Applying Chaid to Identify the Accounting-Financial Characteristics of the Most Profitable Real Estate Companies in Spain

Aplicación del Chaid para identificar las características económico-financieras de las empresas inmobiliarias más rentables en españa

Salvador Rayo[1]

Antonio M. Cortes[2]

Abstract

The aim of this study is the determination, from an empirical perspective, of the accounting and financial features which could condition financial profitability of real estate companies, to identify the performances that guarantee its permanency in the current marketplace, characterized by the world economic crisis, specially in Spain, whose housing sector represents an important contributor to the economic growth. Although at a theoretical level the DuPont Model establishes the relationships between a group of accounting ratios and financial profitability. This paper uses a sample of 5,484 Spanish real estate companies to quantify these relationships and to extract the most relevant ones and to obtain the patterns of the most profitable companies. We use ROE to measure profitability and we analyze various independent variables about solvency, liquidity, activity, turnover, financial equilibrium and investment structure. The main contribution is of methodological nature, as we have applied statistics tools that do not require initial hypotheses on the distribution of the variables, by using a data mining technique of classification and regression tree based on rule induction algorithms known as CHAID. The study provides quantitatively success profiles by means of a set of rules describing the patterns of the most profitable companies.

Keywords: CHAID; financial profitability; classification trees; accounting ratios; Spain.

Resumen

El objetivo de este estudio es la determinación, desde una perspectiva empírica, de las características económico-financieras que podrían condicionar la rentabilidad financiera de las empresas inmobiliarias, para identificar las actuaciones que garanticen su permanencia en el entorno actual, caracterizado por la crisis económica mundial, y especialmente, en España, cuyo sector inmobiliario supone una importante contribución al crecimiento económico. Aunque a nivel teórico, el Modelo DuPont establece las relaciones entre un conjunto de ratios contables y la rentabilidad financiera. Este trabajo usa una muestra de 5,484 empresas inmobiliarias españolas para cuantificar esas relaciones y para extraer las más relevantes, con el propósito de obtener los patrones de las empresas más rentables. Se utiliza el ROE para medir la rentabilidad financiera, y se analizan un conjunto de variables independientes relativas a solvencia, liquidez, actividad, rotaciones, equilibrio financiero y estructura económica. La principal contribución es de índole metodológica, ya que se aplican herramientas estadísticas que no requieren hipótesis iniciales sobre la distribución de las variables, aplicando una técnica de minería de datos de árboles de clasificación y regresión basada en algoritmos de inducción de reglas conocida como CHAID. El estudio ofrece cuantitativamente perfiles de éxito definidos con un conjunto de reglas que indican los patrones de las empresas más rentables.

Palabras clave: CHAID; rentabilidad financiera, árboles de clasificación, ratios contables, España.

1. Associate Professor. Department of Accounting and Finance, University of Granada (Spain) <srayo@ugr.es>.
2. Assistant Professor. Department of Accounting and Finance. University of Granada (Spain) <amcortes@ugr.es>.

INTRODUCTION

In the current marketplace, which is characterized by the financial crisis in the developed world, the consequences in Spain increased due to the crisis in the housing sector that had been an important contributor to Spain's economic growth. However, at present, it is yielding lower rates of profitability. The GDP in Spain is expected to contract by 3.3% in 2009 and 0.6% in 2010, down from 1.2% growth in 2008. A modest recovery will only begin during the second half of 2010, although there is a possibility that this will be delayed. This economic contraction has an important influence in the housing sector, illustrated for example with the last information published by the Bank of Spain that points out that the real estate assets of Spanish banks and saving banks were rising, at the end of March 2009, up to 20.541 million Euros, 2% more than the previous month and 10% more than one year before. The aim of this study is to determine and evaluate, from an empirical perspective, the accounting and financial features that could condition the financial profitability of companies in the real estate sector, identifying the performances that guarantee their permanency.

At a theoretical level, the DuPont Model establishes the relationships between financial profitability and a group of different variables and accounting ratios, such as asset turnover, sales margin or financial leverage. Firstly, the objective of this research is to perform an empirical contrast of this model by analyzing the relationships between the profitability and the accounting ratios, and extracting the most relevant explanatory variables of the profitability. Secondly, the paper aims at quantifying those relationships and their explanatory variables with the purpose of obtaining the patterns or profiles — that is to say, the combinations of economic-accounting features— of the most profitable companies in the housing sector.

The sample includes 5,484 Spanish real estate companies. Return on Equity (ROE) is used to measure the profitability as dependent variable. As explanatory variables the study uses various independent variables related to activity, turnover, financial equilibrium and investment structure, solvency and liquidity, most of them defined in the DuPont Model.

This work begins with a review of the main empirical studies, which have analyzed the relationships between financial profitability and different accounting ratios. Next, we outline our methodological proposal to achieve the aims. To that effect, we illustrate the DuPont Model that is used as a reference, describing the sample and the variables, and finally explaining the analysis technique that is applied. Subsequently, the main results of the analysis are exhibited: in the first place by means of a descriptive and exploratory analysis, and later from an explanatory point of view. Finally, the paper illustrates the most relevant discussions on this research.

REVIEW OF EMPIRICAL EVIDENCE

The importance of the profitability as an essential factor for the long-term survival of the companies has motivated the appearance of a high number of empirical works to evaluate the profitability of the Spanish companies, particularly real estate firms, fundamentally from a descriptive point of view. The review of the empiric literature shows the existence of two research subjects: one with a descriptive character, the other with an explanatory nature.

In papers of research subjects of descriptive character, we can distinguish two groups: (a) those referring to the whole of the Spanish firms, and (b) those analyzing particular branches of the Spanish economy or related to a specific geographic area. Within the first group, the following works stand up: Maroto (1993; 1998), Rodríguez (1989), Bueno et al. (1990), Huergo (1992), Lucas & González (1993), Sánchez (1994), and Gonzalez & Correa (1999). Also, at an institutional level, several organizations issue reports, such as the Research Service of the Mayor Council of Chambers of Commerce of Spain. The council periodically publishes reports about the situation of Spanish companies, such as the study about the profitability of the Spanish firms during the period 2000-2004 (Lizcano, 2004) or the financial report for the year 2006 (www.camaras.org). In the same

descriptive vein, albeit by sectors, some works have studied the profitability of Spanish companies, such as firms in the automobile industry (Rodríguez, 2002). Specifically to the real estate sector, many associations of realtor firms, institutions[1] and banks publish annual reports on the evolution of the sector and its perspectives. Also, many authors, such as Bermudez (2008) and Ferruz (2007), have studied the situation and the main characteristics of the housing sector, with analysis of strengths and weaknesses. However, in general, all these descriptive studies use a traditional methodology, focused fundamentally on the analysis univariable of ratios, applying it on account information too much aggregated, which is obtained from the database of the Statement Central of the Spanish Central Bank. This information introduces problems of representativeness of the Spanish entrepreneurial environment, made explicit by the prevalence of big companies; this approach runs into trouble with the analyses and conclusions from those studies.

With regard to papers with explanatory nature, we have found various documents that make use of statistical techniques of multivariate analysis from an empirical perspective: Fariñas & Rodriguez (1986), Aguilar (1989), Antón, Cuadrado & Rodriguez (1990), Fernández & García (1991), and González (1997). The analysis of these works suggests that size has been the variable which has received a bigger attention from the researchers of profitability. Nevertheless, it is not possible to establish a clear relationship between both variables, since the conclusions from those studies are heterogeneous. Thus, some papers indicate a positive relationship between size and profitability (Galvé & Salas, 1993; González, 1997). However, other authors show the existence of a negative relationship, confirming the results obtained through the traditional methodology, as proved by Huergo (1992), Fariñas (1992), Maroto (1993, 1998), Salas (1994) and Illueca (1996), who point out that small and medium companies get higher financial and economic profitability. On the other hand, the studies of Suárez (1977), Rodriguez

(1989) and Galán (1997) suggest that size is not a significant variable to explain the profitability of companies.

The principal limitations of these explanatory studies are fundamentally consequence of three aspects: (a) the difficulties in obtaining a significant sample of companies that brings consistency to the results, mainly in the real estate sector; (b) the biggest complexity that implies the application of the multivariate statistic techniques and the interpretation of their results; (c) the absence of normality in the distributions of the ratios, which limits the validity of some statistic techniques and reduces the explanatory capacity. This research tries to contribute, by means of the empiric analysis, to improve the knowledge of the economic and financial characteristics that determine the profitability of the Spanish companies in the housing sector.

Our main contributions are of methodological nature. In the first place, our study focuses on the real estate sector using a sample of companies. We try to get over the problems other works show by using disaggregated account information (for each firm), with an appropriate representativeness by size, and by jointly analyzing a sufficient number of variables and ratios that can explain financial profitability. In the second place, the analysis applies statistics tools which do not require initial hypothesis on the distribution of the variables (showing greater adjustment to the characteristics of the account information). We have applied data mining techniques of classification and regression tree based on rule induction algorithms such as CHAID.

METHODOLOGICAL PROPOSAL

The following section outlines the methodological scheme we propose to achieve the aims. In this section, we show the DuPont Model, which is used as a reference to verify them; we describe the sample and the variables, and finally we explain the analysis techniques that are applied.

---

[1] Some of these institutions include Asprima, Tinsa, KPMG, whose reports are available at their websites.

Theoretical model: financial profitability

The study of the profitability is usually carried out at two levels: economic profitability and financial profitability; their relationship comes to be defined by the financial leverage.

1. Economic profitability (ROA = Return on Assets)

   The Economic profitability (ROA) is a measure of the capacity of the assets to generate worth with independence of how they have been financed. It is usually obtained as follows[2]:

   $$ROA = \frac{\text{Earnings Before Interest and Taxes (EBIT) or Operating Profit Loss}}{\text{Total Average Assets (TA)}}$$

   ROA may be decomposed into return on sales multiplied by asset turnover:

   $$ROA = \frac{EBIT}{TA} = \frac{EBIT}{Sales} \; x \; \frac{Sales}{TA} = \text{Return on Sales x Asset Turnover}$$

   Return on sales represents the profit obtained for each sold monetary unit, that is, the profitability of the sales. The components of return on sales can be analyze through the decomposition into costs of goods sold, depreciation and cost of employees.

   Asset turnover measures a firm's efficiency at using its assets in generating sales. The amount of sales is generated for every monetary unit's worth of assets. It is calculated by dividing sales by total assets.

b) Financial profitability (ROE = Return on Equity)

   Financial profitability (ROE) is a measure of a corporation's profitability that reveals how much profit a company generates with the money shareholders have invested. It is defined as:

   $$ROE = \frac{\text{Net Profit (NP)}}{\text{Average Shareholders Funds or Equity (E)}}$$

At a theoretical level, the DuPont Model was a method of performance measurement that was started by the DuPont Corporation in the 1920s. The analysis breaks financial profitability among various factors that represent the explanatory variables to contrast in this research:

$$ROA = \frac{EBIT}{TA} \; x \; \frac{Sales}{TA} \; x \; \frac{\text{Profit and Loss (P/L) before Taxes}}{EBIT} \; x$$

$$\frac{TA}{E} \; x \; \frac{NP}{\text{P/L before Taxes}}$$

As a result, ROE can be disaggregated into the following components:

- Economic Profitability (ROA), determined by dividing return on sales by assets turnover:

  $$ROA = \frac{EBIT}{Sales} \quad x \quad \frac{Sales}{TA}$$

- Financial Leverage, as a product of a leverage indicator (TA/E) and another one about the cost of debt:

  $$\text{Financial Leverage} \; = \; \frac{TA}{E} \; x \; \frac{\text{P/L before Taxes}}{EBIT}$$

- Taxes Effect, determined by dividing net profit by the profit and loss before taxes[3]:

  $$\text{Taxes Effect } ct = \frac{NP}{\text{P/L before Taxes}} =$$

  $$\frac{\text{P/L before Taxes} - \text{Taxes}}{\text{P/L before Taxes}} = \quad 1 - t$$

Alternatively, an expanded decomposition of financial profitability is shown as follows, the equation that is usually known as the Financial Leverage Equation[4]:

---

[2]   To calculate ROA, it is only considered the operating profit and loss, as the influence of the extraordinary profit and loss is separated, which will be added later at the formulation of financial profitability.

[3]   «*Taxes*» means the corporation *taxes* and «t» denotes the *tax rate*.

[4]   «D» represents the nominal debt (current liabilities plus long-term liabilities), and «i» refers to the average cost of the debt (Interest/Debt).

$$ROA \;\; = \;\; \frac{Net\ Profit}{Equity} = \left[ROA \;\; + \;\; (ROA - i) \;\; x \;\; \frac{D}{E} \right] \;\; x$$

$$(1-t) \;\; + \;\; \frac{Extr.\ Profit}{E} \;\; (1-t)$$

This formula[5] allows completing the explanatory variables extracted at the first decomposition.

Variables

The DuPont Model shows the main theoretical variables that affect financial profitability. Additionally, other ratios and indicators that have traditionally been studied by the entrepreneurial analysis are added to those variables of the DuPont Model, completing the group of independent variables to be contrasted empirically in this research.

As a result, Return on Equity (ROE) is used to measure the financial profitability (dependent variable). As explanatory variables this study uses various independent variables related to different aspects of the entrepreneurial environment: asset structure, liability structure, financial balance, profitability and productivity, turnover and activity, and growth. The definitions of the variables used are described on Tables 1 and 2.

The dependent variable (ROE) has been categorized into quartiles (low, low medium, high medium, high), because our main interest is to focus on the first and fourth quartiles. They represent the best and the worst profitability situations (success and failure profiles). This categorization into quartiles is applied by many authors in studies that apply the CHAID technique, such as Santín (2006), Dills (2005) and Gonzalez, Correa and Acosta. (2002).

Sample characteristics

The data for this research is based entirely on the SABI database (Iberian Balance Sheet Analysis System), which is offered jointly by INFORMA D&B and Bureau Van Dijk. This database records the financial statements (balance sheet and profit and loss account) of companies in Spain and Portugal, provided by the Trade Registers of every geographical area.

INFORMA D&B was the first European company to supply commercial and financial information over the internet (15/9/96), and it was also the first Spanish commercial and financial information database to achieve AENOR, at present updated to the standard ISO 9001. In particular, the SABI database shows general information and annual accounts for more than 1,2 million of Spanish companies as well as more than 350,000 Portuguese firms.

The website of INFORMA D&B include the approximate price list for an annual subscription to the database, ranging from € 6,000 to more than € 9,000, which depends on the number of companies that are available, the update frequency, the geographical scope and the type of access (DVD or network). Although INFORMA owns this information, there is no problem in using it for research or academic purposes.

SABI allows making multi-criteria searches defining the variables that are required in each case; configures lists of companies, establishing personal formats, and creating particular ratios and add new financial information in any given report. Through these tools, the information obtained was debugged to avoid errors and to allow the statistical analysis. The sample comprised 147,299 companies, and the housing sector included 5,484 firms. Companies with negative shareholders funds, bankruptcy or negative net assets, or incomplete information in some of the variables defined, such as not disaggregated data, were removed from the sample. Additionally, using the Clementine software (SPSS Inc.), companies with outliers or extreme values (more than 3 or 5 times the average, respectively) were also

---

[5] Because of Extraordinary P/L was not considered into the formulation of ROA, and even though ROE includes net profit, now we must add it (after tax) into the equation of ROE.

Table 1.

Assets, Liabilities and Financial Balance

ASSET STRUCTURE	
ASSETS STRUCTURE:	
Total Assets	Fixed Assets + Current Assets
Sales	
Number of employees	
FIXED STRUCTURE:	
Fixed Asset Ratio (%)	Fixed Assets / Total Assets
Tangible Asset Ratio (%)	Tangible Assets / Total Assets
Intangible Asset Ratio (%)	Intangible Assets / Total Assets
Financial Asset Ratio (%)	% Financial Assets / Total Assets
CURRENT STRUCTURE:	
Current Asset Ratio (%)	Current Assets / Total Assets
Stock Ratio (%)	Stocks / Total Assets
Debtor Ratio (%)	Debtors / Total Assets
Cash Ratio (%)	Cash / Total Assets
LIABILITY STRUCTURE	
COST OF DEBT:	
Interes Rate (i) (%)	Interest / (Current Liabilities + Long-Term Liabilities)
Interest / Sales (%)	Interest / Sales
LEVERAGE:	
Debt Ratio (%)	(Current Liabilities + Long-Term Liabilities) / (Liablities + Equity)
LIABILITY STRUCTURE:	
% Current Liability Ratio (%)	Current Liablties / (Liabilities + Equity)
% Long-Term Liability Ratio (%)	% Long-Term Liabilities / (Liabilities + Equity)
FINANCIAL BALANCE	
WORKING CAPITAL:	
Working capital	Current Assets – Current Liabilities
Working Capital Ratio (%)	Working Capital / Total Assets
SOLVENCY:	
Solvency Ratio (%)	Net Profit + Depreciation / (Long-Term Liabilities + Current Liabilities)
Short-Term Solvency Ratio (%)	Net Profit + Depreciation / Current Liabilities
Asset Coverage Ratio (%)	(Tangible+Intangible+Financial+Current Assets) / (Long-Term + Current Liabilities)
LIQUIDITY:	
Liquidity Ratio (%)	Current Assets / Current Liabilities
Short-Term Liquidity Ratio (%)	Cash / Current Liabilities

Table 2.

Profitability, Activity and Turnover Ratios

PROFITABILITY AND PRODUCTIVITY	
PROFITABILITY:	
Financial Profitability (ROE) (%)	Net Profit / Equity
Economic Profitability (ROA) (%)	Earnings Befote Interest and Taxes (EBIT) / Total Assets (TA)
ROA minus Interest Rate (%)	ROA – i
PRODUCTIVITY:	
Labor Productivity (%)	(Sales – Cost of Goods Sold – Depreciation – General Expensens) / Costs of Employee
ACTIVITY AND TURNOVER	
SALES MARGIN:	
Return on Sales (%)	Earnings Before Interest and Taxes (EBIT) / Sales
Cost of Goods Sold / Sales (%)	Cost of Goods Sold / Sales
Cost of Employee / Sales (%)	Cost of employee / Sales
Depreciation ratio (%)	Depreciation / Sales
TURNOVER:	
Asset Turnover (%)	Sales / Total Assets
Fixed Asset Turnover (%)	Sales / Fixed Assets
Current Asset Turnover (%)	Sales / Current Assets
Collection Period	(Creditors / Sales)*360
OTHERS COMPONENTS:	
Taxes / Sales (%)	Taxes / Sales
Extraordinay Profit and Loss / Equity (%)	Extraordinary Profit and Loss / Equity
Financial Revenue / Equity (%)	Financial Revenue / Equity
GROWTH	
Sales Growth Rate (%)	
Cost of Employee Variation (%)	
Fixed Asset Variatin (%)	Variation from year 2005 to year 2006 in percentage
Total Asset Variation (%)	
ROE Growth Rate (%)	

filtered[6]. Finally and aiming at capturing the recent evidence, the information is only from the year 2006 on annual basis[7], with some variables calculated by variations from 2005 to 2006.

Analysis technique: CHAID

By means of the Clementine (SPSS Inc.) software, the CHAID rule induction algorithm (Chi-squared Automatic Interaction Detector) was applied, a highly efficient statistical technique for segmentation, or tree growing that derives a tree of rules that attempts to describe distinct segments within the data in relation to the output variable (ROE). This allowed us to classify companies according to the different values of the accounting ratios and their profitability.

In fact, a great many algorithms are capable of generating rules based on decision trees, including CLS (Hunt et al., 1966), ID3 (Quinlan, 1979), CART (Breiman et al., 1984) and C4.5 (Quinlan, 1993). In the present study, we implemented the algorithm known as CHAID (Chi-squared Automatic Interaction Detector), which is simple to apply and widely used. This classification mechanism, originally proposed by Kass (1980), has been used extensively by many authors in different studies to derive a tree of rules, which helped the understanding of many phenomena (Santín, 2006; Galguera, 2006; Grobler, 2002; Strambi, 1998).

As a segmentation tool, CHAID presents important benefits. First of all, the technique is not based on any specific probabilistic distribution, but solely on chi squared goodness-of-fit tests, from contingency tables. These tests, given an acceptable sample size, almost always function well. In the second place, it makes it possible to determine a variable to be maximized. This is indeed desirable, and not always achievable with other segmentation techniques. Moreover, classification by segments is always straightforward to interpret, as its results provide intuitive rules that are readily understood by non experts – which, for example, is not the case, with Cluster Analysis. And finally, this technique ensures that the segments always have statistical meaning; they are all different, and are the best possible, given the data provided. Accordingly, the classifications made using the rules found are mutually exclusive, and so the decision tree identifies a single response based on a calculation of the probabilities of belonging to a certain class. Last of all, CHAID, unlike other algorithms such us CART (Breiman et al., 1984), is capable of constructing non-binary algorithms; for example, it can present more than two branches, or data divisions, according to the categories to be explained, for each node.

Using the significance of a statistical test as a criterion, CHAID algorithm evaluates all of the values of every potential explanatory variable. Let us examine in three steps the methodological process to be followed when applying the technique (a complete description is showed by Kass 1980; Biggs, 1991; and Goodman, 1979):

1. Binning of continuous explanatory variables: Continuous explanatory variables are discretized or binned into a set of ordinal categories. It can be done through various machine learning algorithms for building decision trees or decision rules, in particular by the CHAID algorithm, which we apply[8].

2. Merging categories for explanatory variables: CHAID algorithm merges those values that are judged to be statistically homogeneous (similar) with respect to the dependent variable and maintains all other values that are heterogeneous (dissimilar). All the explanatory variables are merged to combine categories that are not statistically different with respect to the dependent variable, and each final category of an explanatory variable X

---

[6] To test the possible impact of the number of firms on the results, we are working on a new sample extended in time, and our preliminary analysis suggests similar results, with only small variations in the quantitative levels of the rules obtained.

[7] It should be noted that SABI offers data on an annual and monthly basis.

[8] We are aware that there are several methods for binning into a set of categories, for example, the one proposed by Berka (1998), which will be studied in future research, to compare results with those described in this paper.

represents a leaf node if that variable is used to split the node. For each explanatory variable X, the algorithm finds the pair of categories of X that is least significantly different (indicated by the largest p-value) with respect to the dependent variable Y. The method used to calculate the p-value is the chi-squared test:

$$N^2 = \sum_{j=1}^{J} \sum_{i=1}^{I} \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$$

where $n_{ij} = \Sigma_n f_n I(x_n = i \wedge y_n = j)$ is the observed cell frequency and $\hat{m}_{ij}$ is the expected estimated cell frequency for cell $(x_n = i, y_n = j)$ under the null hypothesis of Independence. The corresponding p value is given by $p = Pr(N_d^2 > X^2)$, where N_d^2 follows a chi-squared distribution with degrees of freedom $d = (J - 1)(I - 1)$. The frequency associated with case n is noted by f_n.

Then, it merges into a compound category with the pair that gives the largest p-value, and calculates the p-value based on the new set of categories of X. This represents one set of categories for X. The process is repeated until only two categories remain. Then, the sets of categories of X generated during each step of the merge sequence are compared, to find the one for which the p-value in the previous step is the smallest. That set is the one of merged categories for X to be used in determining the split at the current node.

3. Splitting nodes: Each variable is evaluated for its association with the dependent variable, based on the adjusted p-value of the statistical test, and the algorithm selects the best predictor to form the first branch in the decision tree, that is, the explanatory variable with the largest association with the dependent variable (the one for which the chi-squared test has the smallest p-value). If this value is less than or equal to the α split (the split threshold), then that variable is used as the split variable for the current node. Each of the merged categories of the

split variable defines a child node of the split. After the split is applied to the current node, the child nodes are examined to see if they warrant splitting by applying the merge/split process to each in turn. This process continues recursively until the tree is fully grown and no further splits can be made.

The main results of the model are described in the following items:

Support

The support for a scored record is the weighted number of records in the data in the scored record's assigned terminal node (t), i.e., the number of records of each rule. It can be defined $N_{w,j}(t) = \Sigma_{i \in t} w_i f_j(i)$ as the weighted number of records in node t with category j, and $N_{w,j}(t) = \Sigma_{i \in t} w_i f_j(i)$ as the weighted number of records in category j (any node).

Response (or confidence):

The confidence for a scored record is the proportion of weighted records in the data in the scored record's assigned terminal node (t) that belong to a selected category j, modified by the Laplace correction (Margineantu, 2001), with k being the number of categories. It is computed as $(N_{f,t}(t) + 1)/ (N_f(t) + k)$. Thus, the level of confidence (%) of each rule (terminal node) shows the proportion of records of each rule that belong to a selected category j; and, the level of confidence of a set of rules can also be defined as the proportion of records of this rule set belonging to a given category j.

Index:

The index of each of the rules obtained for a given category j is obtained as the ratio between the level of confidence for each rule (terminal node) and the level of confidence of the category j in the total sample (i.e., 25%, as the sample is divided into quartiles). Therefore, it is obtained by dividing the proportion of records that present category j in each terminal node (rule) into the proportion of records presenting category j in the total sample (25%). Thus, it represents the increased probability of belonging to the selected category j

that contains the records presenting the characteristics defined for each rule. Therefore, by accumulation, the index of a set of rules can be obtained as the ratio between the proportion of records presenting category j in this rule set and the corresponding proportion to be found within the total sample (25%).

Gain:

The gain for each terminal node (rule) can be defined as the number of records in a selected category j, in absolute terms. For a set of rules or terminal nodes, and in percentage terms, the gain summary provides descriptive statistics for the terminal nodes of a tree, and shows the weighted percentage of records in a selected category j, noted as $g(t,j) = \sum_{t \in t} f_i x_i(j) / \sum_{t \in t} f_i$ where $x_i(j) = 1$ if record x_i is in category j, and 0 otherwise.

Risk:

It represents the risk of error in predicted values for specific nodes of the tree and for the tree as a whole. The risk estimate of a node (i.e. rule) t is computed as $r(t) = (1/N_f) \sum_j N_{f,j}(t)$, where $N_{f,j}(t)$ is the sum of the frequency weights for records in node t in category j, and N_f is the sum of frequency weights for all records in the sample. Anyway, the risk estimate $R(T)$ for the tree (T) is calculated by taking the sum of the risk estimates $r(t)$ for the terminal nodes, computed as $R(T) = \sum_{t \in T} r(t)$, where T is the set of terminal nodes in the tree.

RESULTS

Descriptive analysis

Table 3 shows means for each one of the variables defined for the year 2006, making a comparison between the real estate sector and the all other activities. In the first place, the analysis of the profitability exhibits similar figures for housing companies, with ROE near 12% and ROA between 5%-6%. Focusing on asset structure, real estate companies were characterized by a higher fixed asset ratio (39.14%) than the mean of all the other activities (33.74%), and, consequently, it shows a lower current asset ratio (60.85% instead of

66.25%), mainly due to the fact that the debtor and cash ratios were much lower (around 21% less). However, it is important to emphasize that the stock ratio was much higher (near 14% more).

Concerning liability structure, the interest rate was similar for the realtor companies and the other ones. Nonetheless, the cost of debt per unit of sales was much higher for the real estate sector (4.62% vs. 1.58%), even taking into account that debt ratio was similar in both groups. The debt structure was different for real estate companies, with a much higher presence of long-term liabilities, resulting in a higher working capital ratio (27.75% instead of 17.05%). As a consequence of that structure of assets and liabilities (with similar leverage ratio but higher fixed assets and long-term liabilities), solvency and liquidity ratios were much better for the housing sector, e.g. the asset coverage ratio or the liquidity ratio were almost two times the ones of the whole sectors.

On the other hand, return on sales was better in this sector (13.11% vs. 4.75%). This was mainly because of the higher productivity of labor (exhibited also by a lower cost of employee ratio) and the higher growth of sales. However, the assets turnover was lower due to the fact that the stock ratio was much higher and, then, the current asset turnover was much lower. This explains that the figures for ROA were similar to those for all the activities, as mentioned above. The growth rate of fixed assets and total assets were also higher for housing companies, which demonstrates the superior dynamism of the housing sector in comparison to the whole economy.

Exploratory analysis

The exploratory analysis of correlations between each of the explanatory variables and ROE is shown on Table 3, including the mean values for each category of the variable ROE and the F-Test statistics of independence. This table allows analyzing the differences between the most and least profitable companies in order to find the accounting and financial features that would explain the profitability of the real estate firms.

Table 3.
Descriptive and Exploratory Analysis of Real Estate Sector [9]

VARIABLES	DESCRIPTIVE ANALYSIS (MEANS)		EXPLORATORY ANALYSIS (MEANS ROE)				CORRELATIONS	
ASSET STRUCTURE	TOTAL SECTORS	REAL ESTATE	LOW	LOW MEDIUM	HIGH MEDIUM	HIGH	F-Test	p-level
Total Asset	3,655.35	9,712.03	8,525.43	10,161.63	9,970.94	10,190.11	0.57	0.368
Sales	3,659.63	2,533.30	920.65	1,951.50	3,093.04	4,168.04	35.96	1
Number of employees	19.62	7.18	5.06	7.22	8.13	8.32	9.08	1
Fixed Asset Ratio (%)	33.74	39.14	55.93	45.97	30.31	24.34	278.65	1
Tangible Asset Ratio (%)	25.64	30.34	44.55	36.96	22.47	17.37	235.44	1
Intangible Asset Ratio (%)	5.03	2.3	2.59	2.02	2.25	2.36	1.20	0.692
Financial Asset Ratio (%)	2.78	6.2	8.49	6.75	5.24	4.32	24.11	1
Current Asset Ratio (%)	66.25	60.85	44.07	54.00	69.68	75.67	278.60	1
Stock Ratio (%)	20.57	34.24	22.75	28.83	41.09	44.31	128.63	1
Debtor Ratio (%)	30.38	13.55	10.42	12.51	14.68	16.59	30.96	1
Cash Ratio (%)	12.15	8.17	6.15	7.05	9.45	10.02	25.72	1
LIABILITY STRUCTURE								
Interest Rate (i) (%)	2.57	2.27	2.30	2.26	2.21	2.30	0.32	0.186
Interest / Sales (%)	1.58	4.62	6.03	5.02	4.06	3.35	92.12	1
Debt Ratio (%)	65.9	63.88	51.40	55.97	70.59	76.75	269.65	1
Current Liability Ratio (%)	49.2	33.1	25.34	27.86	38.14	41.07	139.43	1
Long-Term Liability Ratio (%)	16.7	30.57	26.06	28.11	32.45	35.68	37.66	1
FINANCIAL BALANCE								
Working Capital	396.31	3,488.18	1,970.10	3,138.78	4,195.94	4,647.88	5.94	1
Working Capital Ratio (%)	17.05	27.75	18.72	26.14	31.54	34.59	57.64	1
Solvency Ratio (%)	18.52	20.18	16.43	25.12	18.17	21.02	5.76	0.999
Short-Term Solvency Ratio (%)	30.08	53.33	50.49	75.42	43.15	44.25	15.37	1
Asset Coverage Ratio (%)	223.26	433.77	871.22	497.85	207.48	158.54	121.43	1
Liquidity Ratio (%)	214.62	519.43	777.26	613.05	362.26	325.17	36.14	1
Short-Term Liquidity Ratio (%)	50.46	92.96	142.93	118.51	64.43	45.98	28.88	1

(continues) ⟶

[9] The table shows a descriptive analysis of the means variables of real estate sector compared with the total sectors, and the exploratory analysis of real estate sector by ROE of the companies. Finally, it exhibits the correlations of each explanatory variable with ROE.

→ Table 3

PROFITABILITY / PRODUCTIVITY								
Financial Profilatibity (ROE) (%)	11.64	11.81	-4.40	4.86	14.28	32.50	8.238.77	1
Economic Profitability (ROA) (%)	5.9	5.28	-0.12	3.40	6.60	11.26	910.95	1
ROA minus interest Rate (%)	3.33	3.02	-2.43	1.14	4.39	8.96	847.09	1
Labor Productivity (%)	169.54	533.37	295.02	417.15	604.40	816.91	50.35	1
ACTIVITY AND TURNOVER								
Return on Sales (%)	4.75	13.11	2.85	16.31	16.32	16.96	275.80	1
Cost of Good Sold Ratio (%)	51.56	48.02	49.12	40.08	49.48	53.39	1.04	0.627
Cost of Employee Ratio (%)	26.48	21.41	31.01	20.66	18.73	15.22	127.31	1
Depreciation Ratio (%)	3.86	7.43	15.62	7.93	3.77	2.39	318.68	1
Asset Turnover (%)	167.06	59.96	33.24	45.83	69.60	91.16	163.20	1
Fixed Asset Turnover (%)	4,022.97	5,724.39	3,262.81	2,987.60	4,762.01	11,885.15	12.55	1
Current Asset Turnover (%)	354.37	186.59	175.05	188.72	210.86	171.11	0.53	0.341
Collection period (%)	89.1	142.87	205.00	153.15	117.33	96.02	112.85	1
Taxes / Sales (%)	1.18	3.33	0.61	4.17	4.12	4.42	150.98	1
Extraordinary P and L / Equity (%)	2.23	1.55	0.54	1.28	1.37	3.03	5.28	0.999
Financial Revenue / Equity (%)	1.1	1.61	1.01	1.24	1.87	2.32	21.93	1
GROWTH								
Sales Growth Rate (%)	14.91	27.28	10.38	17.73	28.18	52.81	52.23	1
Cost of Employee Variation (%)	15.41	25.43	25.37	20.86	23.37	32.11	4.92	0.998
Fixed Asset Variation (%)	23.32	35.85	25.14	27.58	35.76	54.93	17.72	1
Total Asset Variation (%)	15.99	21.83	16.11	17.74	23.05	30.41	22.95	1
ROE Growth Rate (%)	-14.18	-23.39	-89.21	-16.84	1.78	14.27	114.03	1

In the first place, the analysis of asset structure shows a negative relationship between fixed asset ratio and profitability, because of the lower fixed asset ratio the higher ROE and ROA. On the one hand, the most profitable companies reduce their fixed asset ratios by means of lower tangible asset ratios, which imply that they achieve higher turnover ratio and better levels of efficiency in the production. On the other hand, firms increase their current asset ratios not only with higher debtor and cash ratios, but with higher stock ratios, which is very important because it could suggest that housing companies own an important problem of oversupply. Summing up, this asset structure allowed them to reach higher asset turnover ratios, explaining the higher ROA, as the DuPont Model points out.

Secondly, there is a positive relationship between leverage and profitability. Moreover, it is shown that the lower cost of debt, the higher profitability. Current liability ratio represents a high percentage among the most profitable companies, which means that they are usually financed by trade providers that offer a financial product cheaper than loans with banks or long-term liabilities, explaining the lower cost of debts. The most profitable companies own bigger sales and allow them to get the best rates when negotiating the conditions with the providers. With respect to long-term liability ratio, the most profitable companies also presented higher figures, and therefore, both ratios explain the high leverage of these companies.

Last, but not least, the rates of activity show that the return on sales was higher for the most profitable companies, as it is expected at the DuPont Model, which, together with higher asset turnover, explains the better figures for ROA. The main variable that determines the high return on sales among the most profitable companies was the cost of employees; to be sure, the lower cost of employee ratio, the higher return on sales. In fact, labor productivity was also higher for the most profitable companies. Additionally, the decline on depreciation ratio for these companies (caused by the low percentage of fixed assets) contributed to improve return on sales and, therefore, ROA; besides, the growth rate of sales was also higher for those companies.

Consequently, as Financial Leverage Equation predicts, when ROA exceeds interest rate, leverage contributes to increase ROE.

$$ROE = \frac{Net\ Profit}{Equity} = [ROA + (ROA - i) * D / E] *$$

$$(1 - t) + \frac{Extr.\ Profit}{E} (1 - t)$$

As a conclusion, the most profitable companies are characterized by:

- Higher ROA explained by:
  - Lower fixed asset ratios, increasing fixed asset turnovers.
  - Higher return on sales due to savings in labor costs and higher productivity, lower rates of depreciation because of getting lower fixed asset ratios, and lower cost of debts. Additionally, the growth of sales explains the higher ROA.

- Higher leverage ratio, mainly because current liability ratio is higher in these companies as they are usually financed by trade providers with long periods of payment.

To sum up, as it is confirmed by the F-test statistics, the main explanatory variables of the profitability of the real estate companies were the fixed asset turnover, the return on sales and the level of fixed assets and debt (included into the asset coverage ratio), all of them provided by an expansive economic cycle characterized by a high level of sales and a high leverage of the companies. It verifies what it is expected in the DuPont Model. Undoubtedly, the analysis of size indicates that the level of assets was not specifically an important explanatory variable of the profitability, as there were not significant differences in the figures of total assets between companies with low and high ROA, confirming some previous studies (Rodríguez, 1989; Galán, 1997). However, the figure of sales, as noted above, was the variable which allowed companies to reach those higher ROA and ROE. It is also important to point out that the main weakness of the real estate companies lies in

their high stocks and, therefore, lower fixed asset which, together with the high debt, brings the real estate sector into a vulnerable status to face the crisis because of the worse solvency and liquidity ratios.

Explanatory analysis: predictive analysis with CHAID

1. Results and rules obtained with the model. Success and failure profiles.

In the previous section, DuPont Model has been contrasted empirically by analyzing the relationships between the profitability and the accounting ratios, and extracting the most relevant explanatory variables of the profitability. This section aims at quantifying those relationships and their explanatory variables with the purpose of obtaining the profiles, that is, the combinations of accounting ratios, of the most profitable companies.

With CHAID modelling, the sample is segmented taking into account the different levels of the explanatory variables, building a classification tree finishing in a set of terminal nodes —with routes from the origin node (the whole sample) to each terminal node (t)—, which constitute the profiles or rules for each of the categories defined in the variable to be explained (ROE). Thus, there are so many rules as terminal nodes. However, the tree segmentation results in a large number of rules or company profiles, so that for the purposes of the present study only the most important have been selected[10]. Therefore, the rules obtained for ROE=high medium and ROE=low medium has been omitted, since the most interesting and useful to study are the extreme quartiles, indicatives of success and failure profiles. Moreover, only the most important rules for the quartiles studied

---

[10] It should be taken into account that the main aim of this study is not mainly to classify and to predict ROE of every company, but to offer profiles and recommendations to real estate companies regarding the main variables and accounting ratios that may influence ROE, as well as to ascertain the most suitable values for them.

have been selected, which are those presenting the highest classificatory and predictive capacities in terms of the level of confidence.

Accordingly, we filter out the rules obtained for the categories ROE=high and ROE=low, and after ordering them by level of confidence, the most important rules in each category are selected. The final result, thus, is that we have the rules for the highest sampling decile in each category, representing around 600 companies for each category studied.

The rules selected, in both cases, are illustrated on Table 4, which shows in brackets the corresponding support (number of firms of the sample with the profile detailed in the rule) and confidence (percentage of these companies which belong to the category studied). Special mention deserves the rules (profiles) for the most profitable companies, those ones with ROE=high. These rules indicate the figures within which these variables should be situated in order to ensure good levels of ROE, with a high level of probability. We can stand out several profiles of companies (rules) which are likely to obtain ROE=high, because they own higher confidence percentage than the whole sample (25%).

As an example, with a support of 109 firms, rule 14 shows that 84.4% of the companies with return on sales over 7.88% and asset turnover over 144.22% obtained ROE=high. Also, rule 12 exhibits that when the return on sales is higher than 14.8%, the asset turnover ratio does not need to be so high as noted above, but even exceeding 90.94%, 86.7% of the 105 firms achieve high levels of ROE (upper quartile, ROE=high), exceeding 25% in three times, i.e. the percentage of firms with ROE=high in the whole sample before segmentation. On the other side, rule 1 exhibits that 69.7% of the companies with asset turnover between 16,11% and 23.42%, return on sales over 14.8% but debt ratio over 90.65%, also obtained ROE=high. Therefore, it is possible to obtain higher ROE not only with better asset turnover, but as well with lower asset turnover

Table 4.

Principal rules selected for ROE=high and ROE=low by CHAID algorithm

Rules for ROE=HIGH	
Rule 1 (66; 0.697)	if Asset Turnover Ratio > 16.11 and <= 23.42 and Return on Sales > 14.8 and Debt Ratio > 90.65
Rule 2 (57; 0.702)	if Asset Turnover Ratio > 23.42 and <= 32.31 and Return on Sales > 14.8 and <= 26.98 and Asset Coverage Ratio <= 116.6
Rule 3 (99; 0.697)	if Asset Turnover Ratio > 32.31 and <= 43.73 and Return on Sales > 10.97 and <= 26.98 and Asset Coverage Ratio <= 125.77
Rule 8 (70; 0.8)	if Asset Turnover Ratio > 61.26 and <= 90.94 and Return on Sales > 10.97 and <= 20.15 and Debt Ratio > 62.74
Rule 9 (103; 0.806)	if Asset Turnover Ratio > 61.26 and <= 90.94 and Return on Sales > 20.15
Rule 12 (105; 0.867)	if Asset Turnover Ratio > 90.94 and <= 144.22 and Return on Sales > 14.8
Rule 14 (109; 0.844)	if Asset Turnover Ratio > 144.22 and Return on Sales > 7.88
Rules for ROE=LOW	
Rule 2 (111; 0.937)	if Asset Turnover Ratio <= 6.61 and Asset Coverage Ratio > 110.03 and <= 202.79 and Solvency Ratio <= 1.32
Rule 3 (174; 0.966)	if Asset Turnover Ratio <= 6.61 and Asset Coverage Ratio > 202.79 and Return on Sales <= 14.8
Rule 5 (105; 0.895)	if Asset Turnover Ratio > 6.61 and <= 10.74 and Debt Ratio <= 79.61 and Return on Sales <= 7.88
Rule 6 (65; 0.862)	if Asset Turnover Ratio > 6.61 and <= 10.74 and Debt Ratio <= 79.61 and Return on Sales > 7.88 and <= 26.98 and Tangible Asset Ratio > 66.4
Rule 10 (87; 0.782)	if Asset Turnover Ratio > 16.11 and <= 23.42 and Return on Sales <= 2.29
Rule 16 (68; 0.75)	if Asset Turnover Ratio > 144.22 and Return on Sales <= 2.29 and ROE Growth Rate <= 91.45

ratios, providing that the debt ratio and return on sales are higher. There are multiple combinations which are represented by the other five rules.

It would be possible to analyze all the selected rules in the table in the same way, and thus to obtain a series of profiles and/or recommendations providing real estate companies with quantitative control measures for obtaining high levels of ROE. In summary, it is observed that there are two groups of profiles of high profitability: the first one, refers to companies with a moderated leverage but with good ratios of turnover and return on sales (rules 9, 12 and 14); the second one, those with worse asset turnover ratios and return on sales, though assuming higher financial risks while supporting low asset coverage ratios and high leverage (rules 1, 2, 3 and 8). In the current marketplace, which is characterized by the financial crisis, the companies in the first group will be able to deal with the crisis with more guarantees, although they will not be

free of difficulties, especially because, as exploratory analysis showed, the high asset turnover ratios mainly come from high fixed asset turnover, and it would be advisable that they were caused by higher stock turnover instead.

At the other extreme, we have the profiles of the companies with the lowest levels of ROE. In the same way, six rules for ROE=low can be extracted to identify clearly the profiles of the least profitable companies. For example, rule 2, for a sampling support of 111 companies, indicates that there is a 93.7% probability that the firms with an asset turnover ratio of less than 6.61%, an asset coverage ratio between 110.03% and 202.79%, and a solvency ratio under 1.32% will present low levels of ROE (lower quartile, ROE=low).

As a conclusion, the study of these rules shows that asset turnover, return on sales, and asset coverage ratio (also measured indirectly with the debt

ratio or the solvency ratio) are the variables which determine the profitability, as exploratory analysis and DuPont Model showed, but additionally, this explanatory analysis allows quantifying the levels of these variables to achieve the highest figures of profitability. Thus, it provides the main accounting ratios, to ascertain the most suitable values for them, which managers of companies should monitor in order to ensure good levels of ROE.

2. Goodness of the model

To illustrate the goodness of the model, the following matrix of incorrect classification shows the companies correctly and incorrectly classified with all the rules obtained by CHAID segmentation (see Table 5). The total risk R(T), that is, the sum of all the risks for the set of terminal nodes (rules), is 40.54%, and it measures the percentage of cases classified incorrectly when all the rules generated by the model are used for classification or prediction. This also enables us to determine the overall level of confidence provided by the entire tree of rules (59.46%). The error rate is much lower than the initial 75% found with the unsegmented sample (the 75% represents the proportion of cases that do not belong to a specific selected category). Therefore, the model of rules provides an improvement of the explanatory and predictive capacity by reducing this risk from 75% to 40.54%.

However, if we make our prediction using these rules exclusively, the error rate is reduced considerably as our interest mainly lies in the rules for ROE=high and ROE=low, specially for the ones

described above and selected within each of these categories,. Thus, Table 6 shows that for the seven rules selected for ROE=high, with a sampling support of 548 firms (a decile of the entire sample), the probability of an accurate prediction increases up to 79.29% (confidence or response). This is equivalent to an index of 317.14%, i.e. more than three times higher than with the 25% of the total sample (the percentage of companies with ROE=high in the unsegmented sample). In other words, 548 companies showed the above-stated levels of variables for those seven rules described, and 79.29% of them achieved high ROE. This figure measures the level of confidence in the set of seven rules defined for ROE=high, while the individual level of confidence for each of these rules was shown on Table 4. This set of rules has a sampling support of 548 firms, which represents a decile of the entire sample, and it explains 31.69% (gain) of the companies with ROE=high.

Furthermore, Table 6 also illustrates the level of confidence for the set of six rules obtained for ROE=low, with the corresponding gain and index indicators, for each of which similar goodness analyses could be made. In fact, results now present better figures, with a percentage of confidence over 90% and an index over 360%, which indicates that predictions carried out with these rules provide around 3.6 times more accurate than those presented at the unsegmented sample into quartiles.

Finally, the following charts in Figure 1 illustrate the gains, responses and indices for the set of rules obtained with the CHAID modelling for

Table 5.
Estimation of Risks with the entire tree of rules

Sample/Predited	ROE=low	ROE=low medium	ROE=high medium	ROE=high
ROE=low	1022	286	44	19
ROE=low médium	256	768	281	66
ROE=high medium	131	287	606	347
ROE=high	77	136	293	865
RISKS complete tree	Correct	3.261 (59.46%)	Wrong 2.223 (40.54%)	

Table 6.
Gain, Response and Index with the rules selected for ROE=high an ROE=low

RULES SELECTED	PERCENTIL (%)	PERCENTIL (n)	GAIN (n)	GAIN (%)	RESPONSE (%)	INDEX (%)
7 RULES Top Decil ROE=HIGH	10	548	434	31.69	79.29	317.14
6 RULES Top Decil ROE=LOW	10	548	495	36.08	90.26	361.05

the classification of the categories ROE=high and ROE=low respectively. It allows evaluating the effectiveness of all the rules for both categories. Note that they show the behaviour of those measurements with respect to the different percentiles. In particular, for the tenth percentile of rules, the values are the same than those previously provided by Table 6, corresponding to the set of rules selected in this table and studied in this paper.

In particular, the Response Chart indicates the level of confidence in the rules; thus, for example, the top decile of rules selected for ROE=high has 79.29% of confidence. In the same way, for the top decile of rules selected for ROE=low, this percentage reaches up to 90.26%, that is, almost 100% of confidence in the rules shown. Thus, it means that the higher the level of the chart over the 25% benchmark (the confidence in the prediction for the categories in the unsegmented sample), the higher the predictive capacity of the model.

On the other hand, the Index Chart also evaluates the effectiveness of this set of rules, because it measures the extent to which the companies with the profile defined by the rule (or a set of rules) are likely to achieve ROE=high when compared with any company of the unsegmented sample. As an example, the index value of 317.14% indicates that companies with profiles defined by the seven rules selected for ROE=high are 3.17 times more likely to achieve ROE=high than any other company in the whole sample. For rules selected for ROE=low, that percentage also achieves an important improvement up to 361.05%.

The Gain Chart is interpreted in a similar way, with the model presenting better goodness as the curve is higher. For example, the top decile of rules for ROE=low has 90.26% of confidence, which represents a probability of accurate prediction 3.61 times higher than the initial 25% (corresponding to the unsegmented sample).

Therefore, in all these charts the elevation of the curve above the initial slope reflects the substantial improvement in predictive and explanatory capacity achieved from applying the rules obtained with CHAID modelling and, in particular, with the rules selected at the first decile (top decile of rules) for each category studied (ROE=high and ROE=low). All the charts allow us to conclude that the set of rules selected implies an important contribution to prediction for the financial profitability of the real estate companies.

DISCUSSIONS

To sum up, the main sources of high profitability within the real estate companies can be summarized into three: (a) the high fixed assets turnover (resulting from the low fixed assets ratios); (b) the high return on sales (made possible owing to high sales and the reduction of the costs of employees); and, (c) the low asset coverage ratio (low fixed assets and high leverage). All of them were provided by an expansive economic cycle that is characterized by a high level of sales and a high leverage of the companies bound up with low interest rates. The level of assets was not specifically an important explanatory variable of the profitability, confirmed by some previous studies about profitability among firms
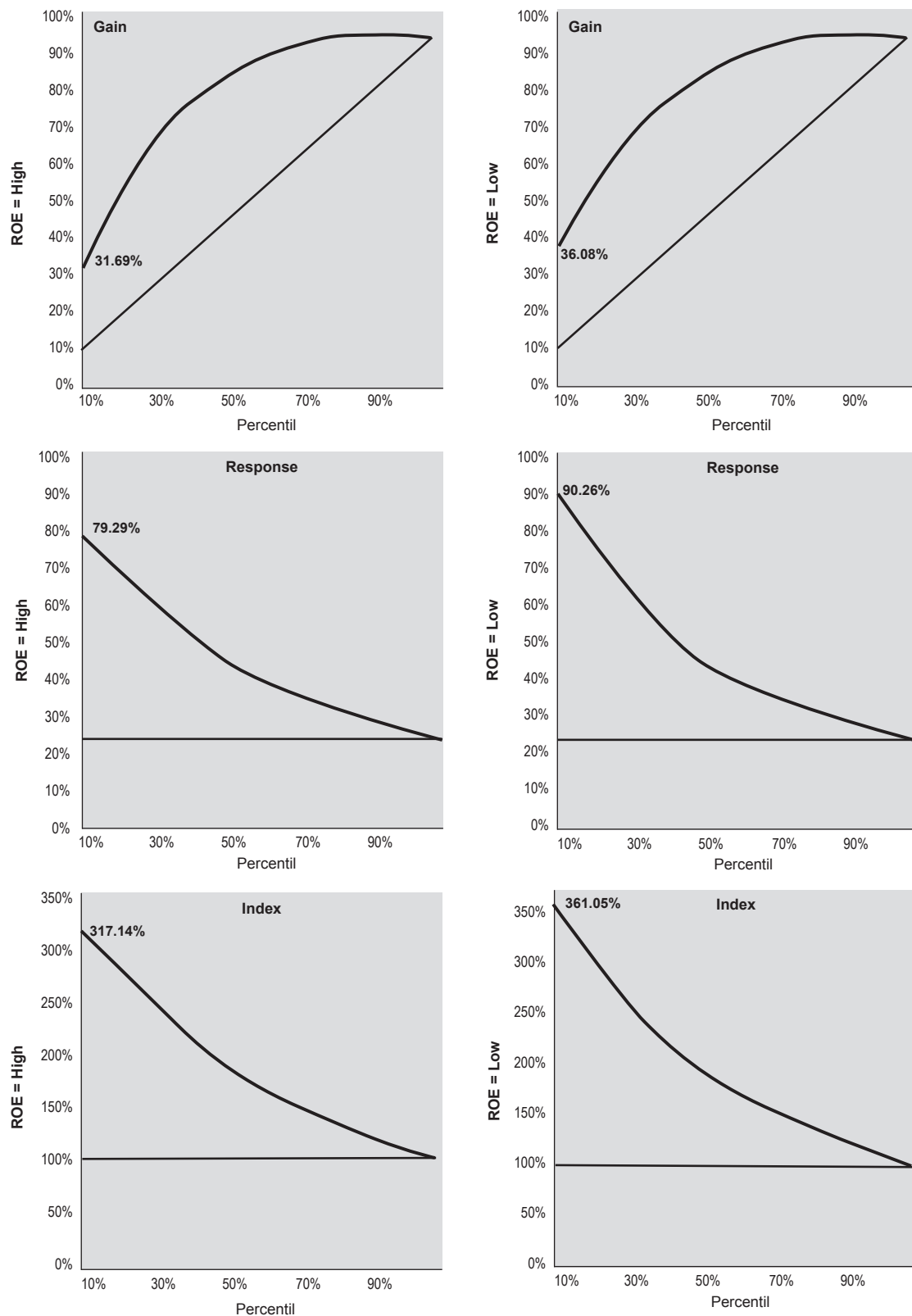
Figure 1. Gain, Response and Index with all the rules obtained for ROE=high and ROE=low

with different sizes (Rodríguez, 1989; Galán, 1997). Nevertheless, sales as a measurement of size appear to be the main implicit explanatory variable of ROE, because it allowed companies to reach higher return on sales and turnover (Galvé & Salas, 1993; González, 1997).

However, in the current marketplace, it will be difficult to maintain those high figures for variables such as return on sales, leverage or turnover. In fact, there are several phenomena endangering the perspectives close to the real estate companies, among which are included the fall of their economic activity and sales, the lack of financing and the mortgage loan restrictions due to the hardening of loan concession criteria. Also, the exorbitant growth of unemployment in families and the bankruptcy of companies loom in the future, resulting in an important increase of the default rate.

The analysis holds up that the main weakness of the housing companies lies in their high stocks and lower fixed assets, which, compounded with high debt, result in risky solvency and liquidity ratios. All of them bring the real estate sector into a vulnerable status to face the economic and financial crisis, with an important collapse of the levels of profitability. Firstly because the high asset turnover ratios are not caused by high current asset turnover, as would be desirable to limit the impact of the crisis on ROA, but by high fixed asset turnover, enabled by the high sales and the possibility to maintain reduced solvency ratios (low fixed assets and high debt). In the current environment, those high sales and low solvency ratios will not be sustainable: on the one hand, it will result in a fall of ROA due to the reduction of turnovers caused by the drop in sales and the need to increase the ratios of fixed assets, and on the other hand, the fall of ROE due to the reduction of high debt ratios to improve solvency. Secondly, because of low labor costs, depreciation and finance costs about sales have led to high sales, but it was only possible due to the high volume and strong growth of sales. As a result, these companies are very sensitive to drops in sales at periods as the current crisis, which also explains the current decline of ROA.

This paper contributes with an exploratory analysis of the real estate sector and provides a set of rules obtained by the CHAID algorithm, which could help companies to know the level to be achieved by the different accounting ratios if they want to obtain high figures of ROE. So, it offers profiles of profitability and recommendations to real estate companies regarding the main variables and accounting ratios that may influence ROE, as well as to ascertain the most suitable values for them. Taking into account these values and in order to obtain high levels of ROE, these firms should aim to achieve those profiles described, in particular, by reducing the stock ratios to increase the current asset turnovers and to compensate the decline in asset turnover caused by the collapse of sales.

In a scenario of general economic contraction, the housing sector must continue its particular and rigorous adjustment of stocks. The statistics confirm the decrease in production, transactions, mortgage loans and, for the first time, property prices, originated by the extensive real estate stocks available for sale that are pushing down prices. This will bring in a traumatic adjustment of supply, but it will be also necessary if real estate companies want to reach sustainable ratios of asset turnover. This scenario implies to reduce stocks in order to increase the asset turnover ratios, far away from excessive historical stocks, as shown in this paper.

Finally, from a methodological point of view, it would be appropriate to apply other algorithms to compare the stability and prediction power of the model created, i.e. the advanced version C5.0 (Chesney, 2009). This is true particularly because we are aware that the discretization of the continuous explanatory variables could represent a strongly impressive pre-processing statement. In this study we have focused on the implementation of the CHAID method to obtain preliminary results as a starting point for future research methodologies on which we are currently working, not only the algorithm C5.0, but also the Neural Networks and the Support Vector Machines (SVM).

References

Aguilar, I. (1989). *Rentabilidad y riesgo en el comportamiento financiero de la empresa.* Las Palmas, España: CIES.

Antón, C., Cuadrado, C., & Rodriguez, J. A. (1990). Factores explicativos del crecimiento y la rentabilidad. *Investigaciones Económicas (Second),* 153-158.

Berka, P., & Bruha, I. (1998). *Principles of Data Mining and Knowledge Discover*. Nantes, France: Springer Berlin.

Bermúdez, J. M. (2008). El sector inmobiliario, básico para reactivar la economía. *Revista del sector inmobiliario, 11*(86)*,* 63-67.

Bigss, D., Ville, B., & Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics*, *18*(1), 49-62.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees.* Monterey, CA: Wadsworth and Brooks-Cole.

Bueno, E. (Comp). (1990). *La empresa española: estructura y resultados.* Madrid, España*:* Instituto de Estudios Económicos.

Chesney, T., & Penny, K. (2009). Data mining trauma injury data using C5.0 and logistic regression to determine factors associated with death. *International Journal of Healthcare Technology and Management*, *10*(1/2), 16-26.

Dills, A. K. (2005). Does cream-skimming curdle the milk? A study of peer effects. *Economics of Education Review*, *24*, 19-28.

Fariñas, J. C., & Rodríguez, L. (1986). Rentabilidad y crecimiento de las grandes empresas industriales españolas en comparación con las de la CEE (1973-1982). *Información comercial española*, (agosto-septiembre), 87-101.

Fariñas, J. C., Calvo, J. L., & Jaumandreu, J. (1992). *La PYME industrial en España.* Madrid, España: Ediciones Civitas.

Fernández, A., & García, M. (1991). Análisis del comportamiento económico-financiero de los sectores empresariales en España. *Esic Market* 72, (abril-junio), 113-128.

Fernández, E., Montes, J. M., & Vázquez, C. J. (1996). Caracterización económico-financiera de la gran empresa industrial española según su rentabilidad. *Revista Española de Financiación y Contabilidad, 87*(XXV), 343-359.

Ferruz, L., & Andreu, L. (2007). El sector inmobiliario en España. Evolución y perspectivas. *Revista de la Asociación Española de Contabilidad y Administración de Empresas, 78.*

Galán, J. L., & Vecino, J. (1997). Las fuentes de rentabilidad de las empresas. *Revista Europea de Dirección y Economía de la Empresa*, *6*(1), 21-36.

Galguera, L.; Luna, D.; & Méndez, M. P. (2006). Predictive segmentation in action: using CHAID to segment loyalty card holders. *International Journal of Market Research*, *48*(4), 459-479.

Galvé, C., & Salas, V. (1993). Propiedad y resultado de la gran empresa española. *Investigaciones Económicas*, *XVII*(2)

González, A. L. (1997). *La rentabilidad empresarial: evaluación empírica de sus factores determinantes.* Colegio de Registradores de la Propiedad y Mercantiles de España. Centro de Estudios Registrales.

González, A., Correa, A., & Acosta, M. (2002). Factores determinantes de la rentabilidad financiera de las PYMES. *Revista Española de Financiación y Contabilidad*, *XXXI*(112), 395-429.

Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, *74*, 537-552.

Grobler, B. R., Bisschoff, T. C., & Moloi, K. C. (2002). The CHAID-technique and the relationship between school effectiveness and various independent vari-

ables. *International Studies in Educational Administration, 30*(3), 44-56

Huergo, E. (1992). Tamaño y rentabilidad en la industria española. *Economía Industrial, 284,* 41-49.

Hunt, E. B., Marin, J., & Stone, P. J. (1966). *Experiments in Induction,* New York: Academic Press.

Illueca, M., & Pastor, J. M. (1996). Análisis económico financiero de las empresas españolas por tamaños. *Economía Industrial, 310,* 41-54.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics,* 29, 119-127.

Lizcano, J. (2004). *Rentabilidad empresarial: Propuesta práctica de evaluación.* Madrid, España*:* Consejo de Cámaras de Comercio de España..

Lucas, P., & González, A. (1993). Rentabilidad de la inversión y recursos propios en la empresa industrial. Análisis en función de la propiedad y el sector. *Economía Industrial, 293* (setiembre/octubre), 19-36.

Maroto, J. A. (1993). La situación económico-financiera de las empresas españolas y la competitividad. Aspectos generales y particulares de la financiación de las PYMES. *Economía Industrial* (mayo-junio), 89-106.

Maroto, J. A. (1998). Central de Balances del Banco de España. *Cuadernos de Información Económica, 140141,* 186-196.

Quinlan, J. R. (1979). Discovering rules by induction from large collection of examples. In D. Michie (Ed.),

Expert Systems in the Microelectronic Age (pp. 168-201). Scotland: Edinburgh University Press.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning.* San Mateo, CA: Morgan Kaufmann.

Rodríguez, E. (2002). Análisis económico-financiero del sector de la automoción en España. *Boletín Económico de ICE, 2747,* 13-22.

Rodríguez, L. (1989). Rentabilidad económica y crisis industrial: evolución de la rentabilidad y sus factores explicativos. *Papeles de Economía Española, 39,* 356-373.

Salas, V. (1994). Economía y financiación de la empresa española según su tamaño. *Situación, 2,* 197-212.

Sánchez, A. (1994). La rentabilidad económica y financiera de la gran empresa española. Análisis de los factores determinantes. *Revista Española de Financiación y Contabilidad, 78*(XXIII), 159-179.

Santín, D. (2006). La medición de la eficiencia de las escuelas: una revisión crítica. *Hacienda Pública Española / Revista de Economía Pública, 177*(2), 57-82.

Strambi, O. (1998). Trip generation modeling using CHAID, a criterion-based segmentation modeling tool. *Journal of the Transportation Research Board, (1645), 24-31.*

Suárez, A. (1977). La rentabilidad y el tamaño de las empresas españolas. *Económicas* y *Empresariales, 3,* 116-132.