Arturo Cifuentes

On the Most Vexing Portfolio Diversification Question: How to Choose N?

www.clapesuc.cl

Documento de Trabajo Nº 114 (junio 2022)

On the Most Vexing Portfolio Diversification Question: How to Choose N?

Arturo Cifuentes
Senior Research Associate, CLAPES UC
Catholic University of Chile, Santiago, Chile
ac@arturocifuentes.com

Abstract

This paper is concerned with estimating the minimum number of positions (N) required to construct a portfolio that achieves an adequate level of risk reduction when compared to holding just one asset of a given universe. The paper proposes a simple but general expression that links the three key variables: (i) the desired level of risk reduction; (ii) N; and (iii) the average correlation within the asset class considered. This expression captures both systematic and non-systematic risk and makes no assumptions regarding the type of asset class or market conditions. The only assumption is that one can estimate the average correlation of returns for the assets considered. The usefulness and validity of this approach are demonstrated with two examples.

Keywords: portfolio diversification; portfolio size; systematic risk; non-systematic risk; risk management; asset correlation; risk reduction

JEL Codes: G11; C63; C65

INTRODUCTION AND MOTIVATION

The merits of portfolio diversification have been promoted at least since the time of the Babylonian Talmud (3rd-6th century). Initially, the case for diversification was based on sound, albeit qualitative, reasoning. It was only in 1952, when Markowitz published his seminal paper, that a solid quantitative argument demonstrating the benefits of diversification was articulated (Markowitz, 1952). The concept has survived well the test of time and today it is widely accepted as one of the bedrocks of modern investment theory. Yet, for all its theoretical merit, there is still a nagging practical matter: there is no agreement in terms of how many assets (N in financial lingo) one must hold to achieve an adequate level of diversification.

The fact that the question remains unsettled has not been for lack of trying: more than 300 academic studies have been published on this topic (most of them in the last twenty years), not to mention the many articles written by market participants (see Figure 1 in Zaimovic et al., 2021). It seems fitting in this context to recall the view of the head of research of a well-known financial institution who opined, in reference to how to choose N, that "the quest for diversification is never ending" (Fragkiskos, 2014). In fact, as of this writing this quest seems to continue unabatedly: a recent study claimed that it is necessary to hold 40-to-50 stocks to reduce diversifiable risk effectively, contradicting the traditional view, in the authors' opinion, that 15-to-20 should suffice (Raju and Agarwalla, 2021).

Notwithstanding the merits of previous studies, they share certain weaknesses. First, they have focused on specific markets (e.g., the Malaysian or U.S. stock market) instead of tackling the problem from a broader, and thus, more conceptual viewpoint. And second, frequently, they have paid too much attention to the non-systematic (diversifiable) risk and have neglected to take into consideration explicitly the systematic (non-diversifiable) risk. We speculate that a direct consequence of these two shortcomings (a point we discuss in detail later) has been the failure to produce an easy-to-use and widely-applicable formula to estimate N while accounting for both, systematic and non-systematic risk. The goal of this paper is to propose a formula for this purpose.

REVIEW OF PREVIOUS WORK

A recent paper by Zaimovic et al. provides a comprehensive review of the relevant literature (Zaimovic et al., 2021). These authors concluded that what constitutes a well-diversified portfolio (in reference to stocks, but the claim can be extended to any asset class) does not have a unique answer for it depends on many factors, namely, the asset class, the risk appetite of the investor, the time-period considered, and the treatment of systematic risk. Nevertheless, it is useful to mention a few studies, just to give a sense of how the treatment of this problem has evolved since Markowitz's 1952 paper.

In 1977 Elton and Gruber presented both, exact and approximate, analytical expressions to calculate the variance of returns of a portfolio of N assets (Elton and Gruber, 1977). They tested their formulas with stock returns data from the 1971-1974 period. However, their approach was rather cumbersome and relied on expressions based on too many parameters to be of practical use. Ten years later Statman did a study using stocks from the S&P 500 and a CAPM-based model and claimed that at least 30 holdings were required to build a well-diversified portfolio (Statman, 1987). This same author argued that his findings challenged the accepted view (his opinion) that ten stocks were sufficient. Later, in 2004, Tang, again, derived an expression for the variance of returns of a portfolio of N assets (Tang, 2004). Yet, he focused then on the diversifiable risk and overlooked the possibility of finding an expression for N. He suggested that 20 positions were required to eliminate about 95% of the diversifiable risk; correlation was not mentioned explicitly in this recommendation. Shawky and Smith analyzed the composition of U.S. mutual funds and concluded that in general they held between 40 to 120 stocks, presumably, a number adequate to achieve a reasonable level of diversification (Shawky and Smith, 2005). Alexeev and Dungey, looked at 2003-2011 returns data based on the S&P 500 companies, and decided that ten positions were good enough to achieve a 90% risk reduction (Alexeev and Dungey, 2015). The year before Zhou voiced a similar opinion (Zhou, 2014). In 2018, de Jong suggested incorporating utility theory into the portfolio diversification conundrum. However, she did not demonstrate how this concept could be inserted into a real-life investment decision (de Jong, 2018). Habibah et al., analyzing the Pakistani stock market, recommended holding 20 securities (Habibah et al., 2018). Again, although they presented an expression for the portfolio variance of returns, they did not progress as far as re-working the formula to arrive at an estimate of N. More recently, and somehow surprisingly, a number of authors have taken the expression for the variance of portfolio returns as the basis of their analyses, but unfortunately, have failed to push it further to arrive at an expression linking N, risk reduction, and both systematic and non-systematic risk (Norsiman et al., 2019; Adamiec and Cernauskas, 2019; Haensly, 2020). These authors approached the problem based on data from specific markets (Malaysia, S&P 500, and, U.S. Large Cap index) and using different techniques arrived at estimates that could be as low as 25 or as high as 300 depending on whether systematic risk was considered.

In summary, most previous studies have focused on estimating the potential risk reduction one can achieve (often neglecting systematic risk) in some specific market and for a given value of N. But have ignored the possibility of arriving at a general expression for N, that took into account both, systematic and non-systematic risks, and that could be used with any asset class.

The following clarification is needed at this point: in recent years the standard deviation of returns has been losing popularity as the risk metric of choice and an increasing number of practitioners have embraced the VaR and/or the CVaR. The reason is that both, the VaR and the CVaR, capture losses, something more useful than volatility in the context of risk management. Nevertheless, since in this study the focus is really relative risk reduction when the comparison is made with a portfolio of one asset, instead of estimating absolute losses, we will use the standard deviation in what follows. This is also consistent with the approach taken by the majority of the studies already mentioned.

ANALYTICAL APPROACH

A more formal statement of the problem is required at this point. Assume an investor has access to a large universe of assets whose returns, and standard deviations of returns, are, on average, μ and σ. Further assume that their corresponding correlations, on average, are equal to ρ. In short, we are assuming that three parameters – (μ, σ, ρ) – are sufficient to describe the key features of the universe of potential assets. (Accordingly, the average covariance can be expressed as $\rho\,\sigma^2$.)

The problem is: what is the minimum-size portfolio (N) such that this portfolio can achieve a reasonable level of diversification? Reasonable, of course, is an undefined term which depends on the investor's risk profile. Clearly, if the investor holds N assets his/her expected return will still be μ, but the corresponding standard deviation (σ_P), will be lower than σ. Hence, we seek to estimate how much reduction in risk, measured as σ_P / σ, we can achieve for a given N. We should note that this formulation is valid for any type of asset class, not only stocks. Finally, we assume that the investor will hold an equally-weighted portfolio (also known as 1/N portfolio). This assumption simplifies the exposition without losing generality.

The expression for the variance of returns of the portfolio, VarP, is

$$VarP = \sum_{i=1}^{N}\sum_{j=1}^{N} \left(\frac{1}{N}\right)^2 \sigma_{ij} \qquad [1]$$

where $\sigma_{i\,j}$ denotes the different entries in the variance/covariance matrix.

After some algebra this expression can be reorganized as follows,

$$VarP = \sum_{i=1}^{N}\left(\frac{1}{N}\right)^2 \sigma_{ii} \quad + \quad \sum_{i-1}^{N}\sum_{j\neq i}^{N}\left(\frac{1}{N}\right)^2 \sigma_{ij} \qquad [2]$$

which leads to (note that $\sigma_{ii} = \sigma_i^2$)

$$VarP = \left(\frac{1}{N}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2\right) + \left(1-\frac{1}{N}\right)\left(\frac{1}{N^2-N}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sigma_{ij}\right) \qquad [3]$$

or, more succinctly

$$VarP = \underbrace{\left(\frac{1}{N}\right) A}_{\substack{\textbf{non-systematic risk}\\ \textbf{(diversifiable)}}} + \underbrace{\left(1-\frac{1}{N}\right) B}_{\substack{\textbf{systematic risk}\\ \textbf{(non-diversifiable)}}} \qquad [4]$$

where

$$A = \left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2\right) \qquad [5]$$

and

$$B = \left(\frac{1}{N^2-N}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sigma_{ij}\right) \qquad [6]$$

The first term in Eq.[4] represents the non-systematic risk. The factor A (see Eq.[5]), is the average asset variance, which in this case is simply σ^2. The non-systematic risk clearly becomes negligible if N is large enough. It is common to find in papers dealing with this subject the familiar plot which shows the non-systematic risk approaching zero asymptotically as N grows (e.g., Solnik, 1974; Norsiman, 2019; Haensly, 2020).

The second term in Eq.[4] represents systematic risk. Note that 1 – 1/N approaches 1 as N grows. Also notice that the factor B (see Eq.[6]) is the average covariance, thus, it can be expressed as $\rho\ \sigma^2$. Hence, an approximate expression for VarP is

$$VarP \approx \left(\frac{1}{N}\right)\ \sigma^2\ +\ \rho\ \ \sigma^2\ =\ \sigma^2\left(\ \frac{1}{N} + \rho\ \right) \qquad [7]$$

which leads to

$$\sigma_P \approx\ \ \sigma\sqrt{\frac{1}{N} + \rho} \qquad [8]$$

and since we are interested in the relative risk reduction with respect to a portfolio of one asset (that is, σ_P / σ), we take Ψ, defined as follows

$$\Psi\ =\ \sqrt{\frac{1}{N} + \rho} \qquad [9]$$

as a useful figure of merit to assess risk reduction. (Note: the above expression is somewhat more useful than writing $N = 1 / (\Psi^2 - \rho)$, since in general what we like to do is exploring how combinations of N and ρ affect total risk.)

Two observations are in order. First, from a practical viewpoint, what an investor cares about is total risk. Therefore, focusing only on diversifiable (non-systematic) risk while neglecting systematic risk is just an academic curiosity with no relevant significance. The usefulness of Eq.[9] is that it captures both risks. And it also links all three relevant variables: N, risk reduction (Ψ), and ρ (a proxy for systematic risk). Despite the obvious simplicity behind the derivation of this equation (it only required the trivial step of finding an approximate expression for the average correlation), we have not seen it anywhere. More surprisingly, many authors started their analyses with expressions similar to that of Eq.[4]. However, instead of finding an approximate expression for the second term, they have concentrated on analyzing actual returns data to estimate the total risk, in some specific markets, and for different values of N (e.g., Habibah et al., 2018; Adamiec and Cernauskas, 2019; Alexeev and Dungey, 2015; Haensly, 2020). Evidently, there is nothing wrong with such exercises. But we reckon that using Eq.[9] is not only simpler but it also offers more insight. Moreover, Eq.[9] makes no reference to either μ or σ. The approaches used in the papers just mentioned required estimates of both, μ and σ, which seems like an unnecessary burden. One might argue that estimating correlations is challenging. After all the process is computationally intensive and correlation estimates are often marred by errors. But to use Eq.[9] we only need to have a reasonable approximation of the average correlation for the assets considered, not a precise value for each ρ_{ij} entry.

EXAMPLES OF APPLICATION

Back to the original question then: what is the appropriate N? Table 1, which shows the value of Ψ (based on Eq.[9]), for different combinations of N and ρ (average correlation) provides some useful guidance. If the average correlation within the universe of assets considered is high, say, 80% or higher, not much risk reduction is achieved, regardless of N. If the average correlation is low, say, between 10% and 20%, a value around N = 30 seems sufficient since for N > 30 the benefits of diversification tend to vanish. (Note that a zero-correlation situation is unlikely to be found in most practical cases.) For correlation values in the middle, a value of N between 10 and 20 is sufficient. Note that this table is valid for any asset class and any time period. In summary, to estimate the value of N, only a rough idea of the value of ρ is required. And in the absence of any correlation estimate, N = 50 is a safe choice.

For anyone skeptical about the usefulness of Table 1, or the validity of Eq.[9], the following example should be convincing. The example is from the study carried out by Elton and Gruber who analyzed returns from 3,290 securities traded in the New York and American Stock Exchange from June 1971 to June 1974 (Table 8 in Elton and Gruber, 1977). Their goal was to estimate for different values of N the corresponding risk reduction, measured by the standard-deviation-of-returns ratio, Ψ in our notation. (Note: this example, with minor variations, was also included as recently as 2014 in the authors' finance book (Elton et al., 2014); it was also discussed in the paper by Statman (1987)). After a fairly involved analysis which required several pages of algebra and statistical derivations in the main body of the paper, three appendices that occupied eleven pages, and estimating a host of unnecessary parameters (e.g., third and fourth moment of CAPM's beta), in addition to the average variance (46.62) and the average covariance (7.05), the authors arrived at the table which is reproduced here (see Table 2, second column).

Had the authors realized that they could have estimated the average correlation using their own data, as 7.05/46.62 = 0.15, they could have used Eq.[9] to calculate Ψ with much less fuss. The third column in Table 2 shows these values. It is clear that the approximate values provided by Eq.[9] are good enough to make a decision regarding the appropriate N. It is also worth noting that in their paper Elton and Gruber reported their results without mentioning the average correlation behind their calculations. Lacking that key information, it is doubtful that an investor could have found that table useful. Note also that the results shown in the third column of Table 2 could have been easily been interpolated from the entries in Table 1.

CONCLUSION

Three conclusions emerge from this study. The first is straightforward: estimating the appropriate N really reduces to estimating the average correlation for the universe of assets considered. This piece of information, in combination with Table 1, is all one needs. Anything else is a distraction. Given the fact that correlation is time-varying (it changes depending on market regimes), settling on a range seems more prudent than agonizing over the "exact" value. After all, in most cases risk reduction becomes stable once N reaches 30 or 40. A value of N = 50, seems sufficient no matter what ρ is.

Second, for all their sound and fury, most previous efforts regarding this topic leave inevitably a much-ado-about-nothing taste. Future efforts should be aimed at finding efficient algorithms to estimate average correlations, especially for assets other than stocks (e.g., bonds, real estate, commodities). The idea of keeping on producing tables showing how portfolio risk decreases as N increases in some very specific market, and especially if systematic risk is not included, seems unwarranted and of little use.

And third, a practical suggestion that has important public policy implications. Pension funds and insurance companies are typically subject to strict investment rules dictated by the regulator, where the degree of diversification is of paramount importance. Frequently, such rules are based on the idea that the portfolio (or some asset class within the portfolio) must have a minimum number of positions to achieve "full diversification." This study challenges this idea for it shows that in general an adequate level of diversification can be achieved with less that 50 positions. Moreover, our study shows that more emphasis should be placed (from a regulatory viewpoint) on looking at the average correlation, since for highlycorrelated assets increasing the number of positions does not decrease the market risk, but instead increases the operational risk (monitoring becomes more onerous).

	ρ (Correlation)										
N	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
10	0.32	0.45	0.55	0.63	0.71	0.77	0.84	0.89	0.95	1.00	1.00
20	0.22	0.39	0.50	0.59	0.67	0.74	0.81	0.87	0.92	0.97	1.00
30	0.18	0.37	0.48	0.58	0.66	0.73	0.80	0.86	0.91	0.97	1.00
40	0.16	0.35	0.47	0.57	0.65	0.72	0.79	0.85	0.91	0.96	1.00
50	0.14	0.35	0.47	0.57	0.65	0.72	0.79	0.85	0.91	0.96	1.00
60	0.13	0.34	0.47	0.56	0.65	0.72	0.79	0.85	0.90	0.96	1.00
70	0.12	0.34	0.46	0.56	0.64	0.72	0.78	0.85	0.90	0.96	1.00
80	0.11	0.34	0.46	0.56	0.64	0.72	0.78	0.84	0.90	0.96	1.00
90	0.11	0.33	0.46	0.56	0.64	0.71	0.78	0.84	0.90	0.95	1.00
100	0.10	0.33	0.46	0.56	0.64	0.71	0.78	0.84	0.90	0.95	1.00

Table 1. Value of Ψ (risk reduction ratio) for different values of N and ρ.

N	**Ψ Elton & Gruber**	**Ψ Eq.[9] with ρ = 0.15**
1	1	1
2	0.76	0.81
4	0.60	0.63
6	0.54	0.56
8	0.51	0.52
10	0.49	0.50
20	0.44	0.45
50	0.41	0.41
100	0.40	0.40
200	0.39	0.39
500	0.39	0.39
1000	0.39	0.39

Table 2. Value of Ψ (risk reduction ratio) for different values of N, based on the example analyzed by Elton and Gruber (1977), using two different approaches.

REFERENCES

Adamiec, L. and Cernauskas, D. (2019). Optimal Number of Assets for Reduction of Market Risk, *International Journal of Economics, Business and Management Research,* 3(3), 45-54.

Alexeev, V. and Dungey, M. (2015). Equity Portfolio Diversification with High Frequency Data, *Quantitative Finance*, 15(7), 1205-1215. DOI: 10.1080/14697688.2014.973898

De Jong, M. (2018). Portfolio Optimisation in an Uncertain World, *Journal of Asset Management*, 19(4), 216-221. DOI: 10.1057/s41260-017-0066-3

Elton, E. and Grober, M. (1977). Risk Reduction and Portfolio Size: An Analytical Solution, *The Journal of Business*, 50(4), 415-437. DOI:10.1086/295964

Elton, E., Gruber, M., Brown, S. and Goetzmann, W. (2014). Modern Portfolio Theory and Investment Analysis, 9th Edition, Wiley.

Fragkiskos, A. (2014). What is Portfolio Diversification?, *Alternative Investment Analyst Review*, 1(3). https://caia.org/aiar/1951

Habibah, U., Sadhwani, R. and Memon, A. (2018). Selecting Optimal Portfolio in Pakistan, *IBA Journal of Economics and Finance*, (2)1, 68-76. DOI:10.30537/SIJEF.V2I1.202

Haensly, P. (2020). Risk Decomposition, Estimation error, and Naïve Diversification, *The North American Journal of Economics and Finance*, 55(c). DOI: 10.1016/j.najef.2020.101146

Markowitz, H. (1952). Portfolio Selection, *Journal of Finance,* 7(1), 77-91. https://doi.org/10.2307/2975974

Norsiman, N., Yakob, N. and McGowan, C. (2019). The Effect of Portfolio Diversification for the Bursa Malaysia, *Accounting and Finance Research,* 8(4), 76-92. DOI:10.5430/afr.v8n4p76

Raju, R. and Agarwalla, S. (2021). Equity Portfolio Diversification: How Many Stocks Are Enough? Evidence From India, available at SSRN: https://ssrn.com/abstract=3790117 or http://dx.doi.org/10.2139/ssrn.3790117

Shawky H, and Smith, D. (2005). Optimal Number of Stock Holdings in Mutual Fund Portfolios Based on Market Performance, *The Financial Review*, 40(4), 481-495. https://doi.org/10.1111/j.1540-6288.2005.00120.xCit

Solnik, B. (1974). Why Not Diversify Internationally Rather Than Domestically?, *Financial Analysts Journal*, 30(4), 48-54. https://doi.org/10.2469/faj.v30.n4.48

Statman, M. (1987). How Many Stocks Make a Diversified Portfolio?, *Journal of Financial and Quantitative Analysis*, 22(3), 353-363. DOI: https://doi.org/10.2307/2330969

Tang., G. (2004). How Efficient Is Naive Portfolio Diversification?, *The International Journal of Management Science*, 32(2), 155-160. https://doi.org/10.1016/j.omega.2003.10.002

Zaimovic, A., Omanovic, A. and Arnaut-Berilo, A. (2021). How Many Stocks Are Sufficient for Equity Portfolio Diversification? A Review of the Literature, *Journal of Risk and Financial Management*, 14(11), 1-30. https://doi.org/10.3390/jrfm14110551

Zhou, Y. (2014). Number of Stocks in Portfolio and Risk Reduction, *Master of Science in Statistics thesis,* University of California, Los Angeles, https://escholarship.org/content/qt78v69630/qt78v69630_noSplash_63a4e31fa02dc4ad7988d734e4da5bb5.pdf

Centro UC
CLAPES UC
Centro Latinoamericano de
Políticas Económicas y Sociales

clapesuc

@clapesuc

claps_uc

clapesuc