

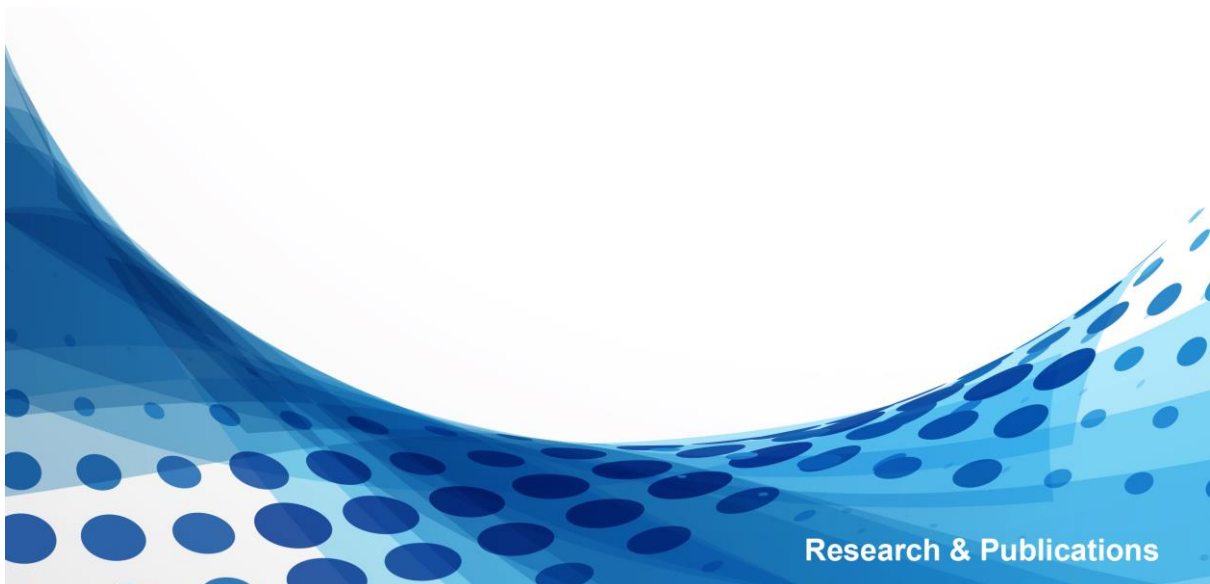


INDIAN INSTITUTE OF MANAGEMENT AHMEDABAD

IIMA
Working Paper

Equity portfolio diversification: how many stocks are enough? Evidence from India

Rajan Raju
Sobhesh Kumar Agarwalla



Research & Publications

Equity portfolio diversification: how many stocks are enough? Evidence from India

Rajan Raju
Sobhesh Kumar Agarwalla

February 2021

The main objective of the working paper series of the IIMA is to help faculty members, research staff and doctoral students to speedily share their research findings with professional colleagues and test their research findings at the pre-publication stage. IIMA is committed to maintain academic freedom. The opinion(s), view(s) and conclusion(s) expressed in the working paper are those of the authors and not that of IIMA.

Equity portfolio diversification: how many stocks are enough? Evidence from India

Rajan Raju*

Sobhesh Kumar Agarwalla†

February 2021

ABSTRACT

How many stocks are required to reduce unsystematic risk significantly is an important question for investors. While there is a large body of research on the subject in the United States, there is little formal work on this question in India. We show that a 15-20 stock portfolio, the traditional market rule-of-thumb for a diversified portfolio, is likely inadequate to minimise unsystematic risk. We show that an investor could target to reduce diversifiable risk by 90% with a 90% confidence with a portfolio of 40-50 stocks. We build a practical framework that serves as a baseline for investors to target a specific reduction in diversifiable unsystematic risk at a chosen confidence level.

Keywords— Portfolio Construction, Diversification, Time series standard deviation

JEL Classification Codes G00, G11, C15

*Director, Invespar Pte Ltd, Email: rajanraju@invespar.com, Phone: +65.62380361

†Professor, Finance and Accounting, IIM Ahmedabad, Email: sobhesh@iima.ac.in, Phone: +91.79.66324865

1 Introduction

Markowitz (1952) formalized diversification using the statistical concepts of mean and variance of returns of the underlying assets in a portfolio. Practically, however, the minimum number of stocks that make a well-diversified portfolio is an important question for long-term investors. Holding too few stocks exposes the investor to unnecessary idiosyncratic risk, and holding too many stocks is costly (both transactionally and ongoing monitoring). Several studies in the US have evaluated the reduction in the portfolio risk as portfolio size increases since Evans and Archer (1968). In the Indian context, research on this subject has been limited (Sanyal and Sen, 1998; Irala and Patil, 2007). Traditionally, a portfolio of 15-20 stocks is seen as being well-diversified.¹ In this paper, we explore how many stocks make for a well-diversified portfolio of Indian stocks and propose a practical heuristic for investors to evaluate the number of stocks in their portfolios.

Using standard deviation as the measure of risk, we find that a risk-averse investor would need to hold *significantly more* stocks than the ‘conventional wisdom’ of 15-20 stocks. The actual number of stocks would depend on the investor’s risk-reduction expectation/preferences, desired confidence level to meet their objectives, and the weighting scheme used for portfolio construction. This number is larger than suggested by Evans and Archer (1968) or Irala and Patil (2007) but similar to findings by others (Statman, 1987).

With the increasing concentration of large-cap indexes and preliminary evidence of herding behaviour amongst the Indian large-cap mutual funds (see Raju et al., 2020), our findings have practical implications for long-term investors, advisors and fund managers looking to manage the risk in their equity portfolios. This paper adds to the body of formal analysis on the ‘how many stocks’ issue in the Indian context.

The rest of the paper is organized as follows: Section 2 covers the literature, in Section 3, we discuss our methodology and data; in Section 4, we present our results; and we conclude in Section 5.

2 Literature Review

Markowitz’s (1952) Capital Asset Pricing Model (CAPM) implies that total risk, measured by the standard deviation of any portfolio, can be split into two components: systematic and unsystematic risk. Increasing diversification reduces the portfolio’s overall risk by reducing the asset-specific unsystematic risk leaving the

¹See for instance

Birani, T. DNA. May 28, 2019. “Just 15-20 stocks can make a well-diversified portfolio”. Retrieved from <https://www.dnaindia.com/personal-finance/report-just-15-20-stocks-can-make-a-well-diversified-portfolio-2754391> on February 15, 2021.

Dharma, P. Livemint. June 12, 2020. “Do all diversified equity portfolios give good returns in the long-term?”. Retrieved from <https://www.livemint.com/money/personal-finance/do-all-diversified-equity-portfolios-give-good-returns-in-the-long-term-11591969263371.html> on February 15, 2021.

Vinta S. R. Ask Value Research. August 17, 2020. “What is the maximum number of stocks I should buy?”. Retrieved from <https://www.valueresearchonline.com/stories/48411/what-is-the-maximum-number-of-stocks-i-should-buy/> on February 15, 2021.

Kayal, P. Financial Express. September 16, 2020. “Risks vs Returns: How much portfolio diversification is too much?”. Retrieved from <https://www.financialexpress.com/money/risks-vs-returns-how-much-portfolio-diversification-is-too-much/2084045/> on February 15, 2021.

investor to deal with only systematic risk. There are decreasing marginal diversification benefits as we add assets. Not diversifying enough has real costs. JP Morgan published a white paper in 2004 titled “Beating the Odds: Improving the 15% Probability of Staying Wealthy” (Zaharoff, 2004) where excessive concentration was the number one reason why the very wealthy lose their fortunes. How many stocks are required in a portfolio to minimise idiosyncratic risk is, therefore, a key question for investors.

Evans and Archer (1968) evaluated the number of stocks required to eliminate the unsystematic risk from a portfolio and concluded that only 8 to 10 stocks were adequate to deliver a diversified portfolio. Since then, there have been many studies in the developed markets (for instances, Statman, 1987, 2004; Domian et al., 2006; Benjelloun, 2010; Alexeev and Tapon, 2013). These studies use a wide variety of risk measures including, but not limited to, the standard deviation of returns, variance, semi-variance, mean absolute deviation, expected shortfall, and terminal wealth standard deviation. There is, however, no consensus on what is the minimum number of stocks that constitute a well-diversified portfolio as yet. Similar formal studies in India have been limited - Irala and Patil (2007) being the most recent. They concluded that a portfolio of 10-15 stocks is ‘appropriate as the reduction in risk is only marginal thereafter’. We examine this conclusion further.

3 Methodology and Data

3.1 Methodology

The risk of any portfolio of stock is a function of the risks of the individual stocks in the portfolio, the covariance amongst the stocks and the allocation to each stock. Following CAPM, any portfolio’s total risk is comprised of systematic risk and unsystematic risk. The variance of a portfolio with N securities can be written as

$$\sigma_p = \sum_{i=1}^N \left(\frac{1}{N}\right)^2 \sigma_i^2 + \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \left(\frac{1}{N}\right)^2 cov(i, j) \quad (1)$$

Taking expected values

$$\begin{aligned} E(\sigma_p) &= N \left(\frac{1}{N}\right)^2 \bar{\sigma}^2 + N(N-1) \left(\frac{1}{N}\right)^2 \overline{cov(i, j)} \\ &= \left(\frac{1}{N}\right) \bar{\sigma}^2 + \left(\frac{N-1}{N}\right) \overline{cov(i, j)} \end{aligned} \quad (2)$$

where $E(\sigma_p)$ is the expected portfolio variance, $\bar{\sigma}^2$ is the average variance for all stocks in the portfolio, $\overline{cov(i, j)}$ is the average covariance across all stocks in the portfolio and N is the number of stocks for the portfolio. One obvious implication of equation 2, is that as the number of stocks, N , increases, the expected portfolio variance reduces and will converge to the average covariance of stocks in the portfolio.

We choose the Nifty-500 as our universe of stocks to represent the market. This index comprises about 96.1% of the market capitalization of the Indian equity market. We, therefore, consider $N=500$ as our total

market.² We construct portfolios by randomly drawing n stocks using the constituents of the Nifty-500 for the relevant six-month period.³ Portfolios are equally weighted and we select portfolios P_i^n , where $n = 1, 10, 20, 30 \dots 500$ is the number of stocks in the portfolio and $i = 1, 2, 3 \dots M$ represents the iteration. We construct $M = 10,000$ n -stock portfolios for each $n = 1 \dots N$. When $n = N$ it is the market portfolio. 10,000 iterations would generate a representative average standard deviations for our $n = 1 \dots 500$ portfolios⁴.

We run our analysis at 6-month intervals⁵ between December 2014 through to December 2020. We compute daily log total returns for each stock and take average returns over the past (one) three years from the period end. If the stock has not been in existence for the full (one) three years, we use all available data to estimate daily returns. For $n = 1 \dots N$ the return of the n -stock equally weighted random portfolio at iteration $i = 1 \dots 10,000$ at time t is defined as

$$r_{i,t}^n = \sum_{j=1}^n \frac{(r_{j,t})_i}{n} \quad (3)$$

where $r_{i,t}^n$ is the return for portfolio i with n equal weighted holdings for period t , j is the $1..n$ stocks, $(r_{j,t})_i$ is the return of stock j in time t for portfolio i and n is the number of stocks in the equal weighted portfolio i for the iteration. The average return for portfolio i over time $t = 1 \dots T$ is

$$\bar{r}_i^n = \sum_{t=1}^T \frac{r_{i,t}^n}{T} \quad (4)$$

Similarly, our risk measure, the standard deviation, is

$$\sigma_i^n = \sqrt{\sum_{t=1}^T \frac{(r_{i,t}^n - \bar{r}_i^n)^2}{T - 1}} \quad (5)$$

where σ_i^n is the standard deviation of portfolio i with n equal weighted stocks, $r_{i,t}^n$ is the return of the portfolio i during time t , and \bar{r}_i^n is the average return of portfolio i across all T . The standard deviation can be averaged across M portfolios each with n equal weighted stocks as

$$\bar{\sigma}^n = \sum_{m=1}^M \frac{\sigma_m^n}{M} \quad (6)$$

where σ_i^n is the standard deviation for portfolio i with n equally weighted stocks. There are two special portfolios: $n=1$, when we have a portfolio consisting of a single stock and $n=N$ when all stocks in the universe hold an equal-weighted position - the market portfolio. The standard deviation of the former, σ^1 represents

²The Nifty-500 has approximately 500 stocks - sometimes marginally more. We constrain the market portfolio to 500 for the paper.

³We choose no-replacement draws to reflect investors' behaviour of holding different stocks in their portfolios. If the number of stocks with available prices is less than 500, we make all portfolios where $n >$ total constituents and $n \leq 500$, as the market portfolio. If there are 501 stocks as index constituents, we make the $n=500$ portfolio as our market portfolio using all stocks. This allows us to construct portfolios with $n = 1..500$ stocks across time. As we shall see, this expediency has no bearing on our conclusions.

⁴As $n = 51$, $M = 10,000$ and $T = 13$, we generate 6,630,000 portfolios in total for each time-series of returns.

⁵The Nifty-500 is reviewed twice a year. Using 6-month intervals allows us to capture the effects of constituent changes.

the idiosyncratic or unsystematic risk while the standard deviation of the latter, σ^N , represents only systematic risk. Or put another way, in terms of equation 2, the former has no covariance term, while the latter portfolio has only the covariance term.

We can now define a number of measures of risk. First, standard deviation is the raw and unscaled measure of total risk, ϕ_1 :

$$\phi_1(n) = \bar{\sigma}^n \quad (7)$$

As $n \rightarrow N$, $\bar{\sigma}^n \rightarrow \sigma^N$. Second, we measure unsystematic risk by subtracting the market risk. We call this measure ϕ_2 :

$$\phi_2(n) = \bar{\sigma}^n - \sigma^N \quad (8)$$

As $n \rightarrow N$, $\phi_2(n) \rightarrow 0$. Third, we scale the measure of risk by dividing portfolio risk by the risk of the market portfolio:

$$\phi_3(n) = \frac{\bar{\sigma}^n}{\sigma^N} \quad (9)$$

In this measure, ϕ_3 , as $n \rightarrow N$, $\phi_3(n) \rightarrow 1$. When $n = 1$, ϕ_3 measures the multiple of market risk that the average unsystematic risk of single stocks are. As market risk is dynamic, this measure helps show the riskiness inherent in the market for that slice of time.

Finally, we can scale and normalize the measure using the range between pure unsystematic risk and pure systematic risk:

$$\phi_4(n) = \frac{\bar{\sigma}^n - \sigma^N}{\sigma^1 - \sigma^N} \quad (10)$$

This measure, ϕ_4 will always be between 1 and 0; 1 reflecting 100% unsystematic risk and 0 reflecting the complete elimination of unsystematic risk, leaving only the systematic risk. ϕ_4 shows the effective reduction in diversifiable risk. If ϕ_4^n is, say, 0.10, we would say that the portfolio with n stock has reduced diversifiable risk by 90%.

So far, we have been working on *average* measures of risk. An individual investor, however, is not likely to hold an ‘average’ portfolio. [Newbould and Poon \(1996\)](#) make a case for using statistical techniques to estimate risk for a given confidence threshold for portfolios. An investor could choose, say, 95% as her threshold. For such an investor, we choose 95% of all generated portfolios rather than relying on the ‘mean’ of all portfolios. 5% of portfolios will still lie outside the threshold. This approach allows us to generate statistical estimates of risk that will be more relevant for an investor. We calculate the quantile standard deviations from our population of simulated portfolios. For every n stock portfolio, from the universe we can compute the q^{th} quantile standard

deviation, σ_q^n and apply it to equation 10:

$$\phi_5(n, q) = \frac{\sigma_q^n - \sigma^N}{\sigma^1 - \sigma^N} \quad (11)$$

We use these quartiles to build tables that an investor could use to determine the diversification they would like and calculate the number of stocks based on statistical confidence intervals. So for our investor choosing a 95% confidence, we would select $q = 97.5\%$ and $q = 2.5\%$ as our upper and lower bounds.

To evaluate the incremental benefit to diversification of adding additional stock, similar to [Statman \(1987\)](#); [Newbould and Poon \(1996\)](#); [Chong and Phillips \(2013\)](#), we compare the difference in the measure of risk of incremental holding ($\sigma^n - \sigma^{n-1}$) scaled by the market portfolio, σ^N .

$$DB_n = \frac{\sigma^n - \sigma^{n-1}}{\sigma^N} \quad (12)$$

This is the diversification benefit of adding additional stock to a portfolio. Given the diminishing benefits of diversification, this measure will asymptote to 0.

3.2 Data

All our data is from Refinitiv and Datastream. We get the constituents of the Nifty-500 from Refinitiv at the end of each month between December 2014 and 2020. We have a total of 760 stocks in the universe (accounting for joiners and leavers for the index during the observation period). We get total returns for each stock from Datastream between December 1, 2011 and December 31, 2020. Table 1 shows the number of constituents at the end of each period. We have prices for all the constituents, including stocks de-listed subsequently, except for six stocks. These six stocks were part of the Nifty-500 for at least one 6-month period between December 2014 and December 2016.⁶ Given the universe of 500 stocks for each period, the missing information for the 6-stocks is not material for our analysis. By construction, as we generate a large number of randomly generated equal-weighted portfolios for every period, the survivorship bias is non-material/small. We show the average variance of the individual stock, and the average variance to the market portfolio stays within a reasonable range except during the Covid period (defined as between February through September 2020), when we note spikes in both values. For this period, we have all stock information and, therefore, fully capture all information.

We use three-year historical daily total returns to estimate the expected returns of each stock as our baseline. For robustness, we use the shorter 1-year historical daily total returns for the analysis to ensure there is no time-series period bias. *A priori*, we expect that while 1-year based returns will have higher variability than 3-year based average returns, there should be no reason for the conclusions to be different: *portfolio expected risk should reduce as the number of holdings increase*. The one-year or three-year return dataset slice is used

⁶For these periods, we use $n = 500$ as the market portfolio even though actual N is slightly lower than 500. As we calculate n in intervals of 10, this approximation does not affect our results

to calculate portfolio standard deviations for the one-year and three-year based analysis, respectively. We use Ledoit Wolf shrinkage in our estimation of covariances (see [Ledoit and Wolf, 2003](#)).

4 Results and Discussion

Table 2 tabulates the risk measures for $n = 1, 10, 20 \dots, 500$ using both 1-year time-series and 3-year time-series of returns. These are the averages for the risk measures across time and iteration. Figures 1, 2, 3 and 4 chart the various risk measures σ^n , ϕ_2^n , ϕ_3^n , and ϕ_4^n plotted against the number of stocks in the portfolio, n , and the reduction in risk arising from increasing diversification. In both the 1-year and 3-year time-series we find that as the number of stocks (n) increases, risks decrease. As expected, the pure unsystematic and systematic risks are higher for the 1-year time-series analysis. When $n=1$, the risk is purely systematic risk. *Average* risk falls steeply as the number of stocks increases before tapering off around $n = 50$. Note that pure unsystematic risk, the average of risks for all portfolios with $n = 1$, is 37.13% and 38.84% for 3-year and 1-year respectively, while pure systematic risk, average risk from an equal-weighted portfolio of all stock across time, is 16.11% and 16.82% : 2.305 times and 2.308 times, the value of $\bar{\phi}_3^1$ for each time-series. The differences between the two time-series are marginal⁷ especially in the scaled measures - ϕ_2 , ϕ_3 , and ϕ_4 - indicating a degree of robustness in the approach. Figure 5 indicates that there are material diversification benefits in having a portfolio of 40 or 50 stocks as compared to a portfolio of 20.

The changes to Nifty-500 constituents arising from the semi-annual review, and consequently the changes in expectations of returns and risk, will affect diversification benefits. To examine the impact of the risk measures over the 13 6-month intervals from December 2014 through December 2020, figures 6 and 7, show ϕ_2 , ϕ_3 , and ϕ_4 over each period for the two time-series used. Not surprisingly, the variation in ϕ_2 and ϕ_3 are more pronounced in the 1-year based time series. This series is more volatile.

Our observation period includes periods of market stress, especially the covid pandemic in the first half of 2020. This is observed as the increase in unsystematic risks as measured by $\bar{\sigma}^1$ and $\bar{\phi}_2$ between December 2019 and December 2020 as shown in figures 8 and 10. Interestingly, $\bar{\phi}_3$ fell between December 2019 and June 2020 indicating that while unsystematic risk increased, the market risk was simultaneously elevated.

Figures 9 and 11 plot the trends of the correlation among securities and with the market portfolio. Again, the volatility is higher for the shorter duration time-series. That said, the spread between average correlations between securities and average correlations with the market portfolio has been relatively stable. Note that the average correlations amongst the stocks and between stocks and the equal-weighted market portfolio spiked in June 2020 (figure 11). The Indian equity market fell almost 30% from early February 2020 value during the pandemic period. It also fell almost 20% in February 2016 from the start of the year before recovering by the middle of the year. The 2016 drawdown is seen as a spike in the 1-year time-series (figure 8) between

⁷Hence the lack of separation in the curves shown in figures 3, 4 and 5

December 2015 and June 2016. The longer-duration 3-year time series was more muted during 2016, however, it shows a spike during the covid period. Alexeev and Tapon (2013) also report similar phenomenon of spikes in correlation in the United States during the 2008 global financial crisis. An immediate implication of spikes in correlation is a potential reduction of diversification benefits. In equation 2, the covariance term will increase when correlations spike up, increasing the expected portfolio variance. However, the relationship is a complex one that involves the correlations between the stocks themselves and with the market portfolio.

The approach so far uses averages. However, an investor would hold only *one* portfolio from the universe of available stock. An average portfolio is, therefore, does not help an investor to make decisions for their unique portfolio. The investor should, most likely, be unwilling to use any analysis based on the *average* results from a large number of equal-weighted portfolios from that same universe of stocks, since every point in this average hides a range of outcomes for portfolios holding the same number of holdings. Following Newbould and Poon (1996), we find upper and lower bounds of risk for a particular confidence interval using equation 11. This method allows us to capture the full distribution of portfolios at every point and thereby tailor the analysis for the investors specific risk appetite. A more conservative investor may choose to have a 99% confidence while a more aggressive one may be comfortable at a 90% confidence (a 10 in 100 chance of that the diversifiable risk reduction target will not be contained in the samples chosen.)

Figure 12 and 13 shows the 99%, 95% and 90% confidence intervals for the 50,000 simulated portfolios for the period ending December 2020 using the two durations for our analysis⁸. We use this to illustrate a few points. First, at the market portfolio, the risk is the same for all confidence levels by construction. This is the terminal point of all the upper and lower bound areas in the left panel of the figures. Second, the reduction of portfolio risk as the number of holdings increases is clear in the figure - all the curves converge. Risk reduction is delayed depending on the confidence interval - the higher the interval, the longer the delay. Third, the more volatile 1-year time series means that as n increases from 1, the risk reduction has a wider spread and is more delayed than the lower volatile 3-year time-series. The right-hand panel shows that the former series has a higher upper bound of risk than the latter series. Fourth, diversification risk is not eliminated at 20 stocks. The slope of the lines only becomes parallel well after $n = 20$.

Table 3 shows the same data from a different perspective. It shows the upper-bounds for ϕ_3 and ϕ_4 across n , confidence levels and the time-series duration of the basis for calculation (1-year and 3-year). ϕ_3 is expressed against the average ϕ_3 for each value of n . Thus, the 1-stock portfolio risk is likely to be 164% of the average risk of a 1-stock portfolio with a 99% confidence and 125% higher with a 90% confidence using the 3-year time-series. The ϕ_3 between the shorter-duration and the longer-duration time-series, becomes similar around $n=30$.

In addition, we show ϕ_4 where the 1-stock portfolio (pure unsystematic risk) as the baseline is scored as 100. A 1-stock portfolio for the period with 99% confidence would be 217% and at 90% 145% higher than

⁸We have run the analysis for other intervals, with the same outcomes/conclusions.

the average for all 1-stock portfolios. As n increases, the diversifiable risk, ϕ_4 , falls rapidly. By $n = 30$, the portfolio has diversified away 89% (100 - 11) of the diversifiable risk for 90% of all the generated portfolios using 3-year returns. Note, using the more static average, at $n = 20$, 92% of the diversifiable risk (in the ϕ_4 Avg column) is reduced for both 1-year and 3-year return series. Using averages, the conventional popular wisdom that a 20 stock portfolio is well-diversified seems reasonable. However, this may not be the right conclusion and an investor is likely to be carrying a significant amount of diversifiable risk if she were to follow conventional wisdom. With a 90% confidence, only about 84% (100 - 16%) or 85% (100 - 15%) would be diversified away depending on which duration of return series we choose.

The table allows an investor to choose the confidence level and then run down the chosen column values until she is comfortable with the riskiness of the portfolio compared to the market portfolio to arrive at the holdings required for the diversification desired. We prefer to use ϕ_4 as our measure of diversifiable risk, but ϕ_3 is also a valid measure. ϕ_4 scales diversifiable risk (the difference between unsystematic and systematic risk divided by the difference between pure unsystematic and pure systematic risk) while ϕ_3 expresses total risk (both unsystematic and systematic) as a multiple of pure systematic risk.

The table is a ready heuristic for an investor who holds a non-equal weighted portfolio as well. Most investors would additionally choose benchmarks to track their portfolios. Such investors can compare the historical variances of the portfolio returns versus the chosen benchmark returns to determine how much diversifiable risk is being carried in the portfolio for a given confidence level, as done in this study.

Figures 14 and 15 shows the number of stocks required to diversify away 90% or 95% of the diversifiable risk over the observation period using the two time-series returns. The number of stocks increases based on the chosen confidence level and the amount of risk the investor expects to diversify away. The data matches intuition: a more risk-averse investor with higher confidence expectations should have more stocks in their portfolio. Table 4 shows the average number of stock across the periods under observation that will diversify away 90% and 95% of the diversifiable risk using Average and 90, 95 and 99% confidence.

The evidence and method show that a 20-stock equal-weighted portfolio would *on average* diversify away 90% of diversifiable risk. However, this is cold comfort to an investor who is likely to find that the realized diversification is very different from the expectation of 90% reduction from holding their *specific* 20-stock portfolio. It would take around 40-50 stocks in an equal-weighted portfolio to give the same investor 90% confidence to diversify away 90% of diversifiable risk. As the confidence level required or the amount of diversifiable risk to be reduced increases, the investor would need to hold a larger number of stocks in their portfolio.

5 Conclusion

Traditional wisdom in India argues that 15 to 20 stocks are sufficient for efficient diversification. Using time-series return data for constituents of the Nifty-500 index between December 2014 and 2020, we show that using averages of randomly generated equal-weighted portfolios, 20 stocks may diversify away about 90% of diversifiable risks. However, in practice, investors would find that their chosen 20-stock portfolio has a low probability of realising the expected diversification benefits. With standard statistical methods, we show that an investor could target a particular level of diversification-related risk reduction and choose the number of stocks for a given confidence interval. Using such a process, an investor could choose a 40-50 stock equal-weighted portfolio to target a 90% diversifiable risk reduction with 90% confidence or a 50-60 stock portfolio to target the same risk reduction with a higher 95% confidence. In all cases, the statistical approach shows that *the number of stocks required for diversification is higher than the oft-quoted 15-20 stocks*.

Our analysis uses a random stock selection and equal-weighting portfolio construction process. In practice, investors would likely have particular reasons for their stock selection and allocation decisions. The choice of the stocks and their weights in a portfolio will determine the reduction of diversifiable risks. Our approach could serve as a baseline for investors in their portfolio construction decisions and help them target a particular amount of risk reduction with a chosen confidence level. This approach can be used to deal with diversifiable risk reduction for any portfolio and its chosen benchmark.

We have focused this paper purely on risk and ignored returns. One logical extension of our work is to add expected portfolio returns to build a risk-adjusted return framework to diversify unsystematic risk.

References

- Vitali Alexeev and Francis Tapon. Equity Portfolio Diversification: How Many Stocks are Enough? Evidence from Five Developed Markets. Working Papers 2013-16, University of Tasmania, Tasmanian School of Business and Economics, November 2013. URL <https://ideas.repec.org/p/tas/wpaper/17313.html>.
- Hicham Benjelloun. Evans and archer - forty years later. *Investment Management and Financial Innovations*, 7:98–104, 01 2010.
- James Chong and G. Michael Phillips. Portfolio size revisited. *The Journal of Wealth Management*, 15(4):49–60, 2013. ISSN 1534-7524. doi: 10.3905/jwm.2013.15.4.049. URL <https://jwm.pm-research.com/content/15/4/49>.
- Dale L. Domian, David A. Louton, and Marie D. Racine. Diversification in portfolios of individual stocks: 100 stocks are not enough, Jun 2006. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=906686.
- John L. Evans and Stephen H. Archer. Diversification and the reduction of dispersion: An empirical analysis*.

- The Journal of Finance*, 23(5):761–767, 1968. doi: <https://doi.org/10.1111/j.1540-6261.1968.tb00315.x>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1968.tb00315.x>.
- Lokanandha Reddy Irala and Prakash Patil. Portfolio size and diversification. *SCMS Journal of Indian Management*, 4(1), Mar 2007. URL <https://ssrn.com/abstract=977763>.
- Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621, 2003. ISSN 0927-5398. doi: [https://doi.org/10.1016/S0927-5398\(03\)00007-0](https://doi.org/10.1016/S0927-5398(03)00007-0). URL <https://www.sciencedirect.com/science/article/pii/S0927539803000070>.
- Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. ISSN 00221082, 15406261. URL <http://www.jstor.org/stable/2975974>.
- Gerald D. Newbould and Percy S. Poon. Portfolio risk, portfolio performance, and the individual investor. *The Journal of Investing*, 5(2):72–78, 1996. ISSN 1068-0896. doi: 10.3905/joi.5.2.72. URL <https://joi.pm-research.com/content/5/2/72>.
- Rajan Raju, Ananya Mittal, and Rajiv Baruah. Diversification Within Large Cap Equity Mutual Funds in India. *SSRN eJournal*, page 18, Jul 2020. doi: <https://dx.doi.org/10.2139/ssrn.3660345>. URL <https://ssrn.com/abstract=3660345>.
- G. Sanyal and S. Sen. Portfolio Diversification and Risk Reduction in India - An Empirical Analysis. *Finance India*, Jun 1998.
- Meir Statman. How many stocks make a diversified portfolio? *The Journal of Financial and Quantitative Analysis*, 22(3):353–363, 1987. ISSN 00221090, 17566916. URL <http://www.jstor.org/stable/2330969>.
- Meir Statman. The diversification puzzle. *Financial Analysts Journal*, 60(4):44–53, 2004. doi: 10.2469/faj.v60.n4.2636. URL <https://doi.org/10.2469/faj.v60.n4.2636>.
- Alexander Zaharoff. Beating the odds: Improving the 15% probability of staying wealthy, 2004.

Table 1: Number of constituents by period

Date	No. Constituents	Missing price information
2014-12-31	500	4
2015-06-30	500	4
2015-12-31	500	3
2016-06-30	501	2
2016-12-31	501	4
2017-06-30	501	0
2017-12-31	501	0
2018-06-30	501	0
2018-12-31	502	0
2019-06-30	502	0
2019-12-31	502	0
2020-06-30	501	0
2020-12-31	501	0

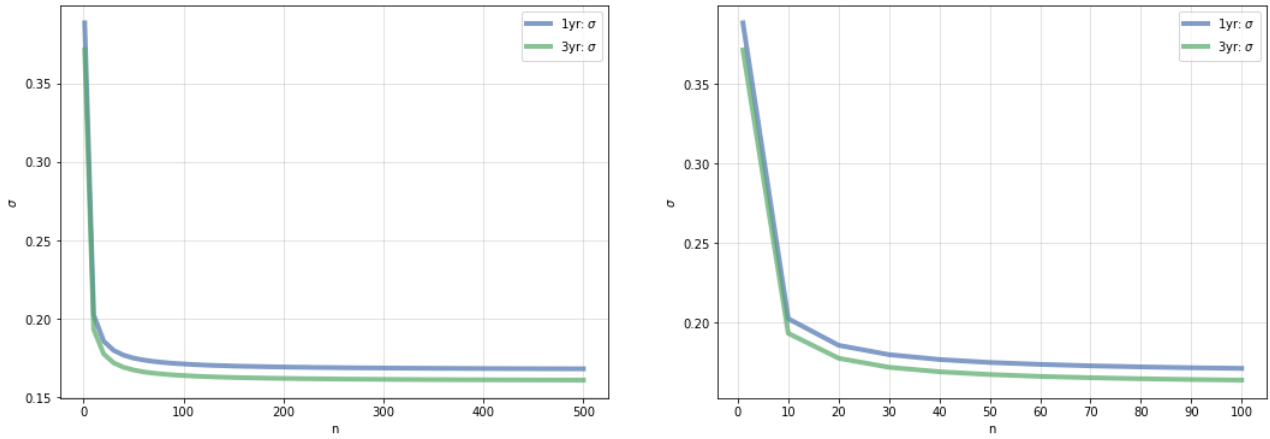
Tickers with missing price information are delisted tickers for which no price information is available on Datastream

Table 2: Average risk measures

N	1yr					3yr				
	σ	$\bar{\phi}_2$	$\bar{\phi}_3$	$\bar{\phi}_4$	Div Benefit	σ	$\bar{\phi}_2$	$\bar{\phi}_3$	$\bar{\phi}_4$	Div Benefit
1	0.3884	0.2201	2.3085	1.0000		0.3713	0.2102	2.3053	1.0000	
10	0.2025	0.0343	1.2039	0.1558	-110.4612	0.1934	0.0323	1.2005	0.1536	-110.4738
20	0.1858	0.0176	1.1045	0.0799	-9.9369	0.1777	0.0166	1.1033	0.0791	-9.7263
30	0.1800	0.0117	1.0697	0.0533	-3.4831	0.1720	0.0110	1.0680	0.0521	-3.5268
40	0.1769	0.0087	1.0515	0.0394	-1.8148	0.1692	0.0081	1.0506	0.0388	-1.7421
50	0.1750	0.0068	1.0404	0.0309	-1.1085	0.1675	0.0064	1.0398	0.0305	-1.0817
60	0.1739	0.0056	1.0334	0.0255	-0.7090	0.1663	0.0052	1.0325	0.0249	-0.7313
70	0.1729	0.0047	1.0280	0.0214	-0.5400	0.1655	0.0044	1.0274	0.0210	-0.5071
80	0.1723	0.0040	1.0239	0.0183	-0.4034	0.1648	0.0038	1.0234	0.0179	-0.3979
90	0.1717	0.0035	1.0207	0.0158	-0.3216	0.1643	0.0033	1.0203	0.0156	-0.3093
100	0.1713	0.0031	1.0184	0.0141	-0.2315	0.1640	0.0029	1.0180	0.0138	-0.2317
:	:	:	:	:	:	:	:	:	:	:
460	0.1683	0.0001	1.0004	0.0003	-0.0092	0.1611	0.0001	1.0004	0.0003	-0.0108
470	0.1683	0.0000	1.0003	0.0002	-0.0122	0.1611	0.0000	1.0003	0.0002	-0.0109
480	0.1683	0.0000	1.0002	0.0002	-0.0083	0.1611	0.0000	1.0002	0.0001	-0.0091
490	0.1683	0.0000	1.0001	0.0001	-0.0119	0.1611	0.0000	1.0001	0.0001	-0.0112
500	0.1682	0.0000	1.0000	0.0000	-0.0089	0.1611	0.0000	1.0000	0.0000	-0.0078

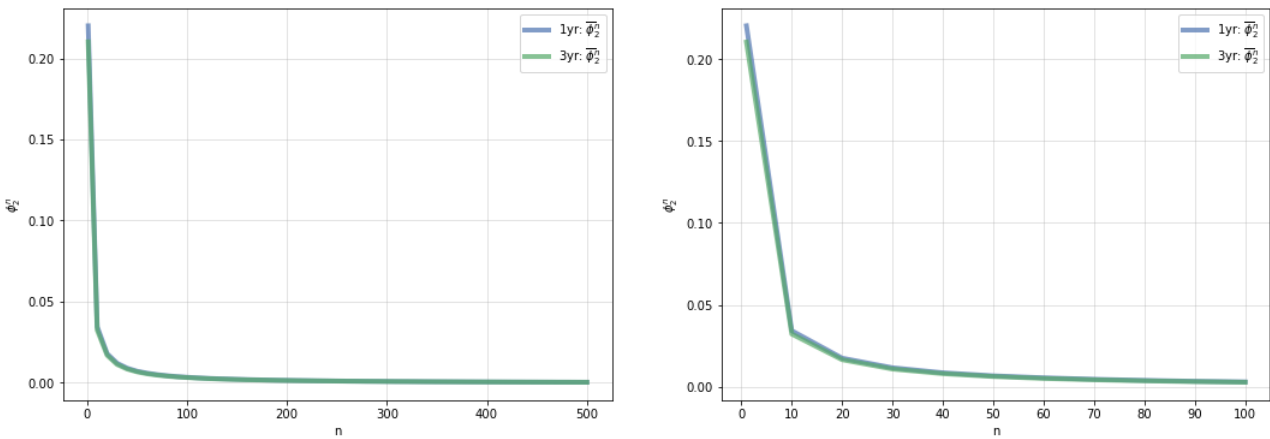
The risk measures are averages for each portfolio of n holdings across the 13 time periods. We show the measures using 1 year and 3 year time series of total returns.

Figure 1: $\bar{\sigma}^n$ for all portfolios



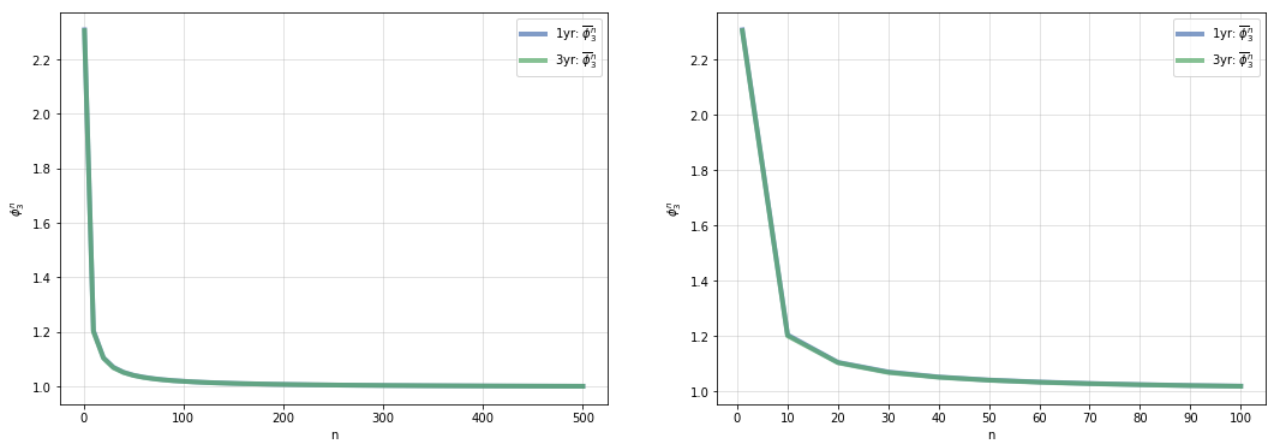
The curves represent the mean standard deviation across all the iterations (10,000 each) for $n = 1,10,20\dots 500$ (51 portfolios) across 13 periods from Dec 2014 through Dec 2020.

Figure 2: $\bar{\phi}_2^n$ for all portfolios



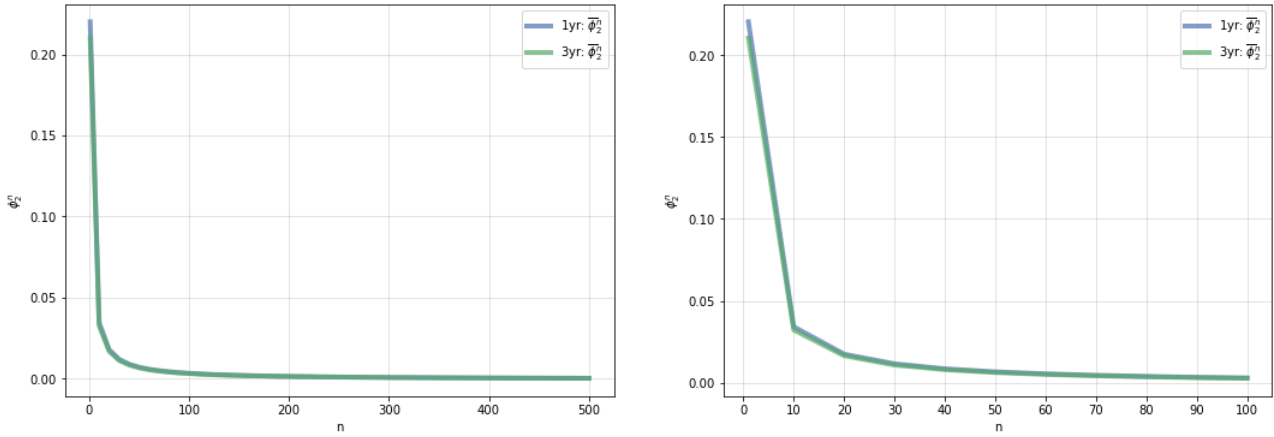
The curves represent the mean ϕ_2 across all the iterations (10,000 each) for $n = 1,10,20\dots 500$ (51 portfolios) across 13 periods from Dec 2014 through Dec 2020.

Figure 3: $\bar{\phi}_3^n$ for all portfolios



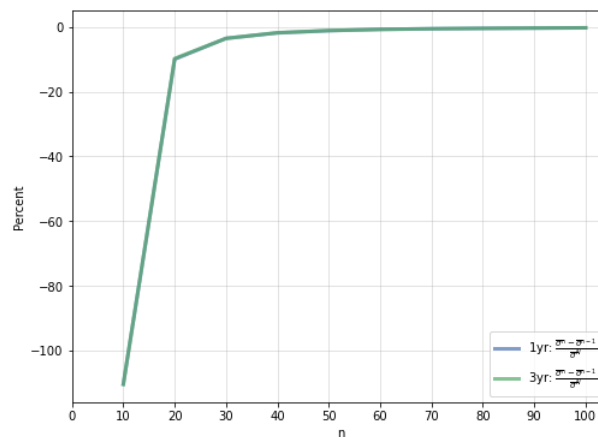
The curves represent the mean ϕ_3 across all the iterations (10,000 each) for $n = 1,10,20\dots 500$ (51 portfolios) across 13 periods from Dec 2014 through Dec 2020.

Figure 4: $\overline{\phi}_4^n$ for all portfolios



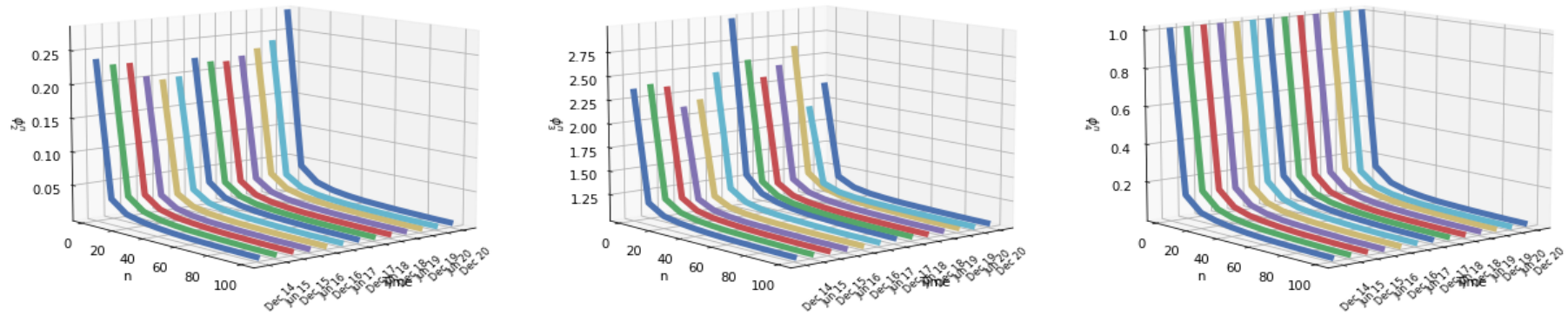
The curves represent the mean ϕ_4 across all the iterations (10,000 each) for $n = 1, 10, 20, \dots, 500$ (51 portfolios) across 13 periods from Dec 2014 through Dec 2020.

Figure 5: Diversification benefit using averages



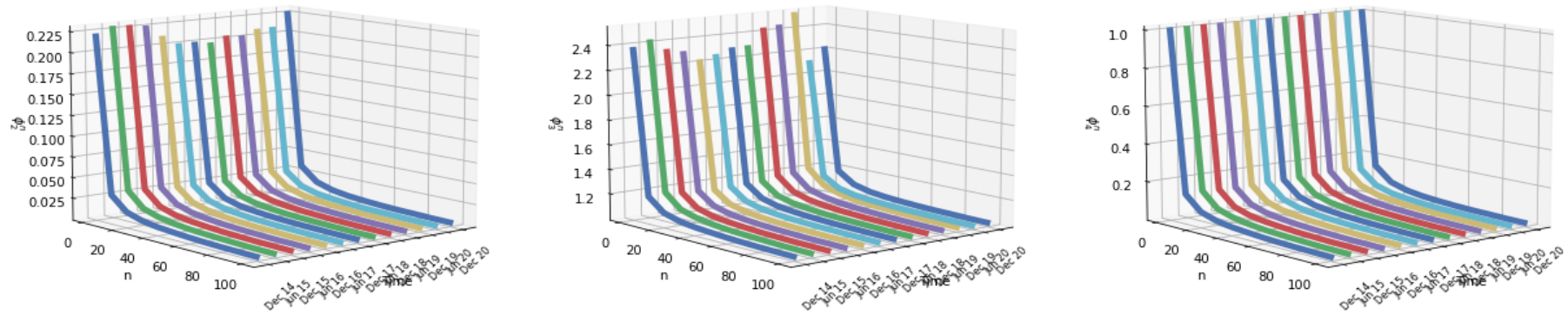
We use $\frac{\overline{\sigma}^n - \overline{\sigma}^{n-1}}{\overline{\sigma}^n}$ to evaluate. As n increases, the measure become asymptotic, implying little additional benefit.

Figure 6: ϕ_2 , ϕ_3 and ϕ_4 over time: : 1-year time-series



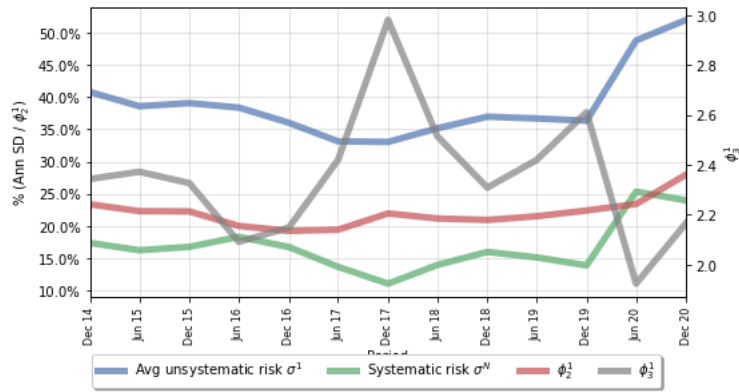
The curves represent the mean standard deviation across all the iterations (10,000 each) for $n = 1,10,20\dots500$ (51 portfolios) for each of the 13 periods from Dec 2014 through Dec 2020.

Figure 7: ϕ_2 , ϕ_3 and ϕ_4 over time: 3-year time-series



The curves represent the mean standard deviation across all the iterations (10,000 each) for $n = 1,10,20\dots500$ (51 portfolios) for each of the 13 periods from Dec 2014 through Dec 2020.

Figure 8: $\bar{\sigma}^1$, σ^N , $\bar{\phi}_2$, and $\bar{\phi}_3$ over time: 1-year time-series



LHS : $\bar{\sigma}^1$, σ^N , $\bar{\phi}_2$, RHS : $\bar{\phi}_3$

Figure 9: Average correlation among securities and with market portfolio: 1-year time-series

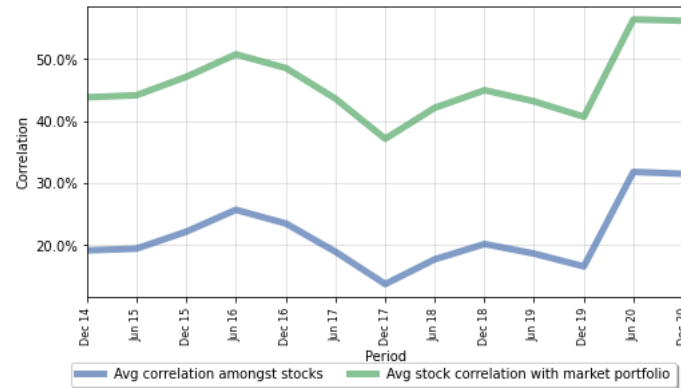
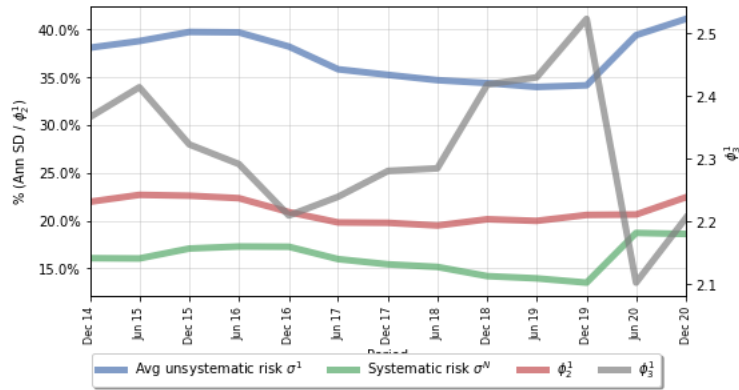
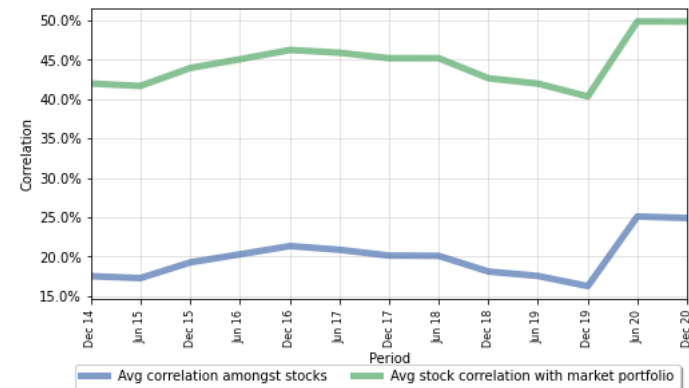


Figure 10: $\bar{\sigma}^1$, σ^N , $\bar{\phi}_2$, and $\bar{\phi}_3$ over time: 3-year time-series



LHS : $\bar{\sigma}^1$, σ^N , $\bar{\phi}_2$, RHS : $\bar{\phi}_3$

Figure 11: Average correlation among securities and with market portfolio: 3-year time-series



The curves on the left panel shows trends over time of $\bar{\sigma}^1$, σ^N , $\bar{\phi}_2$, and $\bar{\phi}_3$. The right hand panel shows the trends in the average correlation among securities and with the market portfolio ($n=500$, equal-weighted) over the same period.

Figure 12: Confidence Limit...Period 31 Dec 2020: 1-year time-series

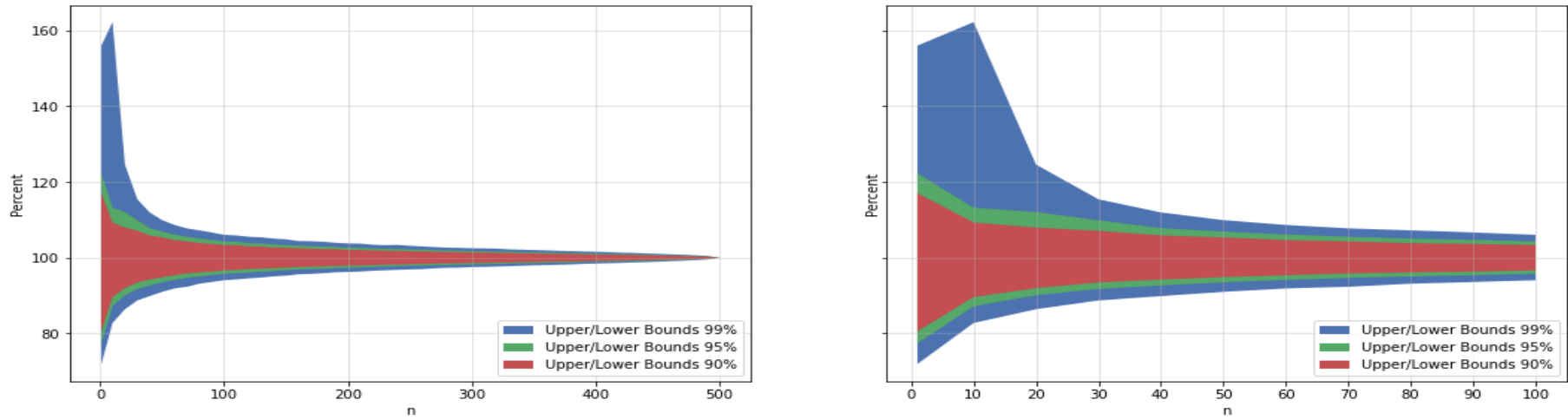
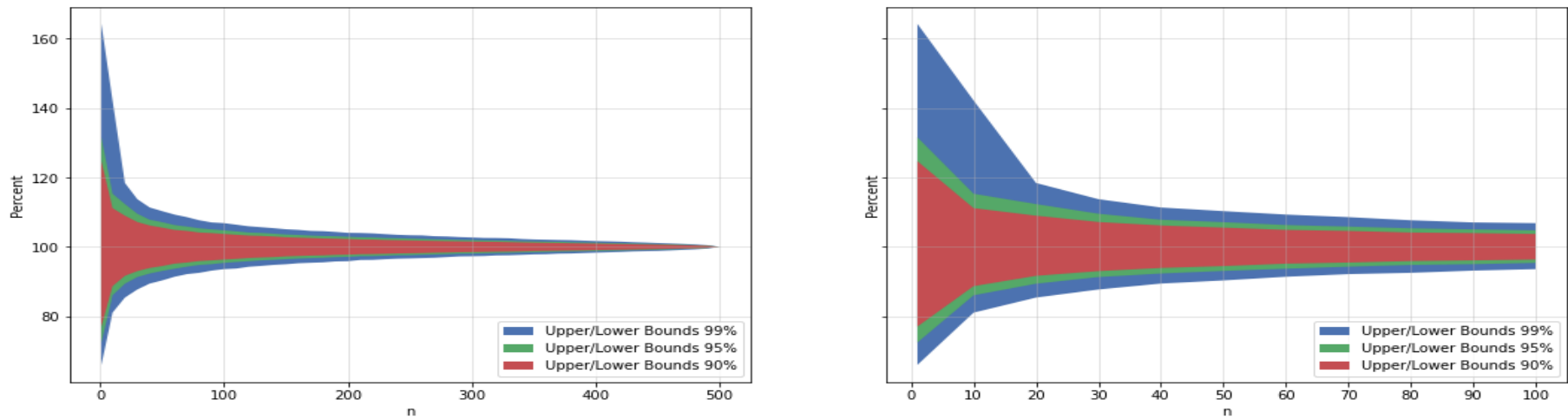


Figure 13: Confidence Limit...Period 31 Dec 2020: 3-yr time-series



The curves represent the upper and lower bounds with 90%, 95% and 99% confidence in the reduction of portfolio risk as n increases from 1 to 500. We scale the pure systematic risk (σ_N) to 100. We show the results for the period ending Dec 2020. The results for the other periods are similar - risk reduces as n increases, the risk reduction is delayed when confidence levels are higher and more volatile time-series have wider initial dispersion than less volatile time-series across confidence levels for the same n .

Figure 14: Number of stock and diversifiable risk reduction over time: 1-year time-series

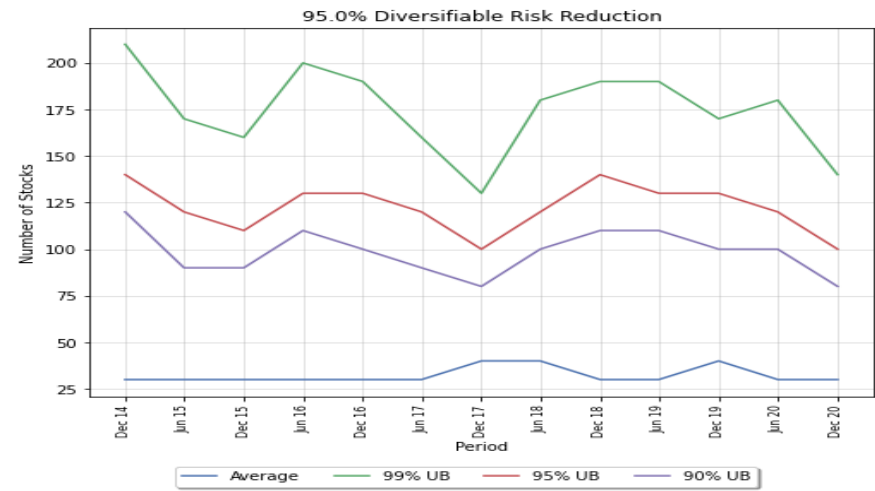
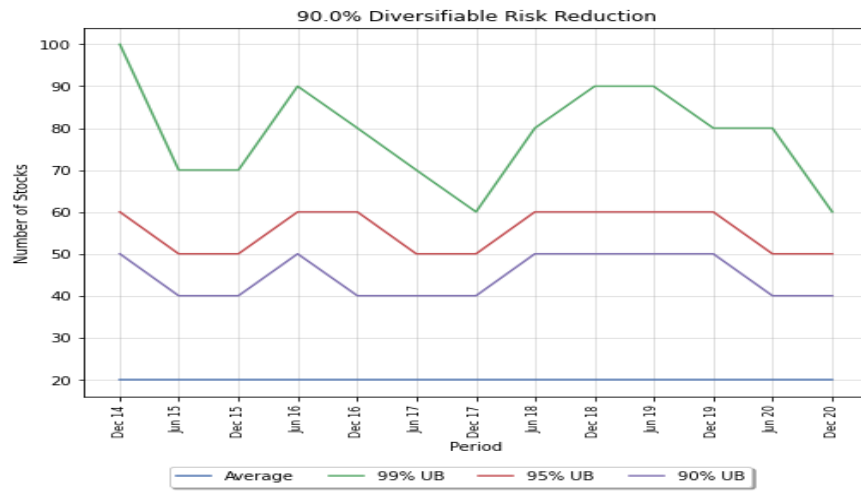
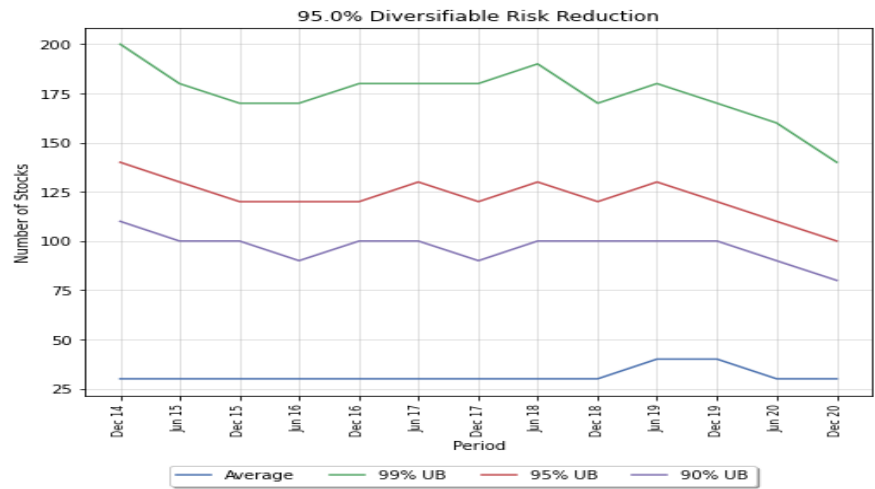
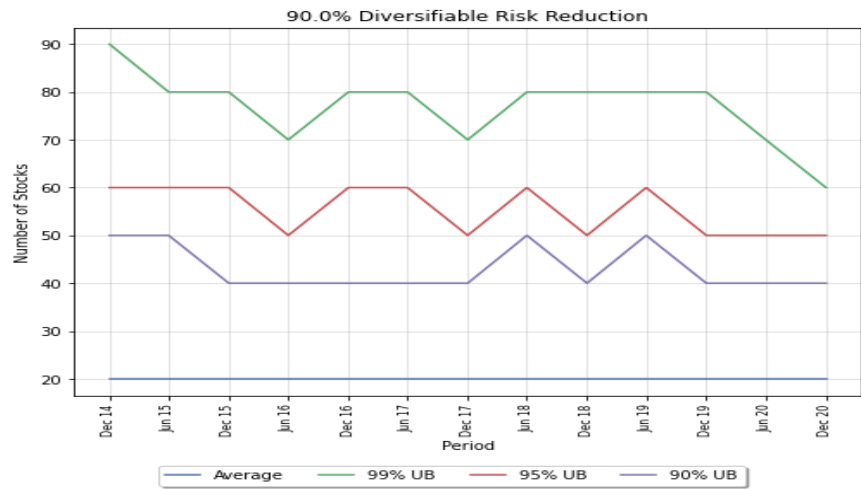


Figure 15: Number of stock and diversifiable risk reduction over time: 3-year time-series



The lines are the number of stock required in an equal weighted portfolio to achieve 90% and 95% diversifiable risk reduction over time using average values and the upper bounds at 99%, 95% and 90% sample coverage.

Table 3: Upper Confidence Limit...Period 31 Dec 2020

N	1yr								3yr							
	$\bar{\phi}_3$				$\bar{\phi}_4$				$\bar{\phi}_3$				$\bar{\phi}_4$			
	99	95	90	Avg	99	95	90		99	95	90	Avg	99	95	90	
1	156.0	122.3	117.0	100.0	204.0	141.4	131.6		164.2	131.6	124.7	100.0	217.5	157.7	145.2	
10	162.2	113.2	109.3	15.3	78.1	28.6	24.7		142.2	115.3	111.1	15.4	56.9	30.5	26.4	
20	124.6	112.0	107.9	7.9	30.9	19.1	15.3		118.4	112.3	109.0	7.9	24.6	19.1	16.1	
30	115.4	109.9	107.1	5.2	19.2	14.2	11.7		113.7	109.5	107.1	5.2	17.3	13.6	11.5	
40	112.0	107.8	105.9	3.9	14.6	10.9	9.2		111.3	107.8	106.2	3.8	13.7	10.6	9.2	
50	109.9	106.9	105.3	3.0	11.8	9.2	7.8		110.3	107.1	105.5	3.1	11.9	9.2	7.8	
60	108.6	106.1	104.7	2.5	10.2	8.0	6.7		109.3	106.3	104.9	2.5	10.4	7.8	6.6	
70	107.7	105.5	104.3	2.1	8.9	7.0	5.9		108.6	105.8	104.6	2.1	9.3	7.0	5.9	
80	107.2	105.0	103.8	1.8	8.2	6.3	5.2		107.6	105.3	104.1	1.8	8.3	6.4	5.3	
90	106.7	104.7	103.6	1.5	7.3	5.7	4.7		107.0	105.0	103.9	1.5	7.5	5.8	4.9	
100	106.0	104.3	103.3	1.4	6.6	5.1	4.3		106.8	104.8	103.7	1.4	7.1	5.4	4.5	
:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	
460	100.9	100.6	100.5	0.0	0.8	0.6	0.5		101.0	100.7	100.6	0.0	0.8	0.6	0.5	
470	100.8	100.6	100.4	0.0	0.7	0.5	0.4		100.8	100.6	100.5	0.0	0.7	0.5	0.4	
480	100.6	100.4	100.4	0.0	0.6	0.4	0.3		100.7	100.5	100.4	0.0	0.6	0.4	0.3	
490	100.5	100.3	100.2	0.0	0.4	0.3	0.2		100.5	100.3	100.3	0.0	0.4	0.3	0.2	
500	100.0	100.0	100.0	0.0	0.0	0.0	0.0		100.0	100.0	100.0	0.0	0.0	0.0	0.0	

* For ϕ_3 the average standard deviation for each portfolio size is scaled to 100. For ϕ_4 , the scaling is done using the average $\bar{\sigma}^1 - \sigma^N$.

Table 4: Required Portfolio Size for Desired Risk Reduction

Risk Reduction*	1yr				3yr			
	Average	99	95	90	Average	99	95	90
90.0%	20	78	55	44	20	76	55	43
95.0%	32	174	122	98	31	174	122	96

* in diversifiable risk defined by ϕ_4 .

The columns reflect the average risk and the upper bounds of the 99%, 95% and 90% intervals. The rows refer to the amount of diversification risk reduced. For instance, for 90% risk reduction, we look for the number of stocks where the ϕ_4 is equal to 10%. The number of stock is the mean number as an integer across all the 13 periods. We are measuring diversification where $n = 1,10,20...500$, so the mean number is just a mathematical average.