Investment decision on the Romanian capital market using data mining

Dissertation paper

The University of Economic Studies

The Faculty of Finance, Assurance, Banks and Stock Exchange

DAFI-master in financial and stock exchange management

June 2014, Bucharest

Author: Eugen-Alexandru Pantea

Coordinator: Prof. Univ. PhD Anamaria Ciobanu

Contents

Abstract

The paper is structured in three chapters. The first chapter begins with the introduction, the motivation of the theme and the research context. Further, I talked about the research objectives and theoretical concepts regarding: fundamental analysis, technical analysis, investor versus speculator and about the value investing concept. The first chapter ends with a review of the main scientific articles about companies' classification by means of discriminant analysis and with the results obtained in research.

In the second chapter we found the empirical research, database selection, descriptive analysis of the data and definitions and assumptions for the fifteen financial ratios used in this research. Further, I executed hierarchical cluster analysis and k-means analysis for 2011 and 2012. Then, I classified the companies by constructing an original index. The predictive nature of the research was realised by constructing the discriminant function: **D=a+v1X1+v2X2+... vnXn,** where D is the function, a is the constant, v_i is the discriminant coefficient and X_i is the variable score. The discriminant function was constructed using 2011 database and tested on the 2012 data base.

Using the classes obtained by k-means analysis, a class with good companies and another class with weak companies (I reduced the number of weak companies thus the number of weak ones to be balanced with the good ones), I constructed two portfolios in order to demonstrate the applicative nature of the research in the real life. The returns of the portfolios demonstrated the utility of the analysis since the good companies portfolio obtained about a return of 60% in two year and three month (including the dividends), the weak companies portfolio achieved a little loss (-0.8%) and in the same period the capital market (measured with BET-C index) gained 24% (without dividends).

Therefore, I reached two main objectives: obtaining two different classes of stocks with respect to a capital market investor (a class with good to invest companies/stock and a class with weak companies/stocks) and obtaining a higher portfolio return than the market, which justify the effort made doing the fundamental analysis and the data mining.

The paper ends with the third chapter which contains the conclusions, discussions and future recommendations.

Introduction

The research focuses on the relation between the value of stocks and the financial ratios. The main purpose of this research is to classify the companies listed on the Bucharest Stock Exchange (annex [4]) in two categories: good and weak companies in which to invest. The main idea is to use fundamental analysis in stocks evaluation by evaluating the financial ratios calculated with data extracted from the annual financial reports released by listed companies.

I am aiming to classify the companies in two groups, though in similar researches the authors used three groups (Jo Vu 2010). The analysis is made for a two year period: 2011 and 2012 and the databes from 2011 is used to calculate the discriminant function upon which I made the class prediction for the 2012 database. Before start the cluster, hierarchical, k-means and discriminant analysis, I classified the companies by my own method. This method consist of an index which may take values between zero and fifteen points. To every financial ratio from all fifteen I give a point if this ratio is situated favorable compared with the average value of that financial ratio. Besides D/E and D/A which are given a point if they are situated bellow the average, all the financial ratios are given a point if they are situated above the average.

Thus, a company if it have all fifteen ratios situated favorable compared with mean have an index value of fifteen and it is considered a good company, and on the other side, the worst company can have a zero index value. I considered that a good company should have an index value greater than eight (which means eight financial ratios out of fifteen situated favorable compared with the mean) and all the companies with a lower value than eight should be considered weak companies.

Methods and Results

I intended to analyse from the fundamental point of view the companies listed on Bucharest Stock Exchange. I take into consideration companies listed on the first and the second tier, around 70 companies. I excluded the banks, insolvent companies and companies with incomplete financial statements. After this exclusion remained around 65 companies from 11 sphere of activity. I collected annual financial data from financial statements (balance sheets, profit and loss statements and cash flow statements) for 2011 and 2012. I choose annually data taking into consideration the stability of the annually figures and the completeness of the annulay reports.

In the paper, I used fifteen financial ratios selected from a larger number of ratios which I calculated initially. I decided to use these fifteen ratios considering: 1) the popularity of the ratios in the financial literature and 2) personal decision (I selected the ratios which I follow when I take investment decision regarding my real stocks portfolio). After the construction of the database with companies and financial ratios I eliminated outliers and I standardised the data in order to eliminate the influence of measures and the size of the data.

The fifteen financial ratios used in analysis:

1. Tobins'Q
2. Dividend yield
3. Dividend
4. ROE
5. ROA
6. Net margin
7. D/E
8. D/A
9. Liquidity of the shares
10. EPS
11. PER
12. P/BV
13. Free-float
14. Operational cash flow per share
15. Cash per share

Hierarchical cluster analysis and k-means

I analysed 67 companies (63 after eliminating the objects that presents outliers) listed on the Bucharest Stock Exchange from 2011 and 2012. I grouped the companies using hierarchical analysis (unsupervised method) and by k-means (supervised method) into two classes: a class with good to invest companies and a second class with weak companies.

In generally, clusterization represent the arrangement activity or the association of objects (here companies) into groups, according to the similarity or disparity between objects. Hierarchical cluster analysis represent an unsupervised classification method, because we don't know the number of classes neither the membership of the objects to the classes. In cluster analysis there are two fundamental criteria: the objects classified into classes being as similar as possible in terms of specific characteristics and the objects classified into a class differentiate as much as possible from the objects classified in the other classes (Ruxanda 2011).

K-means is a supervised method for clusterization because it assume that we know before performing the k-means analysis the number of the clusters. After analysing the dendrograms (annex [1] and annex [3], I noticed the formation of two principal classes and before I start the analysis my goal was to split the companies population into two classes: a class of good to invest companies and a second class which contain weak companies who I don't take into consideration in the investment decision.

At this moment, after hierarchical analysis, I obtained two groups of companies, but I did not know the quality of the classes. In order to identify the classes quality, I compared the mean of each financial ratios in every class with the mean of each financial ratios calculated on the entire group of companies.

In 2011 using hierarchical cluster method (Ward1 method with Euclidean squared distance) , I obtained a class with 13 companies: COTE, FP, SIF1, SIF2, SIF3, SIF4, SIF5, TEL, TGN, BVB, COTR and ELGS and a class with the others fifty companies.

[1] Ward's method is the only one who minimise the variability intracluster and maximise the variability intercluster (Ruxanda 2011)

Applying k-means method with two classes I obtained a class with 12 companies and another class with 51 companies. In the first cluster the method grouped: COTE, FP, SIF1, SIF2, SIF3, SIF4, SIF5, SNP, TGN, BVB, CBC and ELGS. I notice that most of the companies remain in the good companies class.

I performed in 2012 the same analysis as in 2011 with the mention that for the hierarchical cluster analysis I used the Ward's method with Chebyshev distance. Analysing the dendrogram for 2012, I noticed the formation of two main classes: in the first class are grouped 13 companies: FP, SIF1, SIF2, SIF3, SIF4, SIF5, BRK, ROCE, RPH, BVB, COTE, TGN and RRC, and in the second class 46 companies. K-means method grouped the companies in two classes: 18 in the first one and 41 in the second one. In the first cluster are grouped: BIO, BRK, COTE, FP, SIF1, SIF2, SIF3, SIF4, SIF5, SNP, TGN, ARTE, BRM, BVB, ELGS, MECF, PTR and SCD. As in 2011, in 2012 the first class contain the good companies.

Discriminant analysis

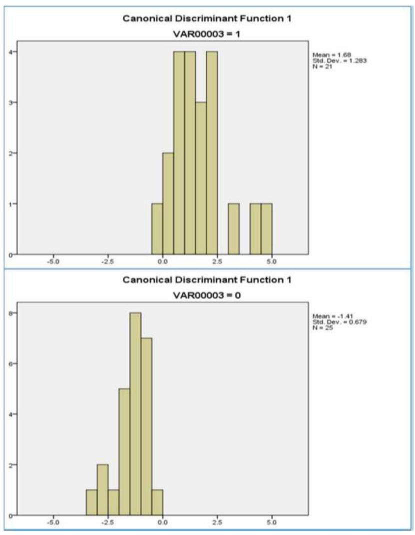
The discriminant analysys is used to detect which variables discriminate between two or more groups of objects. Discriminant analysis is based on a predictive model which groups the objects into classes. The model consist in a liniar combination of predictive variables as: **D=a+v1X1+v2X2+... vnXn,** where D is the function, a is the constant, v_i is the discriminant coefficient and X_i is the variable score.

The purpose of the discriminant analysis is to maximise the distance between classes and minimise the distance between objects into classes. The main advantage of the discriminant analysis is the simultaneous analysis of multiple indicators. Discriminant analysis need a categorical dependent variable which takes two values: one if the company is a good one and zero if the company is weak and the independent variables (fifteen financial ratios used in research). As a categorical variable I used the classification obtained with the own index analysis described above. From my classification I obtained in 2011 21 good companies and 42 weak companies, but for an accurate analysis I decided to balance the classes, taking into consideration a group with 21 good companies and another group with 21 weakest companies after my index analysis. Wilk's Lambda test (0.287;sig.=0.000) for 2011 indicate that the discriminant function discriminate well between companies.

The 2011 discriminant function looks like:

*D=0.172+0.222*Tobin'sQ+0.72*dividend_yield+0.325*dividend+0.58*ROE+0.129*ROA+0.228*net margin-0.613*D/E-0.278*D/A+0.154*liquidity-0.419*EPS-0.116*PER+2.44*P/BV-0.095*free_float+0.213*operational cash flow per share+2.4*cash per share*

The 2011 discriminant function grouped overall correct 95.7% companies, 100% in the weak class and 90.5% in the good class.



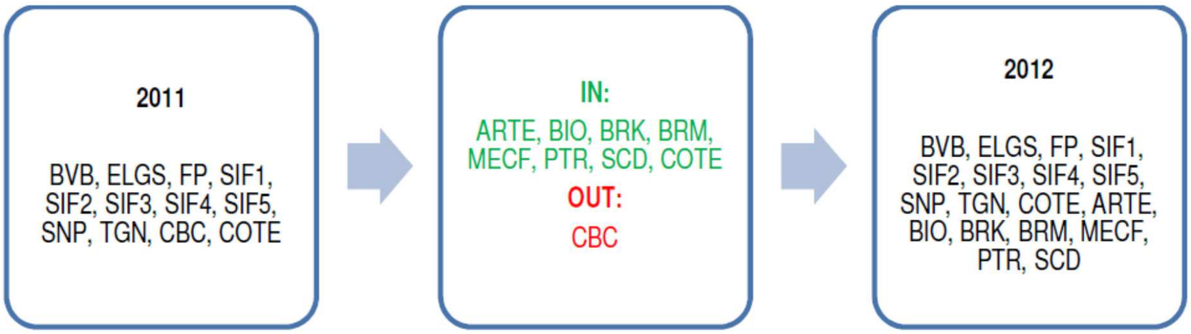
These histograms shows the distribution of discriminants scores for each class. If the histograms does not overlap it means that we have two different classes.

Class prediction 2012

Using the discriminant function obtained with 2011 database (46 companies, 21 good and 25 weak) I calculated the scores for the companies from 2012 database (59 companies). I grouped the 2012 companies in two classes using the scores obtained applying the discriminant function. The good companies have a score greater that 0.3 and the weak companies presents scores lower than -0.2. Thus, in good companies class I have 25 companies and in the weak class I have 34 companies. In order to test the new classification, I apply again the discriminant analysis, this time on the 2012 database and using as a grouping dependent variable the classification obtained applying the discriminant function. Wilk's Lambda test (0.237, sig. 0.000) shows that there are significant difference between the two groups.

The 2012 discriminant function grouped overall correct 94.9% companies, 100% in the weak class and 88% in the good class.

Portfolios



I formed two portfolios, one with good companies (above graph), which I bought at the end of the 2011, and another with weak companies. I invested equally 10.000 ron in each portfolio, and every stock have an equal weight at the beginning of the period. When the first period ends, I redone the k-means analysis and rebalanced the portfolios with new classicifation.

The portfolio with good companies obtained a return (including dividends) of 59% after two year and three months (from December 2011 to march 2014), the portfolio with weak companies lost 0.8% and in the same period the market grow up with 23.7%. An interesting fact is that in 2014 the weak companies portfolio obtained 11.5% and the good companies registered a loss of 8%. This could happened because in 2014 I dindn't perform k-means analysis and I decided to keep the same companies into the two portfolios.

Discussion

Weak points:

- Database size (I analysed around 70 companies, but in similar analyses authors used from 100 to 500 companies)
- Period analysed: I consider in the future to use a larger time horizon
- I retained in analysis companies from different activity sectors. Authors grouped only companies from the same sector.
- Qualitative data: I didn’t use qualitative data in the analysis.

Future recommendations

- To Optimise the financial ratios selection (for example I can use principal components analysis to narrow the data base);
- To Analyse other markets and other time horizons;
- To use qualitative data;

Works Cited[2]

1. Altman I. E.,1968, ‘Financial ratios, discriminant analysis and the prediction of corporate bankruptcy’, *The Journal of Finance*, No.4, pp.589-609.

2. Brown, Caylor, 2006, ‘Corporate governance and firm valuation’, *Journal of Accounting and Public Policy,* No.25, pp.409-434.

3. CFA Program Curriculum Level I, 2013, CFA Institute, John Wiley & Sons, Inc.

4. David D., 1941, ‘Risk elements in consumer instalment financing’, *The National Bureau of Economic Research*, pp. 83-91.

5. Fama F.E., 1970, ‘Efficient capital markets: A review of theory and empirical work’, *Journal of Finance*, Vol. 25, issue 2, pp. 383-417.

6. Graham B, Dodd B. L., 1934, *Security Analysis*, Wjittlesey House, McGraw Hill Book company Inc., New York.

7. James E. W., 1959, ‘A discriminant function for earnings-price ratios of large industrial corporations’, *The Review of Economics and Statistics*, vol. 41, No. 1, pp. 44-52.

8. Joseph P., 2002, ‘Value investing: The use of historical financial statements information to separate winners from losers’, *Journal of accounting research,* Vol.38, pp. 1-41.

9. Keynes J.M., 2009, *Teoria generală a ocupării forței de muncă, a dobânzii și a banilor,* ed.Publica, trad. I. Haita și Corina Mădălina, București.

10. Ruxanda Gh., 2001, *Analiza Datelor*, Editura ASE, București.

11. Schroeder A., 2009, *The snowball Warren Buffett and the Business of Life-Updated and condensed,* ed. Bloomsbury Publishing, London.

12. Spircu L., 2005, *Analiza datelor: aplicatii economice*, Editura ASE, Bucuresti.

13. Vayrynen O., 2007, *Identifying undervalued stocks with multiple financial ratios*, Master’s Thesis, Helsinki University of Technology.

14. Vu J., 2011. ‘Discriminant analysis and the identification of high value stocks’, *The 31st annual international symposium in forecasting*, Praga-Republica Ceha.

[2] This bibliography was used in the main work, this paper being a summary of the main work.

15. www.kmarket.ro/documentare/analiza_actiuni.pdf, Paşol R, 2011'*Mic manual de analiza a actiunilor'* [7.04.2014].

16. www.intercapital.ro/intercapital_start, baze de date [01.02.2014]

17. www.bvb.ro, baze de date [29.03.2014].

18. www.investopedia.com/articles/fundamental-analysis/09/free-cash-flow-yield.asp [15.03.2014]

19. www.cs.uu.nl/docs/vakken/arm/SPSS/spss6.pdf, SPSS Discriminant Analysis [20.05.2014]

20. www.uk.sagepub.com/burns/website%20material/Chapter%2025%20-%20Discriminant%20Analysis.pdf, Discriminant Analysis, [20.05.2014];

21. core.ecu.edu/ofe/StatisticsResearch/SPSS%20Discriminant%20Function%20Analysis.pdf, Bian. H.,'SPSS Discriminant Function Analysis' [21.05.2014];

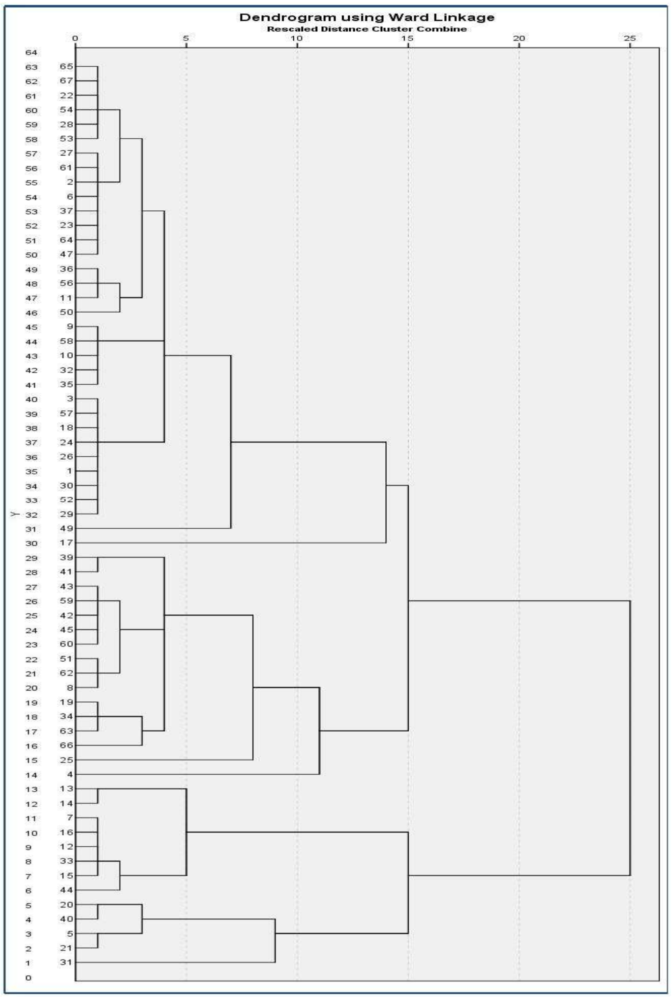
22. www.kmarket.ro/documentare/analiza/EvolutiePiataCapital2013.pdf [05.04.2014]

23. office.microsoft.co, Excel 2013.

24. www-01.ibm.com/software/analytics/spss, SPSS.

Appendices

Annex [1] - Dendrogram for 2011

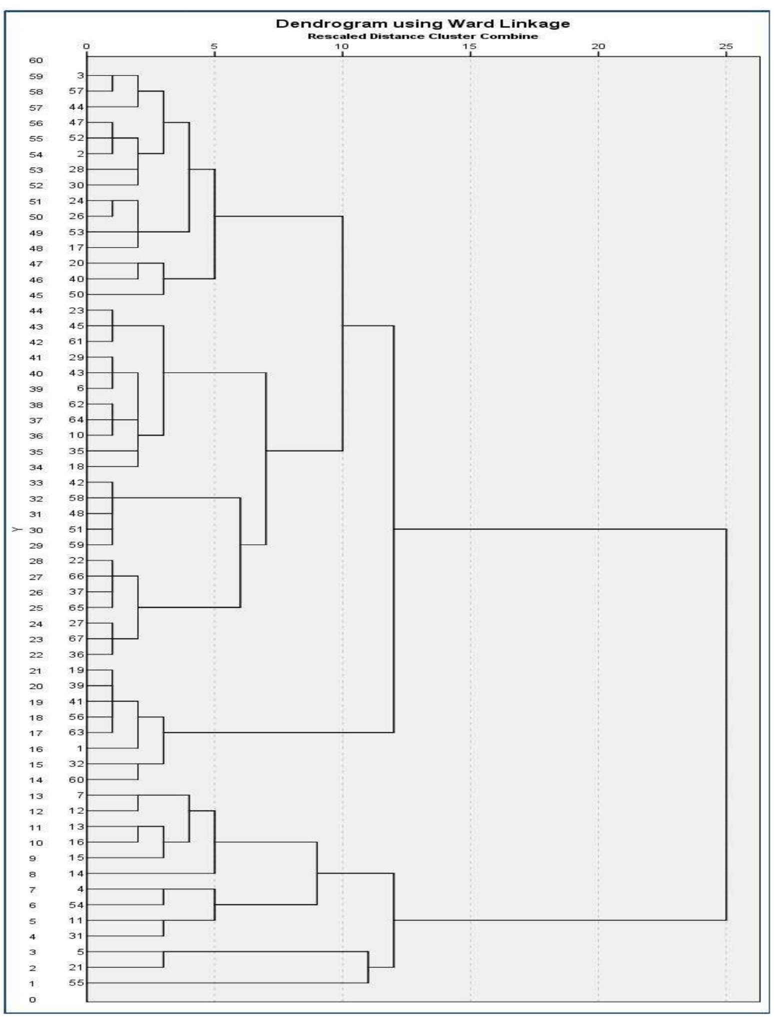


Annex [2] – Hierarchical cluster analysis 2012

Case Processing Summarya 2012					
Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
59	88.1	8	11.9	67	100.0

a. Ward Linkage

Annex [3] – Dendogram for 2012



Annex [4] – The list with listed companies on the Bucharest Stock Exchange (tier I and II)

Symbol	Company's name	Symbol	Company's name
ALR	ALRO S.A.	ALT	ALTUR S.A.
ATB	ANTIBIOTICE S.A.	ALU	ALUMIL ROM INDUSTRY S.A.
BCC	BANCA COMERCIALA CARPATICA S.A.	AMO	AMONIL S.A.
BIO	BIOFARM S.A.	APC	voestalpine VAE APCAROM SA
BRD	BRD - GROUPE SOCIETE GENERALE S.A.	ARM	ARMATURA S.A.
BRK	S.S.I.F. BROKER S.A.	ARS	AEROSTAR S.A.
COFI	CONCEFA SA SIBIU	ART	TMK - ARTROM S.A.
COTE	CONPET SA Ploiesti	ARTE	ARTEGO SA Tg. Jiu
ELMA	ELECTROMAGNETICA SA BUCURESTI	BCM	CASA DE BUCOVINA-CLUB DE MUNTE
FP	SC FONDUL PROPRIETATEA SA - BUCURESTI	BRM	BERMAS S.A.
IMP	IMPACT DEVELOPER & CONTRACTOR S.A.	BVB	SC BURSA DE VALORI BUCURESTI SA
OIL	OIL TERMINAL S.A.	CAOR	CALIPSO SA ORADEA

OLT	OLTCHIM S.A. RM. VALCEA	CBC	CARBOCHIM S.A.
PREH	PREFAB SA BUCURESTI	CEON	CEMACON SA CLUJ-NAPOCA
RPH	ROPHARMA SA BRASOV	CGC	CONTOR GROUP S.A. Arad
SIF1	SIF BANAT CRISANA S.A.	CMCM	COMCM SA CONSTANTA
SIF2	SIF MOLDOVA S.A.	CMF	COMELF S.A.
SIF3	SIF TRANSILVANIA S.A.	CMP	COMPA S. A.
SIF4	SIF MUNTENIA S.A.	CNTE	CONTED SA DOROHOI
SIF5	SIF OLTENIA S.A.	COMI	CONDMAG S.A.
SNG	S.N.G.N. ROMGAZ S.A.	COS	COS TARGOVISTE S.A.
SNN	S.N. NUCLEARELECTRICA S.A.	COTR	SC TRANSILVANIA CONSTRUCTII SA
SNP	OMV PETROM S.A.	DAFR	DAFORA SA
SOCP	SOCEP S.A.	ECT	GRUPUL INDUSTRIAL ELECTROCONTACT S.A.
TBM	TURBOMECANICA S.A.		TURISM, HOTELURI, RESTAURANTE MAREA

		EFO	NEAGRA S.A.
TEL	C.N.T.E.E. TRANSELECTRICA	ELGS	ELECTROARGES SA CURTEA DE ARGES
TGN	S.N.T.G.N. TRANSGAZ S.A.	ELJ	ELECTROAPARATAJ S.A.

TLV	BANCA TRANSILVANIA S.A.	ENP	COMPANIA ENERGOPETROL S.A.
		EPT	ELECTROPUTERE S.A.
		MECF	MECANICA CEAHLAU
		MEF	MEFIN S.A.
		MJM	MJ MAILLIS ROMANIA S.A.
		PEI	PETROLEXPORTIMPORT S.A.
		PPL	PRODPLAST S.A.
		PTR	ROMPETROL WELL SERVICES S.A.
		RMAH	FARMACEUTICA REMEDIA SA
		ROCE	ROMCARBON SA BUZAU
		RRC	ROMPETROL RAFINARE S.A.
		RTRA	RETRASIB SA SIBIU
		SCD	ZENTIVA S.A.
		SNO	SANTIERUL NAVAL ORSOVA S.A.
		SPCU	BOROMIR PROD SA BUZAU (SPICUL)
		SRT	SIRETUL PASCANI S.A.
		STIB	STIROM SA Bucuresti

		STZ	SINTEZA S.A.
		TRP	TERAPLAST SA
		TUFE	TURISM FELIX S.A. BAILE FELIX
		UAM	UAMT S.A.
		UZT	UZTEL S.A.
		VESY	VES SA
		VNC	VRANCART SA