Risk Return Optimization Using the Knapsack Problem in The Formation of a Stocks Portfolio. Case Study of a Brazilian Investment Site.

Nicolas Sampaio Bevilaqua, OCILEIDE Custodio da Silva, GABRIELA DE MATTOS VERONEZE
Federal University of Amazonas
Brazil

Abstract

In this work, the composition of a portfolio was proposed by using the Knapsack problem and verified its effectiveness in comparison to a portfolio of shares on an investment website. The programming variables were based on the Markowitz risk theory of variance and following collaborators for their studies. And from the chosen portfolio, the efficient frontier was elaborated analyzing the performance of the investment site portfolio during 30 days. The portfolio obtained exceeded the percentage performance obtained from the investment site in the same period when considering the maximum possible return, the minimum global variance and also in the naive distribution.

Keywords: Integer programming; Knapsack Problem; Variance Risk; Efficient Frontier.

1. Introduction

Building an investment portfolio is a task that takes into account numerous aspects related not only to the assets themselves, but also to the profile of the investor who will be performing such a task. The common objective, however, is always the same, to obtain the highest return in the face of the lowest possible risk exposure.

The problem initially proposed by (Markowitz, 1958). He presented a model of risk variance where he proved that diversification, observing the correlation of assets, there was an increase in the theoretical portfolio return given the same risk. (Sharpe, 1967) developed the capital asset pricing model (CAPM) and created an index to measure the relationship between risk and investment performance. (Morita et al, 1989) present a stochastic maximization model through the backpack problem with a matrix of variance and covariance. (Konno & Yamasaki, 1990) showed improvements in relation to the Markowitz study introducing a risk calculated as Mean Standard Deviation (MAD). (Speranza, 1991) presented advances in the linear programming model, using the method of mean absolute deviation.

The literature shows numerous advances and studies in determining the portfolio, such as the use of *Fuzzy* and *AHP* methodologies. However, many articles fail to demonstrate the theoretical results compared to the results obtained in practice by the market, due to the great complexity in the parameters used in determining the simulation models.

This study presents a simplistic approach to the application of the *Knapsack Problem* in obtaining assets for the formation of a portfolio and to comparing the theoretical results with the practical ones. First, it was necessary to establish a basic income objective to be obtained. In this case, a Brazilian investment site was chosen that compiles the indications of the collaborating investment brokers. The ten most recommended stocks make up the official portfolio of the website where the performances of these papers were verified in the interval of 30 days. Using the indications of the main brokers, binary programming was applied, using the data of return, risk (standard deviation) and correlation where the objective functions were established as the restrictions to be used in the programming. In a second step, the efficient frontier was assembled using the covariance of the assets obtained to verify the optimal points of resource allocation.

This work is organized into four sections: Section 2 presents methodological research on the problem of knapsack problem, efficient frontier and portfolio selection. Section 3 presents the numerical results obtained and illustrations, while section 4 contains the conclusions obtained.

2. Theoretical foundations

2.1 Integer Programming

An Integer Programming problem is a model in which the constraints and the objective function are identical to those formulated in linear programming. However, in some cases, the decision variables only make sense, as in the case of the article in question, when they have integer values (Hillier & Liebermann, 2006). According to (Render, 2012) throughout the programming, we have three types of solution:

a) Pure, where they receive whole values.

b) Mixed, where some have integer values.

c) Binary, where the decision variables must receive the values of 0 or 1. (Object of study of the article).

The knapsack problem consists of the classic binary programming problem in operational research, where it seeks to determine, among the n possible objects, which of them should be carried in the backpack, taking into account their usefulness and weight. Given a weight restriction, the goal is to maximize the overall usefulness of the backpack. Equation 01 concerns the objective function of maximizing the object's usefulness. Equation 02 refers to a weight capacity restriction that cannot be overcome. Equation 03 refers to the restriction on the values that must be mandatory 0 or 1. Expressed by the functions below:

$$Fobj = Max\ z = \sum_{i}^{n} PiXi \quad (01)$$

$$\sum_{i}^{n} pixi \leq C\ max \ (02)$$

$$x_i \in (0,1) \quad (03)$$

Model parameters:

xi = utility of object i

pi = weight of object i

C max. = backpack capacity

Decision variables: 1 if the object is in the backpack or 0 otherwise.

Analogously for the work, it was proposed to change the parameters for proper application. The objective in the case of the work was to minimize the risk of the portfolios given a restriction in the average correlation of the assets with each other, which can be written as follows:

$$Fobj = Min\ z = \sum_{i}^{n} CiXi \quad (01)$$

$$\sum_{i}^{n} pixi \leq C\ orrel.máx \ (02)$$

$$x_i \in \{(0,1)\} \quad (03)$$

Model parameters:

ci = risk (Standard deviation) of stock i

xi = mean correlation of stock i

Correl. max = minimize portfolio risk given the restricted correlation between assets Decision variables: 1 if the stock is in the portfolio and 0 otherwise.

2.2 Efficient frontiers

Markowitz (1958) introduced in his famous work "Portfolio Selection" terms that are widely used until today as portfolio risk, diversification and optimization (Galiene & Stravinskyte, 2016). In addition, he was the first to prove mathematically that diversification reduced the portfolio's risk (Cibulskien & Grigaliuniene, 2007).

Based on the risk and return estimates of the assets, Markowitz proposes the creation of the so-called efficient average variance frontier (Figure 1), capable of demonstrating the maximum expected return of a portfolio against a given risk.

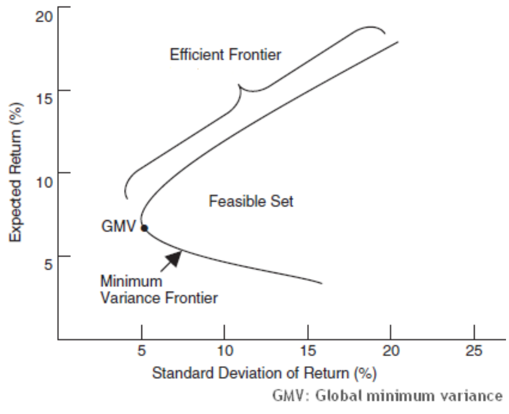

Figure 1: Exemplification of efficient Markowitz frontier

Source: Investment and Finance (2019).

The curve represents the optimal portfolios, which are: 1) The most profitable given a level of risk or 2) The lowest risk given a level of profit. However, as all options are optimal, the choice is up to each investor depending on their level of risk aversion and external factors (Galiene & Stravinskyte, 2016).
The frontier is calculated with the optimal point for each estimated theoretical return interval, in order to minimize the covariance given the following restrictions: maximum and minimum allocation in each asset (guarantee diversification) and the maximum and minimum value of each interval that is desired obtain the optimal allocation point, since without these restrictions we would have a single optimal value, and it would not be possible to set up the efficient frontier. As per the schedule below:

Decision Variables:

$$x_i = covariance\ of\ each\ investment$$

Decision Parameters:

$$R_i = Return\ of\ portifolio$$
$$R_s = Return\ of\ simulation$$

Objective Function:

$$Min = \sum_{i}^{n} x_n\ \ \forall i\ (01)$$

Subject to:
Maximum percentage allocated to each share:

$$\sum_{i}^{n} x_i \leq R_i\ \forall i\ \ (02)$$

Minimum percentage allocated to each share:

$$\sum_{i}^{n} x_i \leq R_s\ (03)$$

Return of the portfolio above the minimum limit for the simulation:

$$\sum_{i}^{n} x_i \geq R_i\ (04)$$

No Negativity:

$$x_i \in (0,1)\ \ \ \forall i$$

Due to the complexity of the calculations that involves several variables, common in problems of optimization of variables, the formulations above were performed in Microsoft Excel®.

3. Methodology

The objective was to obtain a portfolio of 10 shares in order to have a yield higher than that obtained by the official portfolio of the website in question. The *website's* portfolio was obtained from the 10 most recommended stocks by brokers. In this way, a binary model was proposed, in which 10 shares were obtained, which must have a percentage yield higher than that of the site in question.

3.1 Shares chosen

Data were extracted from the daily quotations of 22 shares, resulting from the indications of the 5 brokerages with the best percentage performance accumulated until the month of November 2019, from September to November 2019, from the official Brazilian investment *website*. Thus, the average daily risk and return of the 90 days prior to the beginning of the 30-day simulation was obtained, as shown in Table 1.

Table 1 - Average daily risk image, average monthly return.

	LREN3F	MOVI3F	RADL3F	SULA11F	VIVT4F	EZTEC3F	JBSS3F	KLBN11F	PETR4F	VVAR3F	B3SA3F	LINX3F	PCAR4F	BPAC11F	FLRY3F	GRND3F	LAME4F	WEGE3F	BBDC4F	CPFE3F	CVCB3F	VALE3F
RET	0,0006	0,0011	0,0037	0,0017	0,0006	0,0013	-0,0001	0,0028	0,0023	0,0024	0,0013	0,0004	-0,0010	0,0030	0,0029	0,0051	0,0029	0,0047	0,0007	-0,0004	-0,0043	0,0015
DESV	0,0151	0,0173	0,0142	0,0152	0,0112	0,0243	0,0250	0,0161	0,0149	0,0231	0,0147	0,0247	0,0154	0,0209	0,0136	0,0143	0,0177	0,0154	0,0163	0,0155	0,0259	0,0168

Source: Research data (2020).

The following were considered for optimization: the average risk (standard deviation) and the correlation average of each stock, obtained by Microsoft Excel®, of the 22 pre-selected assets as shown in table 2. The risk minimization program given was applied the mean correlation constraint as shown in table 2 in the appendix.

In table 3 we can see the result of the programming carried out, the 10 actions were obtained where the value of the minimization obtained was 0.1466 with the correlation of 2.39, thus respecting the restriction of 2.40, which was obtained through *benchmarking* of the correlation of the portfolios of the main brokers (average of 0.24 per share).

Table 3 - Result of chosen shares

Share	(Stdev)	Correlation	Object	Result	Stdev_tot
LREN3F	0,0151	0,3698	0	0,0000	0,1466
MOVI3F	0,0173	0,3396	0	0,0000	
RADL3F	0,0142	0,2185	1	0,2185	
SULA11F	0,0152	0,2530	1	0,2530	
VIVT4F	0,0112	0,2864	1	0,2864	
EZTEC3F	0,0243	0,2667	0	0,0000	
JBSS3F	0,0250	0,0829	0	0,0000	

KLBN11F	0,0161	0,2220	1	0,2220	
PETR4F	0,0149	0,2038	1	0,2038	
VVAR3F	0,0231	0,3313	0	0,0000	
B3SA3F	0,0147	0,3278	0	0,0000	
LINX3F	0,0247	0,2186	0	0,0000	
PCAR4F	0,0154	0,3108	0	0,0000	
BPAC11F	0,0209	0,2367	0	0,0000	
FLRY3F	0,0136	0,2722	1	0,2722	
GRND3F	0,0143	0,3091	1	0,3091	
LAME4F	0,0177	0,3575	0	0,0000	
WEGE3F	0,0154	0,1661	1	0,1661	
BBDC4F	0,0163	0,2053	1	0,2053	
CPFE3F	0,0155	0,2543	1	0,2543	
CVCB3F	0,0259	0,2521	0	0,0000	
VALE3F	0,0168	0,2013	0	0,0000	
			10	2,3907	CORREL_TOT
				2,4	RESTRICTION

Source: Research data (2020).

3.2 Assembly of the efficient frontier

From the 10 actions obtained, the covariance matrix (table 4) was obtained by Microsoft Excel®, together with the average monthly return on the actions of the collected data, and it was applied to the programming in order to obtain the minimization of the covariance as explained in item 2.2.

Table 4 - Variance / covariance table

	RADL3F	SULA11F	VIVT4F	KLBN11F	PETR4F	FLRY3F	GRND3F	WEGE3F	BBDC4F	CPFE3F
RADL3F	0,000199	0,000064	0,000042	0,000023	0,000010	0,000010	0,000046	0,000031	0,000006	0,000068
SULA11F	0,000064	0,000227	0,000054	0,000022	0,000011	0,000028	0,000043	0,000048	0,000001	0,000073
VIVT4F	0,000042	0,000054	0,000123	0,000026	0,000010	0,000041	0,000039	0,000027	0,000025	0,000047
KLBN11F	0,000023	0,000022	0,000026	0,000255	0,000077	0,000046	0,000040	0,000126	0,000049	0,000003
PETR4F	0,000010	0,000011	0,000010	0,000077	0,000218	0,000059	0,000076	0,000009	0,000073	0,000006
FLRY3F	0,000010	0,000028	0,000041	0,000046	0,000059	0,000182	0,000064	0,000037	0,000053	0,000036
GRND3F	0,000046	0,000043	0,000039	0,000040	0,000076	0,000064	0,000201	0,000041	0,000047	0,000048
WEGE3F	0,000031	0,000048	0,000027	0,000126	0,000009	0,000037	0,000041	0,000233	0,000072	0,000042
BBDC4F	0,000006	0,000001	0,000025	0,000049	0,000073	0,000053	0,000047	0,000072	0,000261	0,000022

CPFE3F	0,000068	0,000073	0,000047	0,000003	0,000006	0,000036	0,000048	0,000042	0,000022	0,000237

Source: Research data (2020).

For the formation of the efficient frontier, it was necessary to establish the intervals for the construction of the graph. Without the restrictions we would only obtain the point of minimum global variance, which in this case would be an estimated return of 5.12% for an estimated risk of 0.70%. The lowest value obtained respecting the restrictions was a return of 2.82% for a risk of 0.80% and the highest value obtained was 7.62% for an equal risk of 0.80%.

Table 5. Result of efficient border points after programming

St.Dev	Ret[r]	RADL3F	SULA11F	VIVT4F	KLBN11F	PETR4F	FLRY3F	GRND3F	WEGE3F	BBDC4F	CPFE3F
0,80%	**7,62%**	20,00%	2,50%	2,50%	7,50%	2,50%	20,00%	20,00%	20,00%	2,50%	2,50%
0,73%	**6,50%**	20,00%	3,80%	8,70%	2,50%	12,70%	11,40%	10,90%	20,00%	2,50%	7,50%
0,71%	**6,00%**	20,00%	3,50%	12,00%	2,50%	13,90%	10,30%	5,90%	20,00%	2,50%	9,30%
0,70%	**5,12%**	16,00%	4,50%	18,20%	2,50%	14,80%	8,00%	2,50%	17,40%	4,40%	11,80%
0,72%	**4,00%**	8,60%	9,10%	20,00%	3,30%	14,70%	5,70%	2,50%	9,50%	10,70%	15,90%
0,77%	**3,00%**	2,50%	13,30%	20,00%	5,50%	13,30%	2,50%	2,50%	2,50%	17,90%	20,00%
0,80%	**2,82%**	2,50%	20,00%	20,00%	2,50%	7,50%	2,50%	2,50%	2,50%	20,00%	20,00%

Source: Research data (2020).

For comparison purposes, naive distribution was considered, that is, the application of the programming used at the efficient frontier was disregarded and the percentage of share participation was the same for all. Considering this, an estimated return of 5.23% was obtained for a risk of 0.74%, represented by the triangle in figure 2.

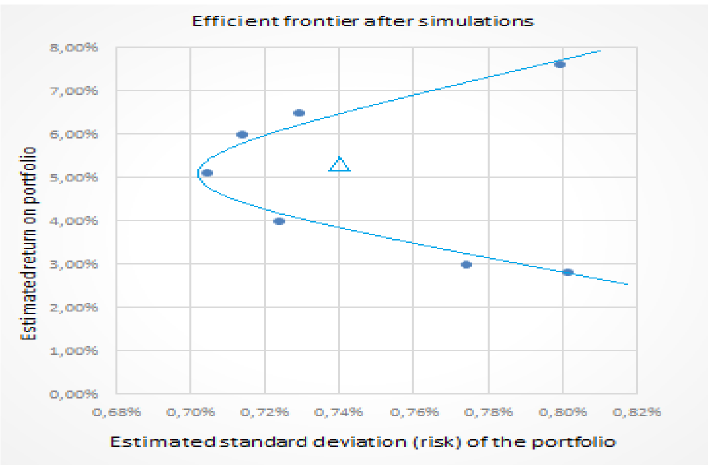

Figure 2 - Illustration of the efficient frontier obtained

Source: Authors (2020).

4. Results and discussion

The period selected for analysis was marked by a sharp rise in the stock exchange due to the momentary animation due to external factors such as the signaling of structural reforms in the country. The percentage performance of 10 shares obtained through binary programming was observed and was compared to a Brazilian investment website portfolio, which for the period analyzed obtained the result of 7.23%, where the naive distribution was used. The application of the Knapsack Problem to obtain a portfolio proved to be a good complementary tool in the analysis of the composition of investment portfolios of variable income. For this, it was necessary a solid fundamentalist and technical analysis of shares for an initial filtering to be carried out on the stocks where the simulations will be carried out. The bold investment profile (Maximum return) was the one with the highest return (8.72%) with a result 14.43% higher than estimated. The conservative profile (Minimum global variance) obtained a return of 7.33% (43.16% higher than initially estimated). And the naive distribution, which corresponds to an equal allocation between the shares (without the need to carry out the LP) obtained a result of 7.87% (50.47% higher than initially estimated). Table 2 presents the summary of estimated results versus results obtained by investment profiles.

Table 2. Summary of estimated versus obtained results from investment profiles.

Invest. profile	Stdev estimated	Return Estimated	Result obtained	Reached the goal? (7.23%)
Bold (Max. return)	0.80%	7.62%	8.72%	YES
Global minimum	0.70%	5.12%	7.33%	YES
Naive distribution	0.74%	5.23%	7.87%	YES

Source: Research data (2020).

Preliminary binary programming proved to be a useful tool in the quantitative formation of the portfolio, since all the arrangements made via PL to obtain the efficient frontier (global minimum and maximum return), as well as the naive distribution, exceeded the result obtained through the Valor.com portfolio in December 2019 (7.23%).

The naive diversification between roles proved to be more effective in the estimated versus obtained item than the allocations for the minimum global variance (conservative profile) and maximum return (bold profile). One of the causes is that a greater allocation in a single asset can increase the portfolio's weight and risk in face of non-diversifiable risks, in this way that allocation via the efficient frontier has great theoretical utility, but not necessarily practical.

5. References

[1]Cibulskienė, D. and Grigaliūnienė, Z. (2007). The Genesis and Development of Modern Portfolio Theory, Economics and Management: Current Issues and Perspectives, p. 52–61, Šiauliai University.

[2]Galinienė, B and Stravinskytė, J.(2016). Constructing an optimal investment portfolio for the bank of Lithuania, Ekonomika, 2016, Vol. 95(1).

[3]Gupta, P, Mehlawat, K and Saxena, A.(2010). "A hybrid approach to asset allocation with simultaneous consideration of suitability and optimality," *Information Sciences*, vol. 180, no. 11, pp. 2264–2285, 2010.

[4]Investment and Finance. Portfolio Management. (2020). Available at: https://www.investment-and-finance.net/portfolio-management/e/efficient-frontier.html. Acessado em 30 de julho.

[5]Konno, H and Yamazaki, H. (1991). "Mean-absolute deviation portfolio. Optimization model and its applications to Tokyo stockmarket,". *Management Science*, vol. 37, no. 5, pp. 519–531.

[6]Malaj, E and Malaj, V. (2016). Portfolio Allocation: An Empirical Analysis of Ten American Stocks for the Period 2010-2015. *International Journal of Accounting, Finance and Risk Management.* Vol. 1, No. 1, 2016, pp. 11-18.

[7]Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, vol.7, no. 1, pp. 77–91, 1952.

[8]Morita, H. (1989). Stochastic linear knapsack programming problem and its application to a portfolio selection problem. *European Journal of operational Research*, vol. 3, no. 4, pp. 329-336, 1989.

[9]Sekar, G. (2012). "Portfolio optimization using neuro fuzzy system in Indian stock market," *Journal of Global Research in Computer Science*, vol. 3, no. 4, pp. 44–47, 2012.

[10]Sharpe, W. (1971). A linear programming approximation for the general portfolio.(1971a).
analysis problem. Journal of Financial and Quantitative Analysis 6, 1263–1275.

[11]Sharpe, W.(1971). Mean-absolute deviation characteristic lines for securities and portfolios. Management Science 18, B1–B13

[12]Speranza, M. (1993). "Linear programming models for portfolio optimization," *Finance*, vol. 14, pp. 107–123, 1993..+

Appendix

Table 2 - Stock correlation table and average correlation

–	LREN 3F	MOVI 3F	RADL 3F	SULA1 1F	VIVT4 F	EZTE C3F	JBSS3 F	KLBN 11F	PETR4 F	VVAR 3F	B3SA3 F	LINX3 F	PCAR 4F	BPAC1 1F	FLRY3 F	GRND 3F	LAME 4F	WEGE 3F	BBDC 4F	CPFE3 F	CVCB 3F	VALE3 F	CORRELATION
LREN3	1,000	0,565	0,411	0,389	0,469	0,513	-0,053	0,135	0,270	0,640	0,366	0,214	0,466	0,234	0,339	0,441	0,560	-0,021	0,180	0,476	0,332	0,207	**0,370**
MOVI	0,565	1,000	0,297	0,246	0,390	0,531	0,009	0,217	0,347	0,338	0,341	0,118	0,411	0,217	0,382	0,462	0,408	0,098	0,145	0,333	0,325	0,291	**0,340**
RADL	0,411	0,297	1,000	0,300	0,270	0,294	0,017	0,100	0,050	0,280	0,268	0,125	0,291	-0,005	0,051	0,231	0,364	-0,143	0,027	0,313	0,083	0,180	**0,219**
SULA	0,389	0,246	0,300	1,000	0,320	0,218	0,126	0,092	-0,049	0,374	0,299	0,116	0,311	0,363	0,137	0,203	0,378	0,211	-0,006	0,313	0,068	0,154	**0,253**
VIVT	0,469	0,390	0,270	0,320	1,000	0,261	0,150	0,145	0,059	0,505	0,388	0,166	0,304	0,175	0,274	0,246	0,496	0,162	0,138	0,278	0,001	0,104	**0,286**
EZTE	0,513	0,531	0,294	0,218	0,261	1,000	-0,003	0,136	0,001	0,487	0,346	0,192	0,360	0,281	0,233	0,246	0,278	-0,051	-0,029	0,392	0,204	-0,020	**0,267**
JBSS3	-0,053	0,009	0,017	0,126	0,150	-0,003	1,000	0,069	0,035	0,112	-0,010	0,015	0,141	-0,034	0,099	0,088	0,051	0,103	0,013	-0,036	-0,155	0,088	**0,083**
KLBN	0,135	0,217	0,100	0,092	0,145	0,136	0,069	1,000	0,325	0,146	0,352	0,194	0,104	0,209	0,212	0,179	0,157	0,517	0,192	-0,012	0,187	0,226	**0,222**
PETR	0,270	0,347	0,050	-0,049	0,059	0,001	0,035	0,325	1,000	0,018	0,238	0,206	0,166	0,045	0,295	0,365	0,144	0,042	0,306	0,027	0,318	0,273	**0,204**
VVAR	0,640	0,338	0,280	0,374	0,505	0,487	0,112	0,146	0,018	1,000	0,343	0,209	0,324	0,235	0,342	0,225	0,556	0,111	0,127	0,448	0,300	0,170	**0,331**
B3SA	0,366	0,341	0,268	0,299	0,388	0,346	-0,010	0,352	0,238	0,343	1,000	0,445	0,293	0,334	0,248	0,223	0,473	0,422	0,343	0,254	0,233	0,011	**0,328**
LINX	0,214	0,118	0,125	0,116	0,166	0,192	0,015	0,194	0,206	0,209	0,445	1,000	0,230	0,153	0,005	0,214	0,410	0,192	0,166	0,162	0,239	0,037	**0,219**
PCAR	0,466	0,411	0,291	0,311	0,304	0,360	0,141	0,104	0,166	0,324	0,293	0,230	1,000	0,236	0,154	0,543	0,482	0,013	0,086	0,361	0,322	0,240	**0,311**
BPAC	0,234	0,217	-0,005	0,363	0,175	0,281	-0,034	0,209	0,045	0,235	0,334	0,153	0,236	1,000	0,272	0,313	0,286	0,396	0,115	0,206	0,195	-0,017	**0,237**
FLRY	0,339	0,382	0,051	0,137	0,274	0,233	0,099	0,212	0,295	0,342	0,248	0,005	0,154	0,272	1,000	0,334	0,313	0,181	0,243	0,173	0,326	0,376	**0,272**
GRN	0,441	0,462	0,231	0,203	0,246	0,246	0,088	0,179	0,365	0,225	0,223	0,214	0,543	0,313	0,334	1,000	0,332	0,191	0,204	0,221	0,312	0,227	**0,309**
LAM	0,560	0,408	0,364	0,378	0,496	0,278	0,051	0,157	0,144	0,556	0,473	0,410	0,482	0,286	0,313	0,332	1,000	0,203	0,263	0,218	0,440	0,057	**0,358**
WEG	-0,021	0,098	-0,143	0,211	0,162	-0,051	0,103	0,517	0,042	0,111	0,422	0,192	0,013	0,396	0,181	0,191	0,203	1,000	0,291	-0,180	-0,026	-0,057	**0,166**
BBDC	0,180	0,145	0,027	-0,006	0,138	-0,029	0,013	0,192	0,306	0,127	0,343	0,166	0,086	0,115	0,243	0,204	0,263	0,291	1,000	0,089	0,247	0,377	**0,205**
CPFE	0,476	0,333	0,313	0,313	0,278	0,392	-0,036	-0,012	0,027	0,448	0,254	0,162	0,361	0,206	0,173	0,221	0,218	-0,180	0,089	1,000	0,325	0,233	**0,254**
CVCB	0,332	0,325	0,083	0,068	0,001	0,204	-0,155	0,187	0,318	0,300	0,233	0,239	0,322	0,195	0,326	0,312	0,440	-0,026	0,247	0,325	1,000	0,271	**0,252**
VALE	0,207	0,291	0,180	0,154	0,104	-0,020	0,088	0,226	0,273	0,170	0,011	0,037	0,240	-0,017	0,376	0,227	0,057	-0,057	0,377	0,233	0,271	1,000	**0,201**

Source: Research data (2020).