Concentration Risk

November 29, 2025

Introduction

One might argue that when forming an optimal portfolio, it would be better to form equal-weighted portfolios rather than value-weighted, which is our default in our analysis in the paper. First of all notice that this is counter to what you observe in practice. Almost all (REF) do value-weighted strategies rather than equal-weighted (see SP500, Russell 2000, MSCI All-world), and for large investors such as pension funds and endowments (REF). Why is that? Well, this is to do with a large cost that is associated in equal-weighting versus value-weighting, which arises due to rebalancing. The only portfolios that do not need to incur trading cost when rebalancing are value-weighted portfolios. To understand this consider as the random returns occur the weights at the end of a trading period automatically line up with the new weights as those that have received a larger weights are those that have grown in size due to the random returns. This is contrary to equal-weights where you have to sell those that have had positive returns and buy those that have decreased in value. Further more this rebalancing cost is growing rapidly as you increase smaller stocks in your investment strategy and the size of you as an investor because price impact is quadratic in the amount you need to buy compared to the outstanding. Therefore it is not feasible for most investors to run an equal-weighted rather than value-weighted investment strategy so focusing on the value-weighted strategy makes sense. For thoroughness we do include some results where we allow for deviations from value-weighting with a cost and confirm our main results.

Hence trading cost plus concentration is why the results of Muir Statman and Elton and Gruber do not hold.

Alternative is then that you can get most of the diversification benefits rather quickly as you include just a few firms, however we argue this is not the right way to view it due to two reasons. First, including more firms do not significantly increase the costs to your portfolio when you do value-weighting due to the arguments given in the previous paragraph. In fact, one can add hundreds of additional stocks with very little portfolio life-time cost. Something that is indeed evident when looking at Index funds and their fees. Nowadays you have have MSCI All Word indixes with fees close to zero (REF and compare to eg SP500). Secondly, this is also not efficient when the largest firms have factor exposures that differ to the market, as we for example see now with seven technology stocks making up 40% of the SP500, as this leads to inefficient diversification across factors, and a larger gain to be made from including more stocks. Hence, instead one should rightly consider how much diversification could be gained from going from the current portfolio to one that included all stocks, or at least hundreds more

Paragraph on concentration in dotcom and also now creating new risks (up to journalists whether priced in), but it is riskier for investors than previous! Even if the average idiosyncratic risk is the same!

Post-Modern Portfolio Theory: Portfolio Optimization with Trading Costs

In this section, we derive the remaining idiosyncratic (or fraction of total volatility) that is left given a random portfolio of N stocks. In the process, we see that this is closely linked to FOMO risk through the square root of the herfindahl index HHI.

1 Setup

Consider an investor that forms a portfolio of N stocks from a universe of K stocks. The investor wants to understand how much idiosyncratic risk is left in his portfolio as he varies N.

We will start off deriving the result for the average investor and then extend it to the case where the investor has a concentrated portfolio, meaning $N < K$, as well as where the investor excludes certain stocks from his portfolio. First off, it is worth noting that as all investors own the whole market by construction, the average investor will have a value-weighted portfolio of all stocks in the market. Normally, this is assumed to mean that the average investor will have no idiosyncratic risk in his portfolio, however we will see that that is not necessarily the case. We will see that the no idiosyncratic risk result comes from stringent assumptions that are at adds with how real markets look like, and relaxing these for more realistic assumptions, means that the average investor will most likely still have idiosyncratic risk in their portfolio.

Investor's problem. Consider an infinite-horizon investor that forms a portfolio of stock weights $\boldsymbol{w}$ from a universe of K stocks. The investor wants to understand how much idiosyncratic risk is left in his portfolio as she varies w_n for the different stocks. The investor targets a specific target portfolio and rebalances to this target portfolio once per period. The trading costs are (quadratic?) and given by c per share of outstanding stocks traded. Hence, the investor wants to choose $\boldsymbol{w}$ to maximize her utility taking into account trading costs. In the following we will assume that the returns are given exogenously and are the same across stocks, and hence the investor only cares about minimizing risk for a given expected return. We can write the investors problem as:

$$\max_{\{\boldsymbol{w}_t\}_{t\geq 0}} U = \sum_{t}^{\infty} \boldsymbol{E}[r_{p,t}] - \frac{\gamma}{2}\mathbb{V}[r_{p,t}] - c\sum_{i=1}^{K} |w_{i,t} - w_{i,t-1}\frac{1+r_{i,t-1}}{1+r_{p,t-1}}|, \tag{1}$$

(or equivalently

$$\max_{\{\boldsymbol{w}_t\}_{t\geq 0}} U = \sum_{t}^{\infty} \boldsymbol{E}[r_{p,t}] - \frac{\gamma}{2}\mathbb{V}[r_{p,t}] - c\sum_{i=1}^{K} |w_{i,t+1} - w_{i,t}\boldsymbol{E}\left[\frac{1+r_{i,t}}{1+r_{p,t}}\right]|. \tag{2}$$

)

To solve this problem, we need to form expectations about the next period weights in relation to next periods' desired weights given the current weights and the returns. We can rewrite this expected difference as:

$$w_{i,t+1} - w_{i,t}\boldsymbol{E}\left[\frac{1+r_{i,t}}{1+r_{p,t}}\right] = w_{i,t+1} - w_{i,t}\left(1 + \frac{\boldsymbol{E}[r_{i,t}] - \boldsymbol{E}[r_{p,t}]}{1+\boldsymbol{E}[r_{p,t}]}\right) = \tag{3}$$

$$= w_{i,t+1} - w_{i,t} - w_{i,t}\left(\frac{\boldsymbol{E}[r_{i,t}] - \boldsymbol{E}[r_{p,t}]}{1+\boldsymbol{E}[r_{p,t}]}\right) = (w_{i,t+1} - w_{i,t}) - w_{i,t}\left(\frac{\boldsymbol{E}[r_{i,t}] - \boldsymbol{E}[r_{p,t}]}{1+\boldsymbol{E}[r_{p,t}]}\right). \tag{4}$$

Here it is worth making two remarks. First, note that if the chosen weights are equal each periods, such as in a equal-weighted portfolio, then there will always be trading costs as the expected returns of the stocks will differ from the expected return of the portfolio, and hence there will always be a need to rebalance. This is because some stocks will have higher expected returns than the portfolio and others lower, and hence to maintain equal weights one needs to sell those that have higher expected returns and buy those with lower expected returns. Second, this is contrary to the special case of a value-weighted portfolio, where the new weights adjust to the new market values of the stocks. In this case, there will actually be no trading costs as the the previous weight times the realized return of each stock is equal to the new market-weight, and hence the trading costs are zero. To see this more clearly, we can define the market weight of stock i at time t as the weight that stock i would have in a value-weighted portfolio:

$$w_{i,t}^m = \frac{v_{i,t}}{v_{p,t}}, \tag{5}$$

where $v_{i,t}$ is the market capitalization of stock i at time t and $v_{p,t}$ is the market capitalization of the portfolio at time t.

Then it is also the case that

$$w_{i,t+1}^m = \frac{v_{i,t}(1+r_{i,t})}{v_{p,t}(1+r_{p,t})} = w_{i,t}^m \frac{1+r_{i,t}}{1+r_{p,t}}. \tag{6}$$

So we can write the expected difference when holding the market weight as:

$$w_{i,t+1}^m - w_{i,t}^m \boldsymbol{E}\left[\frac{1+r_{i,t}}{1+r_{p,t}}\right] = w_{i,t}^m \left(\frac{1+\boldsymbol{E}[r_{i,t}]}{1+\boldsymbol{E}[r_{p,t}]} - 1\right) - w_{i,t}^m \left(\frac{\boldsymbol{E}[r_{i,t}] - \boldsymbol{E}[r_{p,t}]}{1+\boldsymbol{E}[r_{p,t}]}\right) = 0. \tag{7}$$

And hence we can write that for expected difference as a function of deviations from the market weight $\delta w_{i,t} = w_{i,t} - w_{i,t}^m$:

$$w_{i,t+1} - w_{i,t}\boldsymbol{E}\left[\frac{1+r_{i,t}}{1+r_{p,t}}\right] = \delta w_{i,t+1} - \delta w_{i,t}\boldsymbol{E}\left[\frac{1+r_{i,t}}{1+r_{p,t}}\right] + w_{i,t}^m \boldsymbol{E}\left[\frac{1+r_{i,t}}{1+r_{p,t}}\right] - w_{i,t}^m \boldsymbol{E}\left[\frac{1+r_{i,t}}{1+r_{p,t}}\right] = \tag{8}$$

$$= \delta w_{i,t+1} - \delta w_{i,t}\boldsymbol{E}\left[\frac{1+r_{i,t}}{1+r_{p,t}}\right]. \tag{9}$$

Furthermore, as the portfolio return is given by:

$$r_{p,t} = \sum_{i=1}^{K} w_{i,t} r_{i,t}. \tag{10}$$

And portfolio variance is given by:

$$\mathbb{V}[r_{p,t}] = \sum_{i=1}^{K}\sum_{j=1}^{K} w_{i,t}w_{j,t}\mathrm{Cov}(r_{i,t}, r_{j,t}). \tag{11}$$

Having derived the expected portfolio return, variance, and change in weights, we can now return to solving the investors problem. Consider the optimization problem for the choice of stock weights in the portfolio at time t. As mentioned before, we can write the investors problem as:

$$\max_{\{\boldsymbol{w}_t\}_{t\geq 0}} U = \sum_{t}^{\infty} \boldsymbol{E}[r_{p,t}] - \frac{\gamma}{2}\mathbb{V}[r_{p,t}] - c\sum_{i=1}^{K} |w_{i,t+1} - w_{i,t}\boldsymbol{E}\left[\frac{1+r_{i,t}}{1+r_{p,t}}\right]|. \tag{12}$$

This yields the following first-order conditions for the vector of stock weights $\boldsymbol{w}_t$:

$$\nabla_{\boldsymbol{w}_t} U = \boldsymbol{\mu} - \gamma\Sigma\boldsymbol{w}_t - c\cdot \mathrm{sign}\left(\boldsymbol{\delta w}_t\right) \odot \mathbb{E}\left[\frac{\boldsymbol{1}+\boldsymbol{r}_t}{1+r_{p,t}}\right] = \boldsymbol{0}, \tag{13}$$

where $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{r}_t]$ is the vector of expected returns, $\Sigma = \mathrm{Cov}(\boldsymbol{r}_t)$ is the covariance matrix, $\boldsymbol{\delta w}_t = \boldsymbol{w}_t - \boldsymbol{w}_t^m$ is the vector of deviations from market weights, sign$(\cdot)$ is applied element-wise, $\boldsymbol{g}_t = \mathbb{E}\left[\frac{\boldsymbol{1}+\boldsymbol{r}_t}{1+r_{p,t}}\right]$ captures the expected relative gross returns and $\odot$ denotes element-wise multiplication.

Hence, rearranging the first-order conditions gives the optimality condition for the expected returns as:

$$\boldsymbol{\mu} = \gamma\Sigma\boldsymbol{w}_t + c\cdot \mathrm{sign}\left(\boldsymbol{w}_t - \boldsymbol{w}_t^m\right) \odot \boldsymbol{g}_t. \tag{14}$$

This condition states that the expected returns of the stocks must equal the risk-adjusted returns plus a trading cost adjustment that depends on the deviation from market weights.

Key Insight: Value-Weighted Portfolios are Optimal with Trading Costs.

The first-order condition can be rewritten as:

$$\boldsymbol{w}_t^* = \frac{1}{\gamma}\Sigma^{-1}\left(\boldsymbol{\mu} - c \cdot \operatorname{sign}\left(\boldsymbol{w}_t - \boldsymbol{w}_t^m\right) \odot \boldsymbol{g}_t\right). \tag{15}$$

As trading costs increase ($c \to \infty$), the penalty for deviating from market weights dominates, and the optimal solution converges to:

$$\lim_{c\to\infty} \boldsymbol{w}_t^* = \boldsymbol{w}_t^m. \tag{16}$$

That is, when trading costs are sufficiently large, the investor optimally holds the value-weighted market portfolio. This provides a micro-foundation for the prevalence of passive, market-cap-weighted investing: it is the portfolio that minimizes trading costs while still providing market exposure.

Conversely, when trading costs are zero ($c = 0$), the solution reduces to the standard mean-variance efficient portfolio:

$$\boldsymbol{w}_t^*\big|_{c=0} = \frac{1}{\gamma}\Sigma^{-1}\boldsymbol{\mu}, \tag{17}$$

which is the textbook tangency portfolio result. We formalize this in the following proposition:

Proposition 1 (Optimality of Value-Weighted Portfolios under Trading Costs)**.** *Consider an investor solving*

$$\max_{\boldsymbol{w}_t} \ \boldsymbol{\mu}'\boldsymbol{w}_t - \frac{\gamma}{2}\boldsymbol{w}_t'\Sigma\boldsymbol{w}_t - c\left\|\boldsymbol{w}_t - \boldsymbol{w}_t^m\right\|_1 \cdot \bar{g}_t,$$

where $\boldsymbol{\mu}$ *is the vector of expected returns,* Σ *is the covariance matrix,* γ *is risk aversion,* c *is the trading cost parameter,* $\boldsymbol{w}_t^m$ *is the vector of market-cap weights, and* $\bar{g}_t$ *is the average expected relative gross return.*

Then the optimal portfolio weights satisfy:

$$\boldsymbol{w}_t^* = \frac{1}{\gamma}\Sigma^{-1}\left(\boldsymbol{\mu} - c \cdot \operatorname{sign}\left(\boldsymbol{w}_t - \boldsymbol{w}_t^m\right) \odot \boldsymbol{g}_t\right).$$

Moreover:

(i) As $c \to \infty$: $\lim_{c\to\infty} \boldsymbol{w}_t^* = \boldsymbol{w}_t^m$, i.e., the optimal portfolio converges to the value-weighted market portfolio.

(ii) As $c \to 0$: $\boldsymbol{w}_t^*\big|_{c=0} = \frac{1}{\gamma}\Sigma^{-1}\boldsymbol{\mu}$, i.e., the optimal portfolio is the mean-variance efficient (tangency) portfolio.

Interpretation. This result provides a micro-foundation for the prevalence of passive, market-cap-weighted investing: it is the portfolio that minimizes trading costs while still providing market exposure. The value-weighted portfolio is uniquely self-rebalancing, as weights automatically adjust with price changes, eliminating the need for costly trades.

2 Diversification effects under market concentration

Consider the portfolio variance equation derived above. We will now analyze how concentration in the stock market affects the diversification benefits for an average investor holding a value-weighted portfolio.

(OR We now analyze how concentration in the stock market affects the diversification benefits for an average investor holding a value-weighted portfolio.)

The portfolio variance is given by:

$$\mathbb{V}[r_p] = \sum_{i=1}^{K}\sum_{j=1}^{K} w_i w_j \operatorname{Cov}(r_i, r_j).$$

Which for stocks with correlation of ρ, total variance of σ^2 and idiosyncratic variance of

σ_ι^2 can be written as:

$$\mathbb{V}[r_p] = \rho\sigma^2 + \sum_{i=1}^{K} w_i^2 \sigma_\iota^2,$$

where the first term is the systematic variance and the second term is the idiosyncratic variance.

Hence, the idiosyncratic variance can, when the idiosyncratic variance and size are uncorrelated, be rewritten as:

$$\mathbb{V}_\iota[r_p] = \sigma_\iota^2 \sum_{i=1}^{K} w_i^2 = \sigma_\iota^2 HHI = \sigma_\iota^2 h^2 = (\sigma_\iota h)^2,$$

where $HHI = h^2 = \sum_{i=1}^{K} w_i^2$ is the herfindahl index of the portfolio, h is the herfindahl index's square root, and

$$\sigma_{\iota,p} = \sigma_i h.$$

Or more generally, when the idiosyncratic variance and size are correlated, we have that:

$$\mathbb{V}_\iota[r_p] = S,$$

where we have defined S as the size-weighted average of the idiosyncratic variances:

$$S = \sum_{i=1}^{K} w_i^2 \sigma_{\iota,i}^2.$$

We label this S the FOMO or concentration risk, as it captures the risk that arises from concentration in the market due to large stocks with high idiosyncratic risk. And we see how this risk is related to the concentration of idiosyncratic risk in the market, and is therefore strongly related to the herfindahl index of the market.

So in general we have that the amount of idiosyncratic volatility of the portfolio is given by

$$\sigma_p = \sqrt{S},$$

hence the amount of idiosyncratic volatility that remains in the portfolio is directly linked to the concentration of idiosyncratic risk in the market.

In terms of total volatility we have that:

$$\sigma_p = \sqrt{\rho\sigma^2 + S}.$$

We summarize these results in the following proposition:

Proposition 2 (Portfolio Variance Decomposition under Market Concentration)**.** *Consider a value-weighted portfolio of K stocks with weights $\{w_i\}_{i=1}^K$, where each stock i has total variance σ^2, idiosyncratic variance $\sigma_{\iota,i}^2$, and pairwise correlation ρ. Then:*

(i) The portfolio variance decomposes as:

$$\mathbb{V}[r_p] = \underbrace{\rho\sigma^2}_{systematic} + \underbrace{S}_{idiosyncratic},$$

where the concentration risk measure S is defined as:

$$S \equiv \sum_{i=1}^{K} w_i^2 \sigma_{\iota,i}^2.$$

(ii) When idiosyncratic variance is constant across stocks ($\sigma_{\iota,i} = \sigma_\iota$ for all i):

$$S = \sigma_\iota^2 \cdot HHI = (\sigma_\iota \cdot h)^2,$$

where $HHI = \sum_{i=1}^{K} w_i^2$ is the Herfindahl-Hirschman Index and $h = \sqrt{HHI}$.

(iii) The portfolio idiosyncratic volatility is:

$$\sigma_{\iota,p} = \sqrt{S},$$

and total portfolio volatility is:

$$\sigma_p = \sqrt{\rho\sigma^2 + S}.$$

Proof. Starting from the general portfolio variance formula:

$$\mathbb{V}[r_p] = \sum_{i=1}^{K}\sum_{j=1}^{K} w_i w_j \mathrm{Cov}(r_i, r_j).$$

For stocks with pairwise correlation ρ, total variance σ^2, and idiosyncratic variance $\sigma_{\iota,i}^2$, we can decompose the covariance as $\mathrm{Cov}(r_i, r_j) = \rho\sigma^2$ for $i \neq j$ and $\mathrm{Var}(r_i) = \rho\sigma^2 + \sigma_{\iota,i}^2$ for $i = j$. Substituting:

$$\mathbb{V}[r_p] = \rho\sigma^2 \sum_{i=1}^{K}\sum_{j=1}^{K} w_i w_j + \sum_{i=1}^{K} w_i^2 \sigma_{\iota,i}^2 = \rho\sigma^2 + \sum_{i=1}^{K} w_i^2 \sigma_{\iota,i}^2,$$

where we used $\sum_i w_i = 1$. When $\sigma_{\iota,i} = \sigma_\iota$ for all i:

$$\mathbb{V}_\iota[r_p] = \sigma_\iota^2 \sum_{i=1}^{K} w_i^2 = \sigma_\iota^2 \cdot HHI = (\sigma_\iota h)^2.$$

□

Interpretation. The measure S captures *concentration risk*: the undiversifiable idiosyncratic risk that arises when large stocks have high idiosyncratic volatility. Unlike the textbook case where idiosyncratic risk vanishes as $N \to \infty$, concentration in market capitalizations prevents full diversification, leaving residual idiosyncratic risk proportional to the HHI. We label S the "FOMO risk" as it captures the risk investors face from being concentrated in a few large, volatile stocks.

3 Question of how many N needed to get 5% of original ivol under equally weighted assumption

Having set up the general framework, we now show how to answer questions about diversification gains and losses and in the process show the conecction from concentration risk to FOMO risk to portfolio risk.

Specifically, we are curious about the questions of the like: What amount of diversification is lost when going from K to N stocks where $K >> N$ want that, which is equivalent to asking whether there is diversification loss when going from investing in the whole universe to only a subset N of stocks.

To answer these questions, it turns out to be useful to define the fraction of total idiosyncratic volatility σ_ι that is diversified in expectation at a given N. Let this be given by $\alpha(n)$. One can then compare $\alpha(N)$ to $\alpha(K)$ to get the idea of the loss of diversification $(\alpha(N) - \alpha(K))\,\sigma_\iota$.

Let $\alpha(N)$ be defined by

$$\sigma_i = \alpha(N)\sigma_\iota, \tag{18}$$

where σ_i is the remaining idiosyncratic risk in the portfolio.

In general, we have that:

$$\sigma_i = \frac{\sigma_\iota}{\sqrt{N}} = \alpha\sigma_\iota \iff \alpha = \frac{1}{\sqrt{N}} \iff N = \alpha^{-2}. \tag{19}$$

This is useful, as we can ask, for example, how many stocks are needed to reach 95% diversification ($\alpha = 0.05$):

$$N^* = 0.05^{-2} = 400.$$

Or what is the loss from going from a portfolio of 4000 stocks to one with 400:

$$(\alpha(400) - \alpha(4000))\, \sigma_\iota = \left(\frac{1}{\sqrt{500}} - \frac{1}{\sqrt{4000}} \right) \sigma_\iota = (0.050 - 0.016)\sigma_\iota = 0.034\sigma_\iota.$$

Meaning we lose 3.4 percentage points of diversification benefits.

Or 4000 to 1000:

$$(\alpha(1000) - \alpha(4000))\, \sigma_\iota = \left(\frac{1}{\sqrt{1000}} - \frac{1}{\sqrt{4000}} \right) \sigma_\iota = (0.032 - 0.016)\sigma_\iota = 0.016\sigma_\iota.$$

Meaning we still lose 1.6 percentage points of diversification benefits from going from 4000 to 1000 stocks.

A different question would be what is the loss from going from a portfolio of 500 stocks to a fully diversified portfolio ($K \to \infty$)?

$$(\alpha(500) - \alpha(\infty))\, \sigma_\iota = (0.045 - 0)\sigma_\iota = 0.045\sigma_\iota.$$

Hence, we lose 4.5 percentage points of potential diversification benefits from only having 500 stocks.

Now we will relax the assumption that the stocks in the market have the same size. The following section will consider the case where stock sizes are more similar to how they look like in reality.

3.1 Maybe add here for the case of concentrated markets?

It is simply just h instead of $1/\sqrt{N}$. Copy text from appendix here.

4 Maybe still do equivalent for how S is expected to change with K to get alpha s a function of N/K?

5 Putting it all together: The different channels

If we want to compare to total volatility we can see from Equation 38 that we can use $\sigma_\iota = \sqrt{1-\rho}\sigma$ and hence

$$\alpha(N)' = \frac{\alpha(N)}{\sqrt{1-\rho}}, \tag{20}$$

which as $\sqrt{1-\rho} \leq 1$ means the diversification level is smaller compared to total volatility, and that it decreases relatively slower in terms of total volatility as N is increased as compared to the case with no systematic volatility or in terms of its idiosyncratic volatility. When this is the case, more stocks (higher N) is needed to achieve the same amount of diversification (ie an equally low α).

Furthermore, when there is $\sigma_\beta > 0$ we have from Equation 30 that

$$\alpha(N)'' = \frac{\alpha(N)}{\sqrt{1-\rho}}\sqrt{1+\varsigma(\sigma_\beta)}, \tag{21}$$

where ς is given by Equation 31.

Putting it all together, we have that

$$\alpha''(N) = h(N)\sqrt{\frac{1+\varsigma(\sigma_\beta)}{1-\rho}} \sim \begin{cases} \frac{v_1}{\ln N}\sqrt{\frac{1+\varsigma(\sigma_\beta)}{1-\rho}}, & \zeta = 1 \quad \text{(Zipf's law)}, \\ \frac{v_\zeta}{N^{1-1/\zeta}}\sqrt{\frac{1+\varsigma(\sigma_\beta)}{1-\rho}}, & 1 < \zeta < 2, \\ \frac{c}{\sqrt{N}}\sqrt{\frac{1+\varsigma(\sigma_\beta)}{1-\rho}}, & \zeta \geq 2. \end{cases} \tag{22}$$

So, in total, we see that the diversification loss from going to a smaller portfolio N from a fully diversified portfolio ($\alpha''(N)\sigma$) is given by N, the total level of volatility σ, the relative size of idiosyncratic volatility to total volatility (described by ρ), the volatility of average factor betas σ_β, and the degree of stock market concentration h (or equivalently the degree of the power law distribution, ζ).

6 Estimation of the channels

Which can be used to show hypothethical scenarios of how α would look like historically with different values of correlation or concentration. Either use the theoretical coefficients and historical variables or estimate the coefficients in a regression. (Plus possible to take logs and check if coefficients are significant (whether to include channel) and should be around the ones below).

$$log\alpha''(N) = log(h) - 1/2log(1-\rho) + 1/2log(1+\varsigma(\sigma_\beta))$$

Which can be used to show hypothethical scenarios of how α would look like historically with different values of correlation or concentration. Eg if ρ was still its original value.

We can then run a variance decomposition of the log of Equation 22 to see which of these channels has driven changes in $\alpha''(N)$ over time.

$$\begin{aligned}
\mathrm{Var}\big[\log\alpha''(N)\big] = {} & \mathrm{Var}\big[\log h(N)\big] + \tfrac{1}{4}\mathrm{Var}\big(\log\big(1+\varsigma(\sigma_\beta(N))\big)\big) && (23)\\
& + \tfrac{1}{4}\mathrm{Var}\big(\log\big(1-\rho(N)\big)\big) && (24)\\
& + \mathrm{Cov}\big(\log h(N), \tfrac{1}{2}\log\big(1+\varsigma(\sigma_\beta(N))\big)\big) && (25)\\
& - \mathrm{Cov}\big(\log h(N), \tfrac{1}{2}\log\big(1-\rho(N)\big)\big) && (26)\\
& - \tfrac{1}{2}\mathrm{Cov}\big(\log\big(1+\varsigma(\sigma_\beta(N))\big), \log\big(1-\rho(N)\big)\big)\,. && (27)
\end{aligned}$$

Then, by the delta method:

$$\mathrm{Var}\big[\alpha''(N)\big] \approx \mathbb{E}\big[\alpha''(N)\big]^2 \cdot \mathrm{Var}\big[\log \alpha''(N)\big]. \tag{28}$$

7 Descriptive Figures

In this section we show descriptive figures of our variables of interest.

Figure 1 shows the market metrics over time. The underdiversification metric $\alpha''(N)$ is shown in panel (a). We see that it has had peaks and throughs over time, indicating that diversification at times becomes more difficult. Underdiversification from 500 stocks peaks in the 2000's. We also plot the three channels that drive underdiversification over time. Panel (b) shows the average correlation between the stocks in the market. We see that correlation has increased over time, especially during crisis periods of the financial crisis and covid. Panel (c) shows the concentration metric h (the square root of the herfindahl index). We see that concentration has increased over time, especially during the dotcom bubble and the recent tech bubble. Panel (d) shows the idiosyncratic volatility concentration metric S. We see that it has to a large degree follows h however it peaks even more aggresively during the dotcom bubble. Panel (e) shows the market volatility. We see that it has had peaks during crisis periods.

8 Results

Here we document the results from our diversification decomposition tests.

Table 1 shows the results from the decomposition regression between the log-log and linear model specifications.

Table 2 shows the results from the variance decomposition of underdiversification into the different channels. We can see that concentration (S) is the main driver of changes in underdiversification over time, explaining around 57% of the variance. Correlation also plays

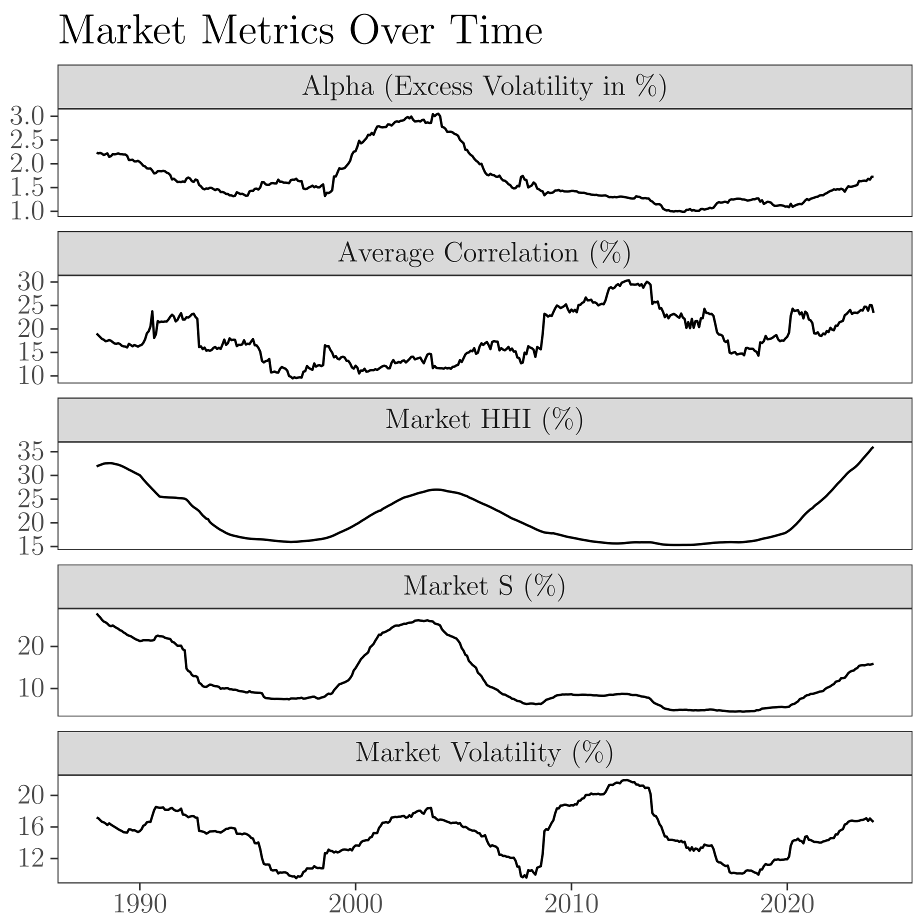

Figure 1: **Market metrics over time**. Panel (a) shows the underdiversification metric $\alpha''(N)$ for $N = 500$ stocks. Panel (b) shows the average correlation between stocks in the market. Panel (c) shows the concentration metric h (the square root of the herfindahl index). Panel (d) shows the idiosyncratic volatility concentration metric S. Panel (e) shows the market volatility. All metrics are computed on a monthly basis from January 1926 to December 2022 using CRSP data. In a specific month, the metrics are computed as the historical values over the past 5 years requiring at least 3 years of data.

Table 1: Decomposition of Underdiversification Metric $\alpha''(N)$.

	Log-Log	Linear
Constant	0.470***	1.381***
	(0.050)	(0.037)
Market Vol	0.295***	0.070***
	(0.040)	(0.005)
Avg Corr	-0.556***	-0.068***
	(0.026)	(0.003)
Market S	0.340***	0.040***
	(0.013)	(0.002)
Num.Obs.	433	433
R^2	0.928	0.916
Adj. R^2	0.928	0.916
Model	Log-Log	Linear

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors in parentheses.

Table 2: Variance Decomposition of log(Alpha) Regression.

Variable	**Relative Importance**	**Percentage (%)**
Market Volatility	0.119	11.9
Average Correlation	0.313	31.3
Market Concentration (S)	0.568	56.8

This table reports the relative importance of each explanatory variable in the log(Alpha) regression using the Lindeman, Merenda and Gold (LMG) method. Relative importance measures the proportionate contribution of each variable to R-squared.

a significant role, explaining around 31% of the variance. The market volatility channel plays a minor role.

Figure 8 visualizes these findings of both high model fit and concentration relevance by showing the predicted versus actual measure of undiversification α, which correspond to the results from Table 1 together with the plot of idiosyncratic volatility concentration S over time from Figure 1.

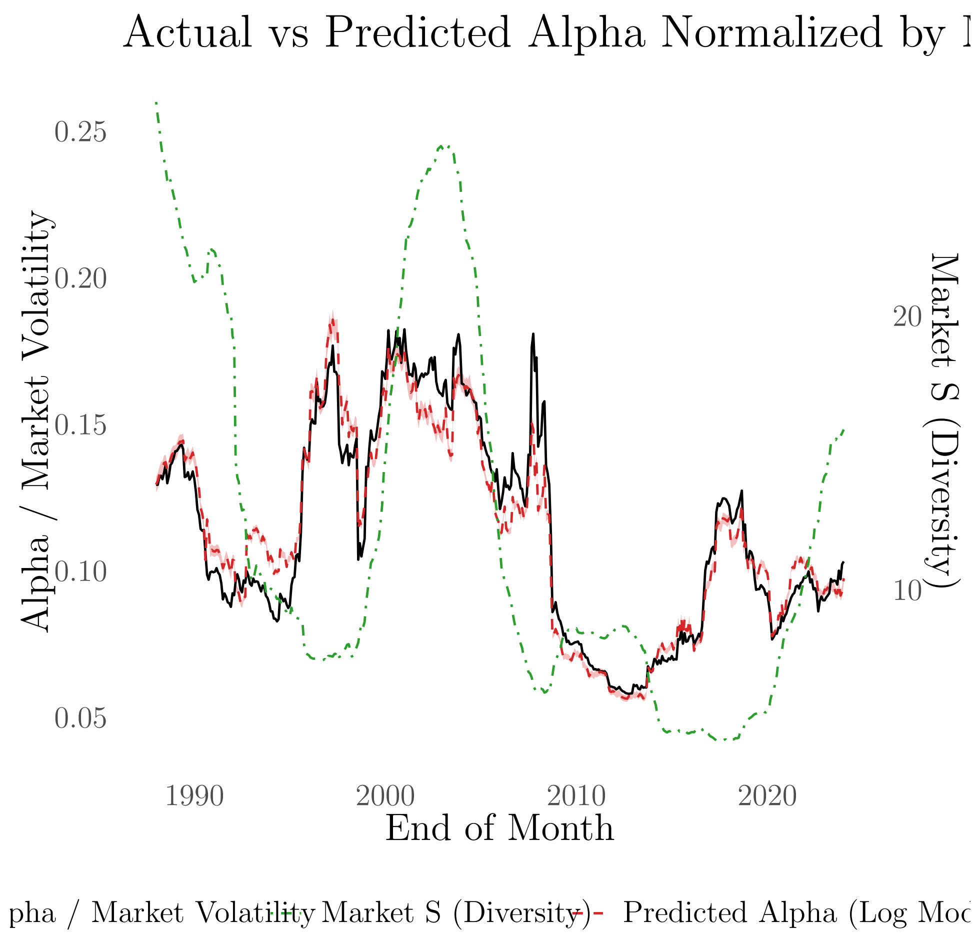

Figure 2: **Actual vs predicted underdiversification metric** $\alpha''(N)$. The actual underdiversification metric $\alpha''(N)$ (black line) is plotted against the historical market S measure and the predicted underdiversification metric $\hat{\alpha}''(N)$ (red line) computed from Equation 22 using historical values of correlation, concentration (S), and market volatility. The S measure is a measure of concentration of idiosyncratic volatility. The metrics are computed on a monthly basis from January 1926 to December 2022 using CRSP data. In a specific month, the metrics are computed as the historical values over the past 5 years requiring at least 3 years of data.

Appendix

9 The minimum idiosyncratic variance portfolio

An object of interest may be the minimum idiosyncratic variance portfolio. This portfolio will be given as the solution to the following optimization problem:

$$\min_w \; w'\Sigma_\iota w,$$

where Σ_ι is an idiosyncratic covariance matrix of stock returns, meaning that it is a diagonal matrix with stock i's idiosyncratic volatility $\sigma_{\iota i}$ given at column i, row i with all off-diagonal entries equal to zero. The solution to this problem is simply given as

$$w^* = \frac{\Sigma_\iota^{-1} 1}{1'\Sigma_\iota^{-1} 1}.$$

which simplifies as

$$w_i^* = \frac{\sigma_{\iota i}^{-2}}{\sum_{j=1}^{N} \sigma_{\iota j}^{-2}}.$$

10 Old version: Factor model setup

Assets. Consider the average investor's portfolio, meaning a weighted portfolio of N stocks, where stock i has return[1] w_i and return r_i. The stock returns follow a factor model with F factors:

$$r_i = \beta_i' f + \varepsilon_i, \qquad \mathbb{E}[\varepsilon_i] = 0, \qquad \mathbb{V}[\varepsilon_i] = \sigma_\iota, \qquad \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0 \;\; (i \neq j).$$

[1]Hence the weights are equal to the stocks' relative market weight, and could be equal if all stocks have the same size (the standard assumption), but will always be value-weighted. Another reason to consider the value-weighted portfolio is that it is the only one that does not incur trading costs when rebalancing, and hence it could be further micro-founded with an investor optimization problem with very large transaction costs.

And factors' $f \in 1, ..., F$ returns follow

$$f = \mu_f + \varepsilon_f, \qquad \mathbb{E}[\varepsilon_f] = 0, \qquad \mathbb{V}[\varepsilon_f] = \sigma_f, \qquad \mathrm{Cov}(\varepsilon_f, \varepsilon_g) = 0 \ \ (f \neq g).$$

Then portfolio returns follow

$$r_p = \sum_i w_i r_i = \sum_i w_i \beta_i' f + w_i \varepsilon_i.$$

As an example consider the equal-weighted portfolio. This then has the following expected return:

$$\mathbb{E}[r_p] = \sum_i w_i \mathbb{E}_i[r_i] = N\frac{1}{N}\mathbb{E}_i[\beta_i]'\mu_f = \mu_\beta' \mu_f = F\bar{\mu_\beta}\bar{\mu_f},$$

when β_i and f are uncorrelated cross-sectionally for the portfolio stocks.

The portfolio standard deviation is likewise given as

$$\sigma_p = \sqrt{\mathbb{V}[\sum_i w_i r_i]} = \sqrt{\sum_i w_i^2 \mathbb{V}[r_i]} = \sqrt{\frac{1}{N^2}[N\sigma_\iota^2 + \mathbb{V}[\beta_i' f]]}.$$

Which, using the law of total variance for the product of two random variables, can be written as

$$\sigma_p = \frac{1}{N}\sqrt{N\sigma_\iota^2 + N^2 \mu_\beta^{2\prime}\sigma_f^2 + N^2\sigma_\beta^{2\prime}\mu_f^2 + N^2\sigma_\beta^{2\prime}\sigma_f^2} = \sqrt{\frac{\sigma_\iota^2}{N} + \mu_\beta^{2\prime}\sigma_f^2 + \sigma_\beta^{2\prime}\mu_f^2 + \sigma_\beta^{2\prime}\sigma_f^2},$$

$\iff$

$$\sigma_p = \sqrt{\frac{\sigma_\iota^2}{N} + F\bar{\mu}_\beta^2\sigma_f^2 + F\sigma_\beta^2\bar{\mu}_f^2 + F\sigma_\beta^2\sigma_f^2}. \tag{29}$$

Or, equivalently, in terms of portfolio variance is given as

$$\mathbb{V}_p = \frac{\sigma_\iota^2}{N} + \mu_\beta^{2\prime}\sigma_f^2 + \sigma_\beta^{2\prime}\mu_f^2 + \sigma_\beta^{2\prime}\sigma_f^2 = \frac{\sigma_\iota^2}{N} + F\bar{\mu}_\beta^2\sigma_f^2 + F\sigma_\beta^2\bar{\mu}_f^2 + F\sigma_\beta^2\sigma_f^2,$$

when σ_β is the same for all factors (else you get higher-order variance terms) and σ_β and μ_f in addition to β_i and f are uncorrelated cross-sectionally for the portfolio stocks and β_i are iid across stocks with mean μ_β and standard deviation σ_β.

Here, we see a new addition which is that the variance of the stock betas in the portfolio matters for the portfolio volatility. The addition is the last two terms. To minimize the risk, the portfolio betas to the different factors should be as similar as possible. We can illustrate the increase in risk from poor factor diversification as follows:

$$\sigma' = \sigma\sqrt{1 + \varsigma(\sigma_\beta)}, \tag{30}$$

where

$$\varsigma(\sigma_\beta) \equiv \frac{F\sigma_\beta^2\bar{\mu}_f^2 + F\sigma_\beta^2\sigma_f^2}{\sigma_\iota^2 + F\bar{\mu}_\beta^2\sigma_f^2}. \tag{31}$$

Hence, $\varsigma(\sigma_\beta)$ shows how much the total risk increases compared to the case with optimal diversification across factors (σ') as σ_β increases.

Of course, an effect could be that due to exclusion one does not have a beta to some factors and more to others and hence the diversification benefit across those factors decreases. This would then be captured from an increase in the σ_β as you move away from the optimal diversification across factors that you get when your beta to all factors is the same in which case the last two terms go to zero. In the worst case where you have all your exposure to just one factor the last three terms become $F^2\sigma_f^2$, which is the diversified case.

11 Effective diversification for concentrated portfolios

So far we have considered the case for equally weighted stocks. This section departs from this assumption and gives particular attention to the case of value-weighted portfolios. (Still with weights uncorrelated with factor exposures).

Diversification effects for power law distributed market capitalizations

Recall from Equation 18 that

$$\sigma_i = \alpha(N)\sigma_\iota,$$

where σ_i is the remaining idiosyncratic risk of the portfolio, σ_ι the original idiosyncratic risk and $\alpha(N)$ describes what fraction of the original idiosyncratic risk is left in the portfolio. Then note that Gabaix introduces his own term h, which he defines as:

$$\sigma_i = \sigma h, \tag{32}$$

and that this h corresponds to our $\alpha(N)$. Furthermore, in his paper he shows that this h is the square root of the size herfindahl of the portfolio:

$$h = \left[\sum_{i=1}^{N} w_{it}^2\right]^{1/2}, \tag{33}$$

where w_{it} is the portfolio-weight of stock i in the portfolio when all stocks have the same size for their idiosyncratic risk. For value-weighted portfolios (the default) this will be given as

$$w_{it} = \frac{v_{it}}{v_{pt}},$$

where v_{it} is the market capitalization of stock i at time t and v_{pt} is the market capitalization of the portfolio.

Another way to understand h, which may be simpler, is to introduce a new term which we label effective N, $\tilde{N}$. Correspondingly, it is defined as:

$$\sigma_i = \frac{\sigma_\iota}{\sqrt{\tilde{N}}}, \tag{34}$$

We can see from comparison between Equations 32, 33, and 34, that $\tilde{N}$ is the inverse of the herfindahl index (HHI) and the inverse of h^2.

From equivalence to the standard case we see that to we get that

$$\tilde{N} = \alpha^{-2} = h^{-2}. \tag{35}$$

and we return to our original finding that

$$h = \alpha.$$

Hence, the number of stocks needed for some ratio α of diversification is hence related to h. We will test this relationship empirically in the results section.

Diversification effects for different degrees of power-distributions (changing concentrations)

Therefore, to understand the diversification loss from adjusting the number of stocks in your portfolio in the case for power law distributed market-values, it is interesting to ask how h, the square root of the HHI, depends on N to compare $\tilde{N}$ with N. The diversity degree $\alpha(N)$ (or the square root of the herfindahl, h) is given from Gabaix (2011 ECMA) as:

$$\alpha(N) = h \sim \begin{cases} \dfrac{v_1}{\ln N}, & \zeta = 1 \quad \text{(Zipf's law)}, \\[2ex] \dfrac{v_\zeta}{N^{1-1/\zeta}}, & 1 < \zeta < 2, \\[2ex] \dfrac{c}{\sqrt{N}}, & \zeta \geq 2. \end{cases} \tag{36}$$

Here v_ζ is N-invariant (a constant for $\zeta > 2$, or a Lévy-stable random variable for $\zeta \leq 2$). Which of these is the case we can estimate empirically.

Equivalently, we can write it as effective N to compare our diversification with the case for equally distributed stocks:

$$\tilde{N}(N) \sim \begin{cases} \dfrac{\ln^2 N}{v_\zeta^2}, & \zeta = 1 \quad \text{(Zipf's law)}, \\ \dfrac{N^{2-2/\zeta}}{v_\zeta^2}, & 1 < \zeta < 2, \\ \dfrac{N}{c^2}, & \zeta \geq 2. \end{cases} \tag{37}$$

Conclusions

We see that when portfolios are not equally weighted with idiosyncratic risks of equal size we have to use effective N ($\tilde{N}$) instead of N. Furthermore:

- If stock values are Zipf-distributed ($\zeta = 1$), the effective number of stocks for diversification benefits only grows very slowly as $\ln^2 N$.
- If values have a fat tail $1 < \zeta < 2$, diversification reduces volatility at rate $N^{-(1-1/\zeta)}$, slower than $1/\sqrt{N}$.
- If $\zeta \geq 2$ (finite variance of V_i), we recover the classical $1/\sqrt{N}$ diversification rate.

12 Idiosyncratic versus total risk considerations

By increasing the number of stocks in our portfolio we can decrease the amount of idiosyncratic risk the portfolio entails but in general not the systematic risk, so we focus on how many stocks it takes for only some fraction α of the original idiosyncratic risk to remain. However, there is an easy relation from the undiversified idiosyncratic risk to the total risk given through the average correlation of stocks (the ratio of systematic to total volatility), namely that:

$$\sigma_i = \alpha \sigma_\iota = \alpha \sqrt{1-\rho}\sigma, \tag{38}$$

where σ_i is the remaining idiosyncratic risk and the proof of which we give in the appendix.

Relationship between idiosyncratic and total volatility. This can also be rewritten in terms of average total stock volatility σ as $[\sigma_\iota^2 = (1-\rho)\sigma^2]$ correlation and total standard deviation defined as:

$$\rho = \frac{\mu_\beta^2\sigma_f^2 + \sigma_\beta^2\sigma_f^2}{\sigma_\iota^2 + \mu_\beta^2\sigma_f^2 + \sigma_\beta^2\sigma_f^2},$$

$$\sigma = \sqrt{\sigma_\iota^2 + \mu_\beta^2\sigma_f^2 + \sigma_\beta^2\sigma_f^2},$$

$$\text{where } \sigma_i = \frac{1}{\sqrt{N}}\sigma_\iota.$$

So

$$\rho\sigma^2 = \mu_\beta^2\sigma_f^2 + \sigma_\beta^2\sigma_f^2$$

and

$$\sigma_p = \sqrt{\frac{\sigma_\iota^2}{N} + \sigma_f^2\mu_\beta + \sigma_f^2\sigma_\beta^2} = \sqrt{\frac{\sigma_\iota^2 + \sigma_f^2\mu_\beta + \sigma_f^2\sigma_\beta^2}{N} - \frac{\sigma_f^2\mu_\beta + \sigma_f^2\sigma_\beta^2}{N} + \sigma_f^2\mu_\beta + \sigma_f^2\sigma_\beta^2},$$

$$= \sqrt{\frac{\sigma_\iota^2 + \sigma_f^2\mu_\beta + \sigma_f^2\sigma_\beta^2}{N} + \frac{N-1}{N}(\sigma_f^2\mu_\beta + \sigma_f^2\sigma_\beta^2)} = \sqrt{\frac{\sigma^2}{N} + \frac{N-1}{N}\rho\sigma^2}.$$

And

$$\rho = 1 - \frac{\sigma_\iota^2}{\sigma^2} = \frac{\sigma - \sigma_\iota^2}{\sigma} = \frac{\sigma_{sys}^2}{\sigma^2}.$$