Experimental Analysis of Diversification Effect on Stock Portfolios

Track: Financial Markets, Investment and Risk Management

Keywords: Portfolio Size, Risk, Diversification

Experimental Analysis of Diversification Effect on Stock Portfolios

Abstract

In the present work we analyze the relationship between risk and individual investors' stock portfolios when shares are chosen from a naive strategy. Following Chance, Skynkevich & Yang (2011) procedure, participants have successively chosen shares to compound their individual portfolio. We have found negative exponential relationship between risk and the number of shares in portfolios. The negative exponential relation occurs in average, but not for most of individuals, who "undiversify" when adding shares. Few shares only diversify a portfolio in a wide sample, and it may not apply to every person.

1. Introduction

How many shares are needed to keep a portfolio diversified? It is important to assertively comprehend the diversification mechanism, once it is known that investors who own higher risk portfolios expect higher returns than traditional investors. It is then relevant to understand this mechanism and characterize the relation between risk and portfolio size, given its importance and utility. Evans & Archer (1968), Fisher & Loire (1970), Elton & Gruber (1977), among others, suggest that up to 20 assets may let a portfolio well diversified. However, there are studies suggesting greater amount of assets for a well diversified portfolio (e.g. Statman (1987), Byrne & Lee (2000), Lee (2005), Bennet & Sias (2011)).

During the selection of assets that will compose the portfolio, as in most of complex tasks, a simple rule may be considered for this allocation. One of those simple rules is named naive diversification ($\frac{1}{n}$). According to the diversification heuristic (BENARTZI & THALER, 2001), some agents have uniformly distributed their resources within the investment options. Throughout this work, the term "naive" and its derivatives are used to represent the means by which shares are chosen by individuals. In this sense, "naive" refers to the selection with no analysis or methodology. The term "random", in turn, usually refers to shares selected by a random number generator.

Towards the context of investment portfolio formation, portfolio diversification, and the use of naive strategy for choosing shares, we have investigated the relation between risk and the number of shares in a portfolio for the individual investor when the naive strategy is used on forming investment portfolios. The methodology of Chance, Shynkevich & Yang (2011) was employed, in which experiments were set forth to investigate if the relation between portfolio risk and number of assets is exponentially negative. We extend Chance, Shynkevich & Yang (2011) work by drawing extended analysis.

2. Portfolios diversification

Starting from the traditional risk measurement in portfolios (MARKOWITZ, 1952), we know that:

$$s_p{}^2(n)^{\square}=\sum_{i=1}^{n} w_i{}^2 s_i{}^2+\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j s_{ij} \tag{1}$$

Where $i \neq j$.

In case of equal weightings, we have:

$$s_p{}^2(n)^{\square}=\left(\frac{1}{n}\right)^2 \sum_{i=1}^{N} s_i{}^2+\left(\frac{1}{n}\right)^2 \sum_{i=1}^{N}\sum_{j=1}^{N} s_{ij}$$

$$s_p{}^2(n)^{\square}=\left(\frac{1}{n}\right)^2\left(\sum_{i=1}^{N} s_i{}^2+\sum_{i=1}^{N}\sum_{j=1}^{N} s_{ij}\right)$$

$$s_p(n)^{\square}=\sqrt{\left(\frac{1}{n}\right)^2\left(\sum_{i=1}^{N} s_i{}^2+\sum_{i=1}^{N}\sum_{j=1}^{N} s_{ij}\right)} \tag{2}$$

Where $i \neq j$.

Mean variance ($\acute{s}_i^2$) is defined as:

$$\acute{s}_i{}^2=\sum_{i=1}^{N}\left(\frac{s_i^2}{n}\right) \tag{3}$$

And the mean covariance ($\acute{s}_{ij}$) as:

$$\acute{s}_{ij}=\frac{\left(\sum_{i=1}^{N}\sum_{j=1}^{N} s_{ij}\right)}{n(n-1)} \tag{4}$$

Where $i \neq j$.

While assets are added and reallocated maintaining equal weightings, n raises. The limit of $s^2_p(n)$ is considered to obtain:

$$s_p^2(n)^{\square} = ¿\, \acute{s}_{ij}$$
$$\lim_{n \to \infty} ¿ \quad (5)$$

As the number of shares increase, the contribution of variance terms decreases, remaining only the mean covariance. Thus, the portfolio variance converges to the mean variance. The issue regarding these usually common results occurs from the fact that this convergence is a characteristic of large samples. The law of large numbers will be applied and the mean variance and covariance will not differ too much from a portfolio to another. However, for a short sample, such as a portfolio chosen by a unique investor, the law of large numbers would not be applicable. The mean variance and covariance are n functions, that is, $s_i^2(n)$ and $s_{ij}(n)$, so that the used general properties of the limit do not apply. In fact, it is not possible to evaluate the general limit, once the relation between mean variance or covariance and n will vary of n for n+1 assets.

Being the relation between portfolio risk and its asset numbers supposedly asymptotic, it is not possible to define an inferior risk limit while asset units may be added. Besides, prior researches focus on cases when the portfolio is equally weighed, that is, allocates assets in a stock portfolio in the same proportion. In practice, real investors may utilize other ponderation schemes. The capital market theory emphasizes great ponderation schemes that minimize risk for a given level of expected return.

Therefore, the risk in a portfolio depends on the proportion of individual shares, its variances and covariances. For cases when assets are related randomly and combined in equal proportions in the portfolio, the risk declines when the number of different assets increases in the portfolio. But after all, how many shares diversify a portfolio? Evan and Archer (1968), for example, have concluded that approximately 10 shares well diversify the portfolio and their result has been widely cited. Statman (1987), in a comprehensive analysis on the theme, relates results that are in accordance to Evans and Archer (1968), with suggested portfolios containing not less than 8 shares and not more than 16.

Elton & Gruber (1977) have investigated the relation between risk and the number of assets in a portfolio, and presented an analytics solution for it. Their results show that 51% of the portfolio's standard deviation is limited after the increment from 1 to 10 assets. Adding ten more assets, (meaning, increasing from 10 to 20), only 5% of standard deviation is reduced. The addition of 10 more (meaning, from 20 to 30) there is a 2% reduction in the standard deviation. It is notable that the diversification effect diminishes when the number of assets increases.

Newbould & Poon (1993) have researched for a number of texts produced in the USA and academic studies that would affirm that a portfolio containing from 8 to 20 shares would be a diversified portfolio; Fisher & Lorie (1970) have concluded that the potential for risk reduction through the increase of portfolio quickly exhausts – they have noted that around 80% of risk reduction potential is reached by a portfolio composed by 8 assets; Bloomfield et al (1977) have found that a portfolio with 20 assets reaches a wide fraction of the total benefits.

On Lee's studies (2005), the number of assets that should be kept for reaching a good diversification in the portfolio of security assets is a puzzle, since the estimated number is considerably higher than the numbers seen in current portfolios. Basing on Statman's point of view (1987), applied to the United Kingdom, it was found that the marginal benefits of security portfolios are so small that investors are, probably, conscious of owning a small portfolio, at least as the reduction of standard deviation. Due to high costs related to the management of a security portfolio, the investors might be acting rationally by keeping a low number of assets.

Few authors have been alerting on the traps of a precipitate conclusion when interpreting risk diagrams and the size of the portfolio. In particular, Tole (1982) highlights the misrepresentation nature of the "mean effect". He disagrees on the used technique of the portfolio size against a "mean" measure of non-systematic risk, aiming to determine the adequate levels of diversification. For him, despite this mean calculation effect statistically foresees regressions relatively simple, its applications may be deceptive to investors who base themselves in studies like that. He has shown, also, that the adequate measurement of diversification does not come from the regression line, but from the dispersion around that line.

Byrne & Lee (2000) have reported similar results for real estate portfolios in the United Kingdom, from the risk analysis of a wide sample, with data from 1981 to 1996. Their goal was to debate on the advantages of developed portfolios in relation to the risk levels in security portfolios, with empiric evidence based on real portfolios. Results show that all that can be said is that large portfolios, in average, tend to have lower risk than small portfolios. Within their results, Byrne & Lee (2000) affirmed that, despite the mean risk to decrease rapidly, the variability around the average decreases in a much lower rate. Authors have concluded that the individual investor who follows the rule based on results of medium portfolios may expose themselves to a higher risk that what is intended; and that the recommendations of around 20 to 40 assets to diversify a portfolio would seem an underestimation of the real number of needed assets.

The fact that almost non-systematic risk is eliminated when the portfolio has among 10 to 100 assets does not show much significance when presented by itself. Following other approaches, diversification must be developed while marginal benefits exceed the costs. Benefits are risk reduction; costs are transaction costs. The usual argument that justifies the limited diversification is the fact that the marginal costs increase faster than the benefits.

An example is the work of Statman (1987), who assumed that the investor chooses randomly the assets to form his portfolio with different numbers of assets, but with identical expected returns. *Ibbotson Associates*' results were used on the risk prize in the case of "reference portfolio", the 500 S&P index, with different weights. It was assumed that the maintenance cost for equal weights is the same that for cases in where the ponderation is different. For him, a well diversified portfolio of assets randomly chosen must include at least 30 assets for an interested investor and 40 for a borrower. This idea contradicts what is widely accepted, that is, the diversification benefits are virtually depleted when the portfolio has approximately 10 assets. Besides, individual observations show that people do not keep healthy portfolios.

Mayshar (1979) has developed a model punctuating that the best to be done is to limit the diversification in the presence of transaction costs. The author shows that, contrary than results obtained up to that moment, transaction fixed costs relatively small may actually result in a substantial restriction in the number of acquired shares. And concludes that transaction and administrative costs do matter, once its assimilation in a simple market model has resulted in substantially different implications from what was obtained from standard-models, where these costs did not exist. Considering that the diversification process has a cost, it is highlighted that the brokerage expenses are inversely proportional to the volumes negotiated. Thus, transaction costs justify the fact that investors have a limited number of shares.

Sanvicente & Bellato (2004) have enumerated other results regarding the number of assets necessary to form a diversified portfolio. Brito & Sancovschi (1980) have certified through daily stock quotations negotiated in Rio de Janeiro Stock Exchange (between 1973 and 1979) that the greatest part of diversification gains may be obtained with a portfolio containing 8 shares, and that for portfolios with more than 15 shares, benefits are almost despicable.

Sanvicente & Bellato (2004) have determined the number of shares necessary to make a diversified portfolio for the Brazilian Stock Market, considering transaction costs of an imperfect capital market. The difference between bid-ask spreads for institutional and individual investors, as Statman did. It was noticed that his result has not differed much from prior Brazilian results. Despite considering market imperfections, this effect was compensated by the high transaction costs from the Brazilian capital market. He has inferred that in the Brazilian capital market a small diversification of portfolios would be done due to transaction costs. If such costs decrease, the Brazilian result would be closer to the North American. Consequently, it is possible the Brazilian investors would be suffering regulamentary interventions, besides forming portfolios beyond a great level of diversification.

Still considering Statman's methodology (1987), the work from Oliveira & Paula (2008), which had as objective determining the number of shares that could diversify Bovespa portfolios of home brokers. This number is found matching the benefit if including one more asset into the portfolio. For that, the benefit was calculated similar to Statman (1987) and the cost was calculated by an original methodology, based on the ponderation of cost function of home brokers by the volume negotiated by each of them. The result was 12.

Bennet & Sias (2011) have enlightened misunderstandings regarding the non-systematic risk diversification. They affirm that there are no evidences that investors may or were already capable of forming portfolios with negligible exposition for systematic returns. As the well diversified portfolios are the base under which most of the financial theory is built, the incapacity of investors in forming it helps on explaining the persistence of anomalies and the possibility of asset price bubbles. Considering that results from Evans & Archer (1968) have become conventional acknowledgement, Bennett & Sias (2011) have disagreed from it. Instead of supporting portfolios of 20 (or 30, or 50) assets are well diversified, they have demonstrated that portfolios with up to 200 (or 3000, or 500) assets have substantial non-systematic risk, that is, the uncertainty not negligible on returns of a specific company.

The explanation for this difference is simple – in most part it results from an ordinary interpretation. It commonly divides de total risk into systematic and non-systematic components. As a result, despite being labeled, the non-systematic risk is not shown. While the variance of the total return is the sum of variations of systematic and non-systematic return, the standard deviation of a portfolio is not the sum of its systematic and non-systematic standard deviations. The non-systematic risk is not equal to the difference between the standard deviation of the total return and the standard deviation of the systematic return.

Conventional knowledge and several didactic books on finance affirm that investors may easily form well diversified portfolios with a relatively low number of assets – estimations vary from 8 to 30 shares, although recent estimations suggest at least 50 shares. However, for Bennett & Sias (2011), conventional knowledge is wrong. By definition, a portfolio is well diversified only when the investor is sure that the non-systematic return will be significantly different than zero or, equivalently, when the uncertainty about the non-systematic return is approximately zero. There are no evidences that investors may, or have been capable of, forming these portfolios.

Campbell et al (2011) examine how the diversification process has changed throughout time, observing that correlations have decreased, what suggests that diversification may be reached with less shares, but the idiosyncratic risk has been increased throughout time, being easier to diversify with a small amount of unities. For Chance, Shynkevich & Yang (2011), there is a vast literature that describes the known exponential decline, leading to the conclusion that diversification passes fast.

3. Methodology and discussion

Following Chance, Shynkevich & Yang (2011), we have conducted an experiment counting with the participation of 126 UFRN undergraduate students. Among them, 49.21% are male, 7.14% have experience with the stock market and 70.63% have a professional experience (in any area). The interviewed participants are students who have been through finance. The research tool was a questionnaire composed by closed questions. Its goal was to collect descriptive information (if the participant had experience in the stock market, for example), and preferential (what shares he would choose to compose his portfolios). The questionnaires were applied in the classrooms of UFRN’s Management course in May 2012 and each last, in average, twenty minutes per group. The questionnaire was filled after a short explanation about risk and the number of shares in a portfolio.

A material about the relation between risk and the number of shares was presented to the students, including the graphic presentation of this function, of exponential and declining format. After it, it was explained that they would participate on an experiment to verify if these results would stand for the mean portfolio of the class. The worksheet created had a menu listing all shares negotiated in BM&FBOVESPA from the five past years. Students were instructed to fill up the list, which had only the names of shares in alphabetical order, with no financial information.

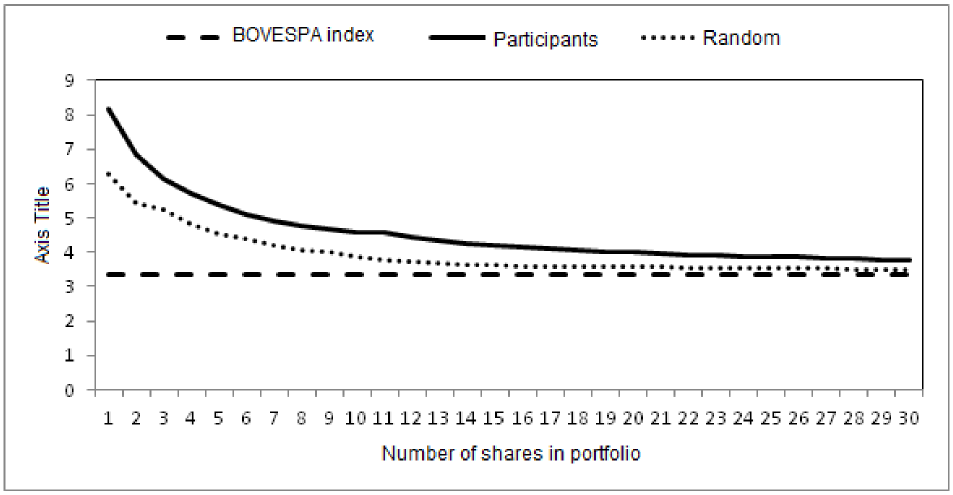
They were also informed that the risk of the first asset chosen would be the average of all standard deviations of portfolios composed by only one asset, and the same procedure would be used to portfolios composed by two units, three shares, and so on until 30 shares. Their only task was to choose shares that they wanted in their portfolios. It was said that no financial analysis would be necessary, and that they should not work in teams, to avoid any bias. Considering the simple task and the small return to participants, there was a strong disincentive for doing any type of analysis. For comparison, a simulation was also done, synthesized by the R software, that selected 126 vectors (each one composed by 30 shares chosen randomly). As each vector corresponds to one participant (virtual), each real participant may be compared to a hypothetic and random investor (component of the control group) facing the same opportunity of historic data. Returns and dispersion measures were taken based on Bloomberg data.

Figure 1 shows the standard deviation of daily returns of the sample in relation to the number of assets, both for participants and for the random portfolios. The horizontal line represents the risk of BOVESPA index that was taken as a proxy for all universe of available assets, if not the market portfolio. The known exponential decline of the relation risk and number of assets is evident for both participants and random portfolios and represents, in this case, the behavior of diversified risk. While the addition of assets for portfolios greater that 20 contribute a little for risk reduction, it is interesting to observe that the risk of a 30-shares portfolio is for both groups beyond the market risk.

For this database, the relation of random portfolios is lower than the participants', what indicate a lower diversifying risk in the composition of random portfolios, for any size. The horizontal line is the standard deviation from BOVESPA index, representing the systematic risk. The total risk would be represented by the sum of systematic and diversifying risks. Thus, the total risk of the portfolio of participants is higher than the risk of random portfolios. Besides, it can be understood from Figure 1 that for both cases, there is few diversification gain after the addition of a fifteenth asset.

Figure 1 – Relation between portfolio mean risk and the number of assets for participants and random portfolios - Through standard deviation of expected returns, it was found risk values for each portfolio size (from 1 to 30), for each

participant. The curves of this Figure represent the mean diversifying risk of participants' portfolios (full line) and random portfolios (dotted line), that is, each curve represents the average of 126 participants for each portfolio size.



Regarding the risk analysis of each participant (or the individual analysis), the curves for their components may be classified according to the form. Curves presented in this work may not be characterized as "positively inclined", "negatively inclined" or "not inclined". In here, portfolios of participants were divided into four groups: Predominantly Exponential Decline (PED), Erratic Exponential Decline (EED), Fast Decline followed by Absence of Tendency (FDAT) and others (no tendency; decrease, increase and no tendency; decrease, increase and decrease; exponential decrease followed by linear decrease; irregular convex; etc.). The classification of these curves is subjective, being this one of the several manners of analyzing patterns of that relation for individual participants. It represents the patterns seen in relations between the portfolio's standard deviation and the number of assets for the 126 participants who selected 30 assets sequentially.

Figures 2 to 5 demonstrate examples of Figures of each group. In Figure 2, it can be seen a representation type "Predominantly Exponential Decline", which is similar to the exponential decline seen for the average of participants calculated as a whole. This image does not show the suavity almost perfect of the global group, but is close to it. This type of curve represents 43.65% of the 126 curves.

Figure 2 – Predominantly Exponential Decline - The curve represents the relation between the risk and the portfolio size of participant 31, suggesting, therefore, and individual relation, calculated by the standard deviation method of expected returns. The curve of this participant presents a negative exponential format.

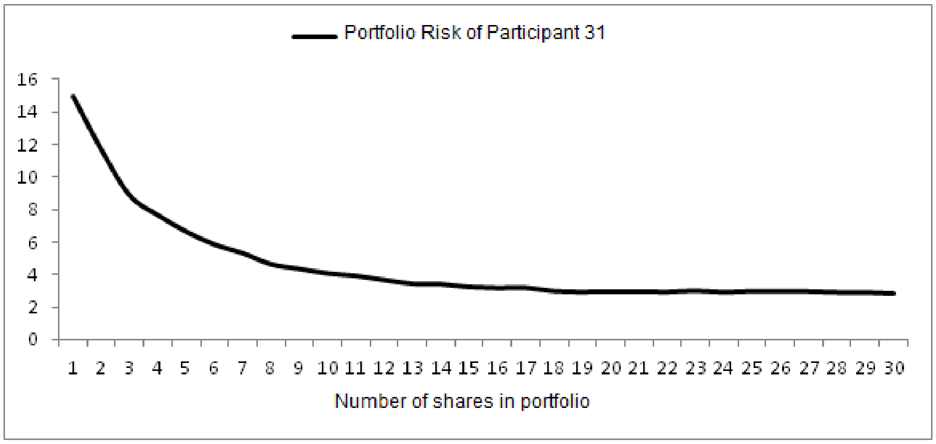


Figure 3 shows an Erratic Exponential Decline curve, a classification that represents 34.92% of curves. Despite relating to studies variables exponentially, the function presents more irregularities than the predominantly exponential.

Figure 3 - Erratic Exponential Decline - The curve represents the relation between the risk and the portfolio size of participant 3, suggesting, therefore, and individual relation, calculated by the standard deviation method of expected returns. The curve of this participant presents an erratic exponential format.

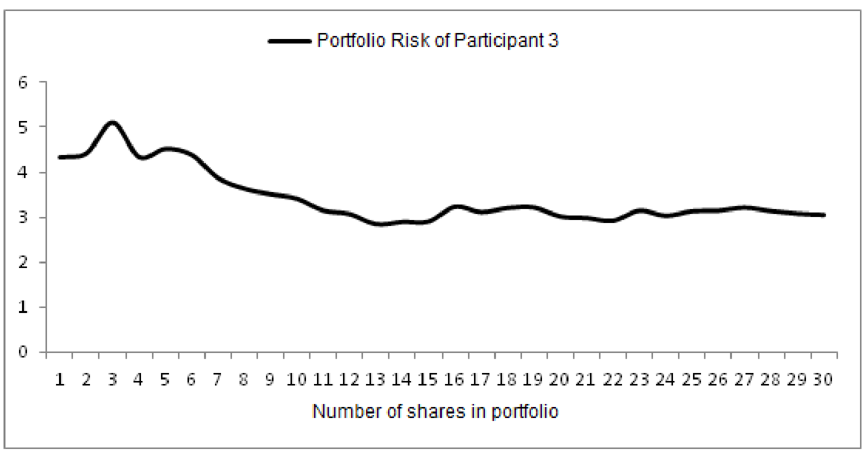


Figure 4 is a "fast decline and no tendency curve". Almost 10% of curves are classified into this category, which means that the portfolio risk does not depend on the number of assets that compounds it.

Figure 4 – Fast Decline followed by Absence of Tendency - The curve represents the relation between the risk and the portfolio size of participant 69, suggesting, therefore, and individual relation, calculated by the standard deviation method of expected returns. The curve of this participant presents a fast decline and, after it, shows no tendency.

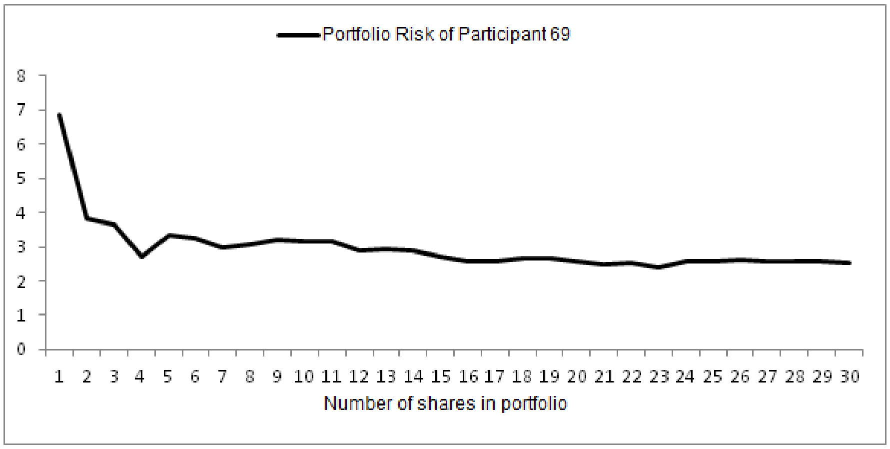


Figure 5 is an example of the "others" category, which portfolio risk may have several behavior less frequent for this sample. I this group there was many profiles, such as "no tendency" or "irregular convex", for example.

Figure 5 - Others - The curve represents the relation between the risk and the portfolio size of participant 121, suggesting, therefore, and individual relation, calculated by the standard deviation method of expected returns. The curve related to this participant shows a low frequency format.

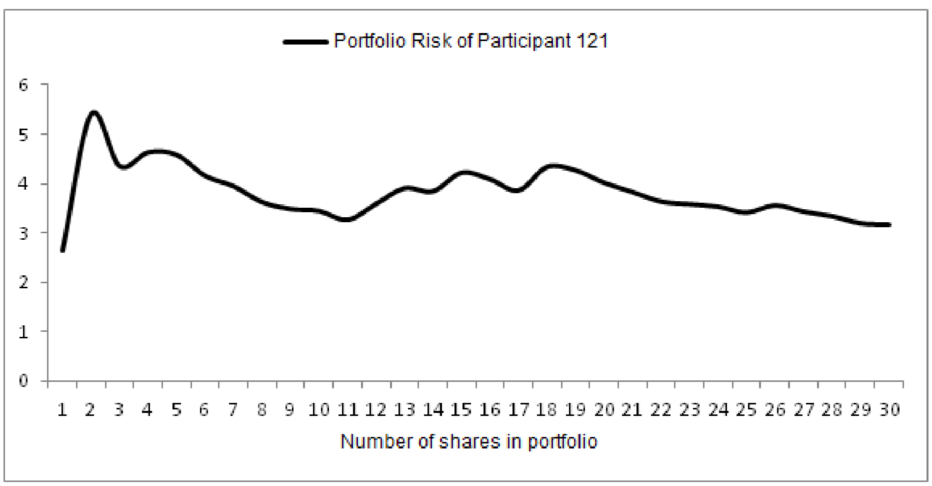


Table 1 represents the results of portfolio classification of participants. In it, in less than half (43.65%) of cases, the relation between portfolio size and its risk is classified as "Predominantly Exponential Decline", which is the classification that fits the most the pattern for all grouped participants and is the expected result for the grouped case. In approximately one third of cases (43.92%), curves may be characterized as "Erratic Exponential Decline". The third group with high frequency is the "Fast Decline followed by Absence of Tendency", with 9.52%. The group "Others" represents 11.90% of cases. Thus, curves that are not in accordance with the expected pattern totalize 56.34% (34.92% + 9.52% + 11.90%) of participants' curves.

Table 1 – Classification of individual participants' curve formats - Classifications represent patterns seen between standard deviation and share number from the 126 participants, who chose 30 shares sequentially. Returns are daily based, considering a period of five years prior to this study.

Classification	Quantity of participants	Percentages
PED	55	43.65
EED	44	34.92
FDAT	12	9.52
Others	15	11.90
Total	126	100

Certainly the groupings are a bit arbitrary, but they show that irregular patterns that do not follow the norm are not uncommon. For obtaining a more objective measure of the curves behavior, another frequently used procedure was used for the present research. A curve will be adjusted representing the exponential decline for each participant through the following relation:

$$s_p(n) = \alpha + \beta \left(\frac{1}{n}\right) + \varepsilon_n \tag{6}$$

That is, the standard deviation (or risk) of the portfolio ($s_p(n)$¿ is receded by the inverse of shares number ($\left(\frac{1}{n}\right)$¿ . This regression was estimated for each of the 126 real participants and for each of the 126 randomly generated by the software. Each individual regression was done considering that the dependent variable "standard deviation (or risk) of the portfolio" is a function of the independent variable $\left(\frac{1}{n}\right)$, where n is the number of shares. From equation 5, values were obtained referring to risk for each portfolio size (from 1 to 30). Thus, each regression was done using 30 observations.

The independent variable $\left(\frac{1}{n}\right)$ derives from the naive strategy and represents the simplicistic rule of a investor who diversifies his investments in proportion $\left(\frac{1}{n}\right)$, that is, quantities of the invested assets are equal or present the same ponderation. Joint regressions (one concerning the participants and another, the random portfolios) were done considering the dependent variance as being the mean standard deviations (or risks) of the 126 individual cases.

The independent variable, as in cases of individual regressions, is the inverse of shares number that compose the portfolio $\left(\frac{1}{n}\right)$. Descriptive statistics on the estimative of parameters are given in Table 2. In it, it is no surprisingly seen that Alphas are highly significative, resulting from the simple fact that the standard deviation of a portfolio with an infinite number of securities is clearly different than zero. Independently of the assets that compose the portfolio, the risk inherent to the speculative activity is positive for both cases (participants and random portfolios).

Beta analysis is very important once it indicates if there is relation between the inverse of assets numbers and the portfolio risk. When there is this relation, Beta also indicates its level (by the curve inclination), that is, suggests is the risk is flexible or not. Betas tend to present quite high t-statistics and few are not positively significative. Thus, the regression variable explains in most of cases the regressor variable. As an example, only 5 from the 126 cases of participants are not significative and none is negatively significant – in contrast of the 16 cases not significative and 3 negative significances from the random portfolios.

The mean R^2 (from all individual cases) is around 62% and 46% for participants and random portfolios, respectively. However, when R^2 generated by group regression (or joint) is compared, the determination coefficient of random portfolios is superior to the participants' (88% against 75%). It means that for the average matter, random portfolios have a higher prevision power than participants' portfolio.

Table 2 – Statistics on the portfolio variance regression in $\left(\frac{1}{n}\right)$ function, where n is the number of assets per participant and random portfolio - Returns have a daily basis, considering the period of 2007 to 2012. Joint regressions (or of group) are grouped throughout time. In case of individual regressions, there are 30 observations of each of the 126

participants. The regression number is the number of participants, that is, 126. Thus, the results' average of alpha, beta, etc, is obtained. Joint regressions use 126 participants times 30 = 3780 observation portfolios.

Results of Regression				
	Participants		Random	
	Individuals	Group	Individuals	Group
Regressions	126	1	126	1
Alpha	3.90	2.99	3.51	2.76
Standard Deviation of Alpha	0.14	0.11	0.14	0.07
t (Alpha)	32.64	31.71	31.47	41.03
Beta	5.01	6.03	3.34	4.24
Standard Deviation of Beta	0.60	0.50	0.59	0.29
T (Beta)	9.75	16.02	7.58	14.61
Significative and positive Beta	110	1	90	1
Beta not significative	15	0	28	0
Significative and negative Beta	1	0	8	0
R^2	0.62	0.75	0.46	0.88

As for the random portfolios, it is seen that more than half of cases (52.38%) R^2 values are higher than 60%. It means that the portfolio risk variability is less explained by the inverse of asset numbers' variability than the participants'' case. Thus, while many regressions adjust themselves to data, less than half of cases fit perfectly to the yield exponential model. Random portfolios show a tendency a bit lower to adjust to the yield curve, once 107 individual cases are significantly positive, *vis-à-vis* 121 of participants' portfolios. Portfolios of participants contain fewer regressions with low R^2, as an example, 7.14% of R^2 lower than 20%, when for the random portfolios there were 22.22% shorter than 20%.

In general, it is seen that random portfolios better distribute the frequency of R^2 values. It must be considered that participants, presenting different levels of risk dislike, consequently choose portfolios with different risk levels. Grouping them is an issue, once the variance of a dependent variable is not constant throughout the sample. It can be concluded then that participants' portfolios do not fit into the relation hypothesis, considering that one of the prior requirements for the application of this model does not work.

For keeping the analysis between risk and share number, it is examined how the risk changes when a share is added. Table 4 informs the percentage of times in where the addition of a single share increases risk. In the left column there are numbers from 2 to 30, representing the number of assets after the addition of an asset into the portfolio (in a minimum size of one). Percentage is the proportion of times (in which the 126 participants and their correspondents from random portfolios) that a portfolio size n has a higher risk of a portfolio size n-1. Thus, in the first line, it is seen that when a second asset is added, around 18% of participants' portfolios increases their risk, while 27% of random portfolios also do so. When a third share is added, around 14% of portfolios raise their risk, and around 34% of random portfolios follow this movement. While a fourth share is added, around 24% of portfolios raise their risk, against 24% of random portfolios.

Some of these results must not be that surprising. As an example, if the first chosen share is a too low-risk asset, it is not hard to imagine that a second asset may increase the global risk of the portfolio. Analyzing the bottom part of Table 4, however, it is seen that the addition of a thirtieth share raises the risk in 26.19% of participants' portfolios and around 45.24% of random portfolios. Among 20 to 30 shares, the percentage of participants' portfolios which risk is raised remains reasonable consistent among 25.40% and 37.30%, but the random portfolios case shows a similar variation, although in a higher level, with percentages ranging from 33.33% to 45.24%.

When analyzed in parts, results found in Table 3 present: the addition of the second up to the eleventh share, the participants' mean is 23.81% diversification, while the random group is 31.75%; from the addition of the twelfth up to the twenty first share, the averages are 24.20% and 39.68%, respectively; from the twenty second to the thirtieth, 27.78% and 42.86%. It is noted that the diversification "diversifies" more frequently in cases of random portfolios, besides occurring with more frequency in larger portfolios for both groups.

The portfolio size average for all groups is 25.12% for participants. It means that the addition of a single share raises the risk for participants, in average, one in every four times. For the random portfolios, risk raises in 36.95% of times. For all portfolio sizes, the frequency of risk raise of random portfolios is higher than the participants. Apparently, participants knew how to combine shares to compose their portfolios better than the computer (randomly). In some situations, the diversification difference was 20% more for random portfolios. These results show two interesting points. The first refers to the number of times in that the addition of a single share raises the risk, a particularly notable result for large portfolios, as the ones containing more than 20 shares. The second point is that for portfolios selected randomly, results are the worst.

A portfolio containing more than 29 shares selected by participants raises the risk with the addition of a thirtieth share almost 26.19% of times, and a portfolio of 29 shares selected by a number generator raises the risk, with the addition of a 30th share, 45.24% of times. There is, certainly, more random noise than the probably known in prior studies. Computers make a slightly better job than people, but not much. Results from Table 3 reflect only the addition of a single share, which may be a higher risk than the combined risk of shares that were already in the portfolio. Adding multiple assets may reduce the noise related to the addition of a single share (which could have raised the idiosyncratic risk), according to Chance, Shynkevich and Yang (2011).

Table 3 – Frequency of the risk raise deriving from the addition of a share in a portfolio, according to participants and the random portfolios - Percentage values represent the relative frequency of diversification, that is, risk raise when the number of shares composing the portfolio also rises. Column "Number of Shares" represents the portfolio size after the addition of one share.

	Percentage	
Number of Shares	Participants	Random
2	18.25	34.92
3	14.29	38.10
4	23.81	32.54
5	16.67	25.40
...	...	...
26	27.78	36.51
27	30.16	38.10
28	37.30	43.65
29	26.19	44.44
30	26.19	45.24
Average	25.12	36.95

Table 4 relates results linked to the addition of several shares. Columns identified as 1, 5, 10, 15, 20 and 25 represent the reference number of securities. Lines identified as 5, 10, 15, 20, 25 and 30 are the final number of assets. Thus, in column 10 with line 25 it is presented a percentage of 3.97%. It means that when adding 15 shares into a portfolio that already has 10 (totalizing, then, 25) the resulting portfolio risk rose in almost 4% of cases.

It is verified that for participants, a portfolio of 30 shares has a higher risk than a portfolio size 1, around 2.38% of times. Said in other words, the risk has risen a bit more than 2% of the times when 29 shares were added in a portfolio. Random portfolios represent 11.11% of times, what means that little more than one in ten cases, the addition of 29 shares into a portfolio of a single share is un-diversified. Once again, it seems that participants have combined assets better than random portfolios.

Table 4 – Frequency of the risk raise derived from the multiple addition of shares in the portfolio, according to participants and random portfolios - Numbers represent the percentage of times in which a portfolio containing a number of assets presented in the columns has a higher risk than a portfolio which contains a number of assets presented in the lines.

	Number of Assets					
Number of Assets	1	5	10	15	20	25

Pannel A: Portfolios of Participants						
5	13.49	NA	NA	NA	NA	NA
10	4.76	15.87	NA	NA	NA	NA
15	3.17	8.73	10.32	NA	NA	NA
20	3.17	5.56	4.76	17.46	NA	NA
25	2.38	3.97	3.97	12.70	21.43	NA
30	2.38	3.17	1.59	9.52	15.87	22.22
Pannel B: Random Portfolios						
5	23.02	NA	NA	NA	NA	NA
10	16.67	15.87	NA	NA	NA	NA
15	13.49	10.32	18.25	NA	NA	NA
20	10.32	10.32	18.25	34.13	NA	NA
25	10.32	9.52	18.25	30.16	35.71	NA
30	11.11	10.32	16.67	27.78	28.57	30.16

Analyzing the column named “1”, it is seen a significant difference between participants and random groups, what means that the addition of shares in a portfolio increases the risk more frequently in cases of random portfolios than of participants. The performance of random portfolios is still worse when the reference is a portfolio of fifteen shares. In this case, the risk raise of random portfolios is at least twice than the participants’. Then, again, a significative percentage of individuals, when forming their portfolios based on the naïve strategy, diversify when adding shares. For random portfolios, results are worse.

When people do not diversify as well as computers, results are normally linked to the lack of randomness. That is, portfolios may not have represented a wide range of industries, and may have offered subtle biases. Even said to a participant to choose from all shares, he would tend to choose those more familiar. Familiarity may arise from current of past employers, friends or relatives, from the exposition of a company by regular consumers, or simply seeing a company in a daily route. Companies with strong local exposition may also generate a bias in a participant. In the case of students from UFRN, as an example, the five most chosen companies had a productive unit in the state of Rio Grande do Norte.

It was also investigated if selections have presented a more subtle bias. The opportunity set was presented in a form of list, alphabetically in an Excel sheet. With around 70 company names, participants did not have a long list of assets to choose from. From Table 5, it is possible to realize that there is no concentration on choosing shares according to the company’s position in the list given with the questionnaire. It means that there is no evidence that a participant would be lazy and would have picket companies from the beginning of the list, for example.

Table 5 – Choice analysis of share according to the position in the list - Values are absolute frequencies of share choices initiating with a certain letter.

Shares	A to E	F to J	K to P	R to U	V to Z	Total
Number of listed shares	31	8	15	13	3	70
Number of times that the share was chosen	1524	472	864	611	289	3760
Proportion (%)	49.16	59	57.60	47	96.33	

One of the main differences between participants and random portfolios is in the general risk level. It means that functions that give a relation type between the risk and the number of shares (for the grouped case) when participants and random portfolios are bought, have similar behaviors. What differs them is the risk volume and not the manner as the risk decreases.

For comparing the risk level in both cases and for each portfolio size, it is needed a statistical test. The first step consists in knowing if data have a normal distribution or not. For that, the Shapiro test for Normality was performed. As p-values of both groups are inferior to the significance level of 5%, the hypothesis of following a normal distribution is rejected. As data does not follow Gaussian distribution, it is recommended the use of a non-parametric test, as for example, the Wilcoxon.

Afterward, with the use of portfolio variance and the mean variance of shares, normalized variances were calculated for each group. For Goetzmann and Kumar (2008), the normalized variance is the variance of a portfolio divided by the mean variance of shares composing a portfolio. The low normalized variance indicates that the assets are well diversified. The percentage difference in normalized variance of participants' portfolios in relation to the normalized variance of random portfolios is the excess of normalized variance. The average for all participants, for each portfolio size, this average gives the excessive normalized variance, where a high value indicates that participants' portfolios are more diversified than random portfolios.

The large bias, however, could be more subtle. When participants add shares, they are able to choose assets of strong correlation to their existing portfolios. It was calculated then the correlation between return of a size n portfolio and the next asset to be added (that will compose the n+1 portfolio). Table 6 presents this information that shows the mean correlation between different size portfolios and the percentage of times in where the correlation is positive.

In more than 70% of cases for participants and random portfolios, the correlation is positive. However, the mean correlation is notably more elevated for participants' portfolios than for random portfolios, a result that does not vary much for different sizes. Analyzing the group of participants in Table 6, it is seen that when the portfolio present a single asset, its correlation with the next asset to be added is, in average, 40% - a moderate correlation. Still considering the single asset portfolio, it was noted that values of this correlation, when analyzed case by case, (that is, for each participant), were positive in 78.47% of times. Random portfolios, for the same portfolio size, shower a mean correlation of 26% and the correlation was higher than zero in 72.22% of cases. For most of portfolio sizes, the mean correlation of participants' portfolio was higher than the random portfolios. The same result is found when the frequency of positive correlations for each group is compared. It may mean that participants have chosen more associating shares.

Table 6 – Correlations of existing portfolios in relation to the marginal asset, according to the participants' portfolios and random portfolios - For an n size portfolio, the correlation between the n portfolio return and the return of the asset to be added into the portfolio n+1. Besides, it is seen the amount of times in where correlation in each line are positive.

	Portfolios of Participants		Random Portfolios	
Portfolio Size	Mean Correlation	% > 0	Mean Correlation	% > 0
1	0.40	78.57	0.26	72.22
2	0.37	80.95	0.34	77.78
3	0.45	84.92	0.34	84.92
4	0.45	88.89	0.40	82.54
5	0.39	80.16	0.39	78.57
6	0.43	90.48	0.42	85.71
7	0.44	89.68	0.35	80.16
8	0.43	89.68	0.43	87.30
9	0.42	87.30	0.39	81.75
10	0.50	94.44	0.44	88.89
11	0.40	84.13	0.44	88.10
12	0.40	87.30	0.39	84.92
13	0.44	89.68	0.42	87.30
14	0.43	89.68	0.46	92.06
15	0.47	93.65	0.46	92.86
16	0.46	93.65	0.48	92.06
17	0.45	89.68	0.45	87.30
18	0.47	91.27	0.49	94.44
19	0.46	95.24	0.48	91.27
20	0.51	96.03	0.48	94.44
21	0.43	89.68	0.42	87.30
22	0.44	91.27	0.45	93.65
23	0.46	94.44	0.48	92.86
24	0.47	92.86	0.47	89.68
25	0.42	94.44	0.45	92.86
26	0.49	90.48	0.49	94.44
27	0.48	91.27	0.50	96.03
28	0.47	94.44	0.48	93.65
29	0.43	90.48	0.51	97.62

Conclusions

People select portfolios by an inconsistent bias. They tend to choose full-sized companies, well known and more correlated to their existing portfolio than the share that may be chosen randomly. Thus, when people try to diversify choosing shares randomly, a subtle bias is set, what may limit their capacity of diversifying with a relatively small number of shares. The negative exponential relation, as documented by Chance, Skynkevich & Yang (2011), occurs in average, but not for most of individuals, who “undiversify” when adding shares. The choice of few shares only diversify a portfolio in a wide sample, and it may not apply to every person. Individual bias must be considered do determine optimal portfolio size.

References

BENARTZI, Shlomo; THALER, Richard H.. Naive Diversification Strategies in Defined Contribution Saving Plans. The American Economic Review, New York, v. 91, n. 1, p.79-98, 01 mar. 2001.

BENNETT, James. A.; SIAS, Richard W. Portfolio Diversification. Journal of Investment Management, Portland, v. 9, n. 3, p.74-98, 2011.

BRITO, Ney Ottoni de; SANCOVSCHI, Moacir. Risco, retorno e betas: o mercado acionário brasileiro. Relatório n. 24, COPPEAD/UFRJ, Rio de Janeiro, 1980.

BYRNE, Peter J.; LEE, Stephen. Risk Reduction in the United Kingdom Property Market. Journal of Property Research, Londres, v. 17, p. 23-46, 2000.

CAMPBELL, John et al. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. The Journal of Finance, Nova York, p. 1-43. 01 fev. 2001.

CHANCE, Don M.; SHYNKEVICH, Andrei; YANG, Tung-Hsiao. Experimental Evidence on Portfolio Size and Diversification: Human Biases in Natıve Security Selection and Portfolio Construction. The Financial Review, n. 46, p.427-457, 2011.

ELTON, Edwin J.; GRUBER, Martin J. Risk Reduction and Portfolio Size: An Analytical Solution. Journal of Business, n. 50, p. 415-437, 1977.

EVANS, John L.; ARCHER, Stephen H. Diversification and the Reduction of Dispersion: an Empirical Analysis. Journal of Finance, v. 23, n. 5, p. 761-767, dez. 1968.

FISHER, L.; LORIE, J. Some studies of variability of returns on investment in common stocks. Journal of Business, v. 47, p. 67-87.

GOETZMANN, William N.; KUMAR, Alok. Equity portfolio diversification. Review of Finance, v. 12, p. 433–463, mar. 2008.

LEE, Stephen. *The marginal benefit of diversification in commercial real estate portfolios.* Working Papers in Real Estate & Planning. University of Reading, Reading, abr, 2005.

MARKOWITZ, Harry. Portfolio Selection. The Journal of Finance, v. 7, n. 1, p. 77-91, mar. 1952.

MAYSHAR, Joram. Transaction Cost in a Model of Capital Market Equilibrium. Journal of Political Economy, n. 87, p. 673-700, ago. 1979.

NEWBOLD, Gerald D.; POON, Percy S. The minimum number of stocks needed for diversification. Financial Practice and Education, v. 3, n. 2, 1993

OLIVEIRA, Fernando Nascimento de; PAULA, Eduardo Lana de. Determinando o Grau Ótimo de Diversificação para Investidores Usuários de Home Brokers. Revista Brasileira de Finanças, v. 6, n. 3, p.437-461, 2008.

SANVICENTE, Antonio Zoratto; BELLATO, Leticia Lancia Noronha. Determinação do grau necessário de diversificação de uma carteira de ações no mercado de capitais brasileiro. In: SEMINÁRIOS DE ADMINISTRAÇÃO, 7., 2002, São Paulo. Anais... São Paulo: FEAC/USP, 2002.

STATMAN, M. How many stocks make a diversified portfolio. Journal of Financial and Quantitative Analysis, n. 22, p. 353–363, 1987.

TOLE, Thomas M. You can’t diversify without diversifying. Journal of Portfolio Management, v. 8, n. 2, p. 5-11, 1982.