-2013-0021

Ferreira, F.; Spahr, R.; Gavancha, I.; Çipi, A. 2013. Readjusting trade-offs among criteria in internal ratings of credit-scoring: an empirical essay of risk analysis in mortgage loans, *Journal of Business Economics and Management* 14(4): 715–740. http://dx.doi.org/10.3846/16111699.2012.666999

Ferreira, F.; Spahr, R.; Santos, S.; Rodrigues, P. 2012. A multiple criteria framework to evaluate bank branch potential attractiveness, *International Journal of Strategic Property Management* 16(3): 254–276. http://dx.doi.org/10.3846/1648715X.2012.707629

Hsu, C.; Sandford, B. 2007. The Delphi technique: making sense of consensus, *Practical Assessment, Research and Evaluation* 12(10). Available online at http://pareonline.net/pdf/v12n10.pdf [accessed June 2013].

Jacobson, T.; Roszbach, K. 2003. Bank lending policy, credit scoring and value-at-risk, *Journal of Banking & Finance* 27(4): 615–633. http://dx.doi.org/10.1016/S0378-4266(01)00254-0

Junior, H. 2008. Multicriteria approach to data envelopment analysis, *Pesquisa Operacional* 28(2): 231–242.

Kowalski, T.; Shachmurove, Y. 2011. The financial crisis: what is there to learn?, *Global Finance Journal* 22(3): 238–247. http://dx.doi.org/10.1016/j.gfj.2011.10.014

Laininen, P.; Hämäläinen, R. 2003. Analyzing AHP matrices by regression, *European Journal of Operational Research* 148(3): 514–524. http://dx.doi.org/10.1016/S0377-2217(02)00430-7

Leow, M.; Mues, C. 2012. Predicting loss given default (LGD) for residential mortgage loans: a two-stage model and empirical evidence for UK bank data, *International Journal of Forecasting* 28(1): 183–195. http://dx.doi.org/10.1016/j.ijforecast.2011.01.010

Lima, S.; Vaz, M.; Baptista, M.; Ramos, A.; Leitão, M.; Cunha, M. 1995. *Operações bancárias II* [Banking operations II]. 6th ed. Instituto Formação Bancária, Lisbon.

Linstone, H.; Turoff, M. 1975. *The Delphi method: techniques and applications*. Reading, Mass.: Addison-Wesley.

Lopez, J.; Saidenberg, M. 2000. Evaluating credit risk models, *Journal of Banking & Finance* 24(1): 151–165. http://dx.doi.org/10.1016/S0378-4266(99)00055-2

Mari, C.; Renò, R. 2005. Credit risk analysis of mortgage loans: an application to the Italian market, *European Journal of Operational Research* 163(1): 83–93. http://dx.doi.org/10.1016/j.ejor.2003.12.007

Olson, D.; Fliedner, G.; Currie, K. 1995. Comparison of the REMBRANDT system with analytic hierarchy process, *European Journal of Operational Research* 82(3): 522–539. http://dx.doi.org/10.1016/0377-2217(93)E0340-4

Pau, L.; Tambo, T. 1990. Knowledge-based mortgage-loan credit granting and risk assessment, *Journal of Economic Dynamics and Control* 14(2): 255–262. http://dx.doi.org/10.1016/0165-1889(90)90020-H

Perez-Gladish, B.; M'Zali, B. 2010. An AHP-based approach to mutual funds' social performance measurement, *International Journal of Multicriteria Decision Making* 1(1): 103–126. http://dx.doi.org/10.1504/IJMCDM.2010.033689

Perini, A.; Ricca, F.; Susi, A. 2009. Tool-supported requirements prioritization: comparing the AHP and CBRank methods, *Information and Software Technology* 51(6): 1021–1032. http://dx.doi.org/10.1016/j.infsof.2008.12.001

Puri, M.; Rocholl, J.; Steffen, S. 2011. Global retail lending in the aftermath of the US financial crisis: distinguishing between supply and demand effects, *Journal of Financial Economics* 100(3): 556–578. http://dx.doi.org/10.1016/j.jfineco.2010.12.001

Saaty, T. 1980. *The analytic hierarchy process: planning, priority setting, resource allocation*. New York: McGraw-Hill.

Saaty, T. 2008. Relative measurement and its generalization in decision making: why pairwise comparisons are central in mathematics for the measurement of intangible factors – the analytic hierarchy/network process, *Real Academia de Ciencias* 102(2): 251–318. http://dx.doi.org/10.1007/bf03191825

Scholl, A.; Manthey, L.; Helm, R.; Steiner, M. 2005. Solving multiattribute design problems with analytic hierarchy process and conjoint analysis: an empirical comparison, *European Journal of Operational Research* 164(3): 760–777. http://dx.doi.org/10.1016/j.ejor.2004.01.026

Twala, B. 2010. Multiple classifier application to credit risk assessment, *Expert Systems with Applications* 37(4): 3326–3336. http://dx.doi.org/10.1016/j.eswa.2009.10.018

Wang, G.; Hao, J.; Mab, J.; Jiang, H. 2011. A comparative assessment of ensemble learning for credit scoring, *Expert Systems with Applications* 8(1): 223–230. http://dx.doi.org/10.1016/j.eswa.2010.06.048

Weber, M.; Borcherding, K. 1993. Behavioral influences on weight judgments in multiattribute decision making, *European Journal of Operational Research* 67(1): 1–12. http://dx.doi.org/10.1016/0377-2217(93)90318-H

Xu, Y.; Zhang, Y. 2009. A online credit evaluation method based on AHP and SPA, *Communications in Nonlinear Science and Numerical Simulation* 14(7): 3031–3036. http://dx.doi.org/10.1016/j.cnsns.2008.10.018

Yeager, T. 2011. Causes, consequences and cures of the subprime financial crisis, *Journal of Economics and Business* 63(6): 345–348. http://dx.doi.org/10.1016/j.jeconbus.2011.05.004

Yu, L.; Wang, S.; Lai, K. 2009. An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: the case of credit scoring, *European Journal of Operational Research* 195(3): 942–959. http://dx.doi.org/10.1016/j.ejor.2007.11.025

Yurdakul, M.; İç, Y. 2004. AHP approach in the credit evaluation of the manufacturing firms in Turkey, *International Journal of Production Economics* 88(3): 269–289. http://dx.doi.org/10.1016/S0925-5273(03)00189-0

Zanakis, S.; Solomon, A.; Wishart, N.; Dublish, S. 1998. Multi-attribute decision making: a simulation comparison of select methods, *European Journal of Operational Research* 107(3): 507–529. http://dx.doi.org/10.1016/S0377-2217(97)00147-1

Zavadskas, E.; Turskis, Z. 2011. Multiple criteria decision making (MCDM) methods in economics: an overview, *Technological and Economic Development of Economy* 17(2): 397–427. http://dx.doi.org/10.3846/20294913.2011.593291

Zhou, P.; Ang, B. 2009. Comparing MCDA aggregation methods in constructing composite indicators using the Shannon-Spearman measure, *Social Indicators Research* 94(1): 83–96. http://dx.doi.org/10.1007/s11205-008-9338-0

APPENDIX 1. Evaluation criteria, descriptors and impact levels

Criteria	Descriptor	Impact levels and partial decisions		
		Impact levels		Decision
Profession (12%)	Sector: primary, secundary and tertiary	Analyzed case by case	1	
			2	
			3	Favorable
			4	
			5	
			6	
			7	
			8	Unfavorable
			9	
			10	
Employment situation (8%)	Effective	>3 years	1	
		]2; 3] years	2	
	Temporary	2 years	3	Favorable
		1 year	4	
		<1 year	5	
	Unemployed	Assumes maximum score	6	
			7	
			8	Unfavorable
			9	
			10	
Marital status (3.75%)	Single, married, divorced and widower	Analyzed case by case	1	
			2	
			3	Favorable
			4	
			5	
			6	
			7	
			8	Unfavorable
			9	
			10	

(Continued)

Criteria	Descriptor	Impact levels and partial decisions		
		Impact levels		Decision
(Continued)				
Age (2.5%)	<24	<24	1	Favorable
	25 to 35	[25;35]	2	
	36 to 50	[36;38]	3	
		[39;41]	4	
		[42;50]	5	
	>51	[51;60]	6	Unfavorable
		>60 assumes maximum score	7	
			8	
			9	
			10	
Household (3.75%)	1 element	1	1	Favorable
	2 to 4 elements	2	2	
		3	3	
		4	4	
		2 to 4 with ascendants	5	
	>5 elements	5	6	Unfavorable
		>6 ascendants	7	
			8	
		>6 ascendants and descendants	9	
			10	
Cross-selling (10%)	Customer who holds several financial products and households his/her salary in the institution	Depends on the typology of products	1	Favorable
			2	
			3	
	Customer who holds a single bank account and households his/her salary		4	
			5	
	Customer who does not have any relationship with the institution	Assumes maximum score	6	Unfavorable
			7	
			8	
			9	
			10	
Deposit portfolio (2.5%)	> 100.000 €		1	Favorable
	50.001 to 100.000 €		2	
	25.000 to 50.000 €		3	
	2.500 to 24.999 €	]12.500; 25.000[	4	
		[2.500; 12.500]	5	
	<2.500 €	Assumes maximum score	6	Unfavorable
			7	
			8	
			9	
			10	

(Continued)

Criteria	Descriptor	Impact levels and partial decisions		
		Impact levels		Decision
(Continued)				
Average balances (2.5%)	>1.000 €	>2.500	1	Favorable
		]1.000; 2.500]	2	
	0 to 1000 €	]500; 1.000]	3	
		]250; 500]	4	
		[0; 250]	5	
	<0 € (there is salary household)	]−500; 0[	6	Unfavorable
		]−1.500; −500]	7	
		]−2.000; −1.500]	8	
		[−3.000; −2.000]	9	
		<−3.000	10	
Rate of effort (15%)	0 to 100%	[0; 30]	1	Favorable
		]30; 60]	2	
		]60; 90]	3	
		]90; 100]	4	
			5	
	>100 assumes maximum score		6	Unfavorable
			7	
			8	
			9	
			10	
Responsabities in BP (15%)	Without responsibilities (age indexed)		1	Favorable
	With responsibilities but without incidents (rate of effort indexed)	Equivalent to 0% (bond)	2	
		Equivalent to 10%	3	
		Equivalent to 20%	4	
		>30%	5	
	With responsibilities and incidents	Implies automatic rejection	6	Unfavorable
			7	
			8	
			9	
			10	
LTV (15%)	<90%	[0; 20[	1	Favorable
		[20; 40[	2	
		[40; 60[	3	
		[60; 80[	4	
		[80; 90[	5	
	≥90%	[90; 100]	6	Unfavorable
		>100	7	
			8	
			9	
			10	

Criteria	Descriptor	Impact levels and partial decisions	
		Impact levels	Decision
(Continued)			
Existence of guarantors (5%)	Yes	1	Favorable
		2	
		3 Analyzed case by case	
		4	
		5	
	No	6	Unfavorable
		7	
		8 Analyzed case by case	
		9	
		10	
Economic situation of the country (2.5%)	Stable	1	Favorable
		2 Depends on the conjuncture	
		3	
		4 Stable with difficulties	
		5	
	Unstable	6	Unfavorable
		7	
		8 Assumes maximum score	
		9	
		10	
Political situation of the country (2.5%)	Stable	1	Favorable
		2 Depends on the conjuncture	
		3	
		4 Stable with difficulties	
		5	
	Unstable	6	Unfavorable
		7	
		8 Assumes maximum score	
		9	
		10	

Source: Administrative information.

APPENDIX 2. Delphi survey – round 1

Obs.: The present survey is composed of a single table and respective items. To ensure the anonymity of institutions and individuals involved, all statements provided, and their statistical treatment, will be fully confidential.

In terms of mortgage loan risk evaluation, what is the degree of importance (i.e. weight) that you give to each one of the following criteria? [From 1 to 100 (1 = Very minor importance, 50 = Moderate importance and 100 = Extreme importance), mark your preference for each criterion. Please, note that the sum of the weights must be 100%].

Criteria	Degree of importance %
Profession	
Employment status	
Marital status	
Age	
Household	
Cross-selling	
Deposit portfolio	
Average balance	
Rate of effort	
Responsibilities in BP	
LTV	
Existence of guarantors	
Economic situation of the country	
Political situation of the country	
Total:	100%

Thank You!

APPENDIX 3. Delphi survey – round 2

Obs.: The present survey is composed of a single table and respective items. To ensure the anonymity of institutions and individuals involved, all statements provided, and their statistical treatment, will be fully confidential.

In terms of mortgage loan risk evaluation, what is the degree of importance (i.e. weight) that you give to each one of the following criteria? [From 1 to 100 (1 = Very minor importance, 50 = Moderate importance and 100 = Extreme importance), mark your preference for each criterion. Please, note that the sum of the weights must be 100%].

Criteria	Degree of importance %			
	Round 1			Round 2
	Mean	Median	Standard deviation	
Profession	5.3	5.5	2.1679	
Employment status	8.6	8	6.3482	
Marital status	1.8	2	1.4832	
Age	2.3	3	1.5652	
Household	6.6	5	4.9295	
Cross-selling	3.8	3	3.3466	
Deposit portfolio	6.3	8	3.3466	
Average balance	9.6	7	4.9800	
Rate of effort	17.2	16	2.5884	
Responsibilities in BP	11.9	10.5	5.7706	
LTV	8	8	4.4159	
Existence of guarantors	7.2	4	6.4576	
Economic situation of the country	7.7	7	4.9193	
Political situation of the country	3.7	4	1.9235	
Total:	–	–	–	100%

Thank You!