

The Downside Risk of Heavy Tails Induces Low Diversification

Namwon Hyung and Casper G. de Vries*

University of Seoul,

Tinbergen Institute, Erasmus University Rotterdam[†]

May 29, 2013

Abstract

Actual portfolios contain fewer stocks than are implied by standard financial analysis that balances the costs of diversification against the benefits in terms of the standard deviation of the returns. Suppose a safety first investor cares about downside risk and recognizes the heavy tail feature of the asset return distributions. Then we show that optimal portfolio sizes are smaller than traditional correlation based diversification analysis suggests.

*We would like to thank Jay M. Chung for introducing us to the paper by Statman (1987). We are grateful to Chen Zhou, Laurens de Haan and M. Statman for helpful discussions and for the remarks by participants at the conference in honor of Mike Wickens at the University of York and seminar participants at the Goethe University, House of Finance.

[†]Mailing address: Casper G. de Vries, Department of Economics, Erasmus Univeristy Rotterdam, PO Box 1738, 3000 DR, Rotterdam, The Netherlands, email <cdevries@ese.eur.nl>, Namwon Hyung, Faculty of Economics, The University of Seoul, Seoul, 130-743, Korea, email <nhyung@uos.ac.kr>.

Keywords: Portfolio diversification, downside risk, heavy tails

JEL code: G0, G1, C2

1 Introduction

The level of diversification in investor's equity portfolios presents a puzzle to the mean-variance based portfolio analysis. The older literature focussed on the benefits of diversification that derive from the reduction in risk. If the risk is measured by the variance of the portfolio return, typical portfolio sizes comprise dozens of different assets. This gives a one sided view of diversification, as it only presents the benefits, not the costs and hence may imply overdiversification.

In standard financial analysis the optimal portfolio size should follow from balancing at the margin the cost of trading and holding different securities against the benefits of diversification. In two perceptive papers Statman (1987, 2004) was the first to explicitly consider the trade-off between the costs and benefits of diversification. To compare the benefits with the costs, Statman translated the benefits into the money metric via the capital market line. Still, his analysis implied a higher level of diversification than is observed in reality. Statman (2004) discusses the importance of behavioral aspects of the investor's decision process for closing the gap.

In this paper, we take the analysis further by explicitly recognizing the behavioral concern for downside risk in the investor's evaluation of portfolios and by incorporating the empirical fact that the loss return distribution is fat tailed distributed. Traditional diversification analysis proceeds on the basis of naive

portfolio selection whereby stocks are selected at random. This can already be regarded as a form of behavioral analysis, where the investor acts under ignorance (Elton and Gruber, 1978); see also Benartzi (2001), Benartzi and Thaler (2001) for evidence in this direction. DeMiguel, Garlappi and Uppal (2009) show that the naive strategy performs about equally well to more refined strategies that require estimation of some of the moments of asset returns. As is well known, the mean log return is notoriously difficult to estimate reliably. Nevertheless, we also consider a more refined diversification strategy that requires estimation of the asset betas.

Several recent papers have shown the relevance for downside risk in financial analysis. For example Ang, Chen and Xing (2006), and Harvey and Siddique (2000) provide evidence that downside risk is priced. We cast the concern for downside risk in Arzac and Bawa's (1977) equilibrium setting of a market with safety-first investors. The downside risk measure is made operational in two alternative ways. One measure is the zero-th lower partial moment or Value-at-Risk (VaR) measure. Given its proliferation in banking and insurance, it is the most direct evidence for the downside risk concern. We also consider expected shortfall as an alternative measure, given its theoretical appeal of subadditivity.

We first investigate the empirical relevance of the downside risk measures and the global variance measure by examining historical data. The variance measure combined with the safety first concern still requires large portfolios. This changes when we turn to the downside risk measures. The naive investor who applies the random stock picking rule and applies the VaR measure, requires about fifty

stocks in his portfolio at a risk level on event that occurs once every five years. The more sophisticated investors who is able to randomly select from the subset of low beta stocks needs a much smaller portfolio of about ten to fifteen stocks.

Subsequently, we back these claims with theoretical results using the fact that return distributions are heavy tailed and contrast this with the counterfactual assumption of normally distributed returns. If agents display concern for downside risk, it becomes important to model this risk adequately. It is by now a well recognized stylized fact that tail risk is not normal. Rather, this risk is fat tailed distributed, see e.g. Jansen and de Vries (1991). We show theoretically that the speed of diversification under fat tailed distributed (loss) returns is, perhaps somewhat surprisingly, higher than under normality. The intuition is as follows. At a given risk level, under normality diversification changes by the square root of the number of assets, since this is how the standard deviation changes. In the case of heavy tails, the tail risk is shaped like the power of the Pareto distribution. Diversification then lowers the scale of the tail risk at the rate of this power minus one (the power is unaffected). Holding the risk level constant, this implies that diversification reduces the loss level at a rate equal to one minus the inverse of this power. Since for stocks and bonds it is an empirical fact that this power is larger than two (consistent with a finite variance), the diversification speed is higher than the square root of the normal case. The apparent, but not real, contradiction derives from the comparison between fat tails and normal tails: By lowering the risk level, any power rate is eventually always beaten by the exponential decline of the normal distribution. But if the

risk level is held constant, the above power arguments apply.

In summary, we extend the cost-benefit analysis to explain the low portfolio diversification by incorporating the concern for downside risk and by recognizing the stylized fact of fat tailed distributed returns. This gives quite a bit of mileage to closing the gap of the portfolio diversification puzzle. The rest of the paper proceeds as follows. Section 2 recapitulates Statman's cost-benefit analysis and presents the alternative measures of risk. Following this, we calculate the optimal levels of diversification empirically by the mean-risk optimization but with different types of risk measures in section 3. In section 4 we review some of the important properties of heavy tail distributions. The implications of the heavy tail property for diversification under the various risk measures are derived in section 5. Conclusions and summarizing comments are provided in the final section. The appendix collects some derivations and a useful result.

2 A Cost Benefit Analysis of Portfolio Size

Early diversification studies such as Evans and Archers (1968) and Elton and Gruber (1978) focussed solely on the benefits from portfolio diversification. In these studies the benefits are measured in terms of the reduction in the portfolio return volatility. It is shown by how much the volatility is reduced if the number of assets in the portfolio is increased. Since different stocks are correlated, as in the CAPM, these studies also show there is a limit to what diversification can attain. Moreover, it is clear from these studies that it takes quite a few extra

stocks to get some volatility reduction.

Even though it is of clear interest to know how much it takes to eliminate 'almost all' or 'virtually all' unsystematic risk through diversification, which is the typical result from the early literature, it is unsatisfactory in an economic sense when left to itself. A first step to meet this criticism was the development of tests to determine the statistical significance of the volatility reduction as the portfolio size is increased, an approach pioneered by Evans and Archers (1968)¹. Such an analysis provides a statistical limit to the benefit of diversification. An economic based limit takes into account the associated costs. Thus a financial economics based analysis weighs the benefits against the costs of diversification. The optimal portfolio size is there where at the margin the cost of adding one extra security is equal to the benefit of the reduction in risk. Statman (1987) was the first to cast the question of diversification in this optimizing framework. We are not aware of any subsequent literature² that has followed this framework, except Statman (2004).

Even though the recognition of the costs of diversification reduces the portfolio size, Statman(1987, 2004)'s estimates nevertheless display the low diversification puzzle. Actual portfolios contain less than 10 different stocks in reality. Polkovnichenko (2005) and Goetzmann and Kumar (2008) find that the average number of stocks³ held in actual portfolios is only 3 to 5. The level of diversifica-

¹For recent literature following this approach, see Beck, Perfect and Peterson (1996), Tang (2004) and Domian, Louton and Racine (2007).

²Shawky and Smith (2005) consider indirectly the cost of diversification by using risk-adjusted returns net of expenses of mutual fund portfolios. But these authors do not calculate explicitly costs such as transaction, monitoring and holding costs as in Statman (2004).

³These analyses only consider directly held stocks. In some of empirical tests Goetzmann

tion in the average investor's portfolio appears to be way below the optimal level as prescribed by random stock picking and mean-variance analysis. To close the apparent gap, we propose to turn to downside risk measures and recognize the heavy tail property of the return distributions.

2.1 The cost of diversification

Although diversification has been accepted as an important element of portfolio construction, it carries several potential costs as well, such as transaction, holding and monitoring costs. If costs for a stock were proportional to the size of the trades, then the total amount of costs would be independent of the number of stocks in the portfolio. The only effect would be a reduction in the risk until this is equal to the average covariance between all stocks, see Elton and Gruber (1978). This is the case the early diversification analysis must have had in mind, as a consideration of the costs would not alter the analysis.

The proportionality assumption is not warranted, however, if there are fixed costs per trade so that costs do increase as the number of different stocks in the portfolio increases, while the risk (diversification benefit) is inversely related to the number of stocks. Thus with fixed costs a trade-off exists. For large wealth portfolios, the cost function may even be U-shaped as large trades usually have negative market impact. Statman (1987, 2004) was the first to consider the

and Kumar (2005) use the proportional holdings of investors in mutual funds but they do not examine the composition of investors' mutual fund holdings. They adopt the idea of layered portfolio structure of behavioral portfolio theory. They argue that an investor makes optimal portfolio selections separately when the portfolios belong to a different layer structure. Although a mutual fund is a sum of stocks, an investor considers it separately from individual stocks.

increasing costs of diversification. Statman uses the concept of "additional net cost". The "additional net cost" is the net cost of increasing diversification from any n -stock portfolio to a fully diversified portfolio. Statman assumed that the additional net costs are constant, i.e. independent of n . This presents some conceptual difficulty, since the fully diversified portfolio in practice contains a large but finite number (m) of different securities. Thus as n approaches m , the additional net costs should go down to zero. A by-product of our paper is that we show that this consistency requirement has only a moderate effect in practice.

2.2 Benefits of diversification

The benefits of diversification obtain from a risk reduction. Alternative risk measures are examined to explain the observed level of diversification in reality. These are the global measure of the variance of the portfolio return, and two downside risk measures Value-at-Risk and Expected Shortfall. Because these models are familiar to most readers, we provide only a brief description of each.

In the case of the mean-variance framework, the benefit of diversification is the reduction of risk, and where risk is measured as the standard deviation of portfolio returns. However, the variance may not capture the (extreme) downside risk, whereas the concern for downside risk is an important aspect of the behavioral portfolio theory.

The literature has suggested several alternative measures to capture the downside risk, see e.g. Danielsson et al. (2006). In banking the Value-at-Risk

(VaR) and Expected Shortfall (ES) are arguably the most popular downside risk measures. The VaR is simply a low probability high loss quantile and the ES is the expected loss below the VaR quantile. The latter measure has the theoretical appeal of being subadditive. Both measures of risk are used alongside the traditional standard deviation measure. The downside risk measures better capture the risk of loss than the standard deviation in the case of non-normal heavy tailed distributed returns.

There are several motives for using a downside risk measure for portfolio selection instead of a global risk measure such as the standard deviation. Even if agents are endowed with the standard concave utility function, practical circumstances such as margin requirements often impose constraints that elicit asymmetric treatment of upside potential and downside risk. Regulatory concerns require commercial banks to report and act on the VaR number. Capital adequacy is judged on the basis of the size of the expected loss. There is, moreover, a wealth of experimental evidence for loss aversion by individuals.

2.2.1 The cost and benefits of diversification in the mean-variance model

Consider an investor who composes an equally weighted n -stock portfolio by randomly selecting n different securities from the universe of m securities, $n < m$. Note that the expected standard deviation of the portfolio declines as the number of stocks in the portfolio increases. The limit of the diversification benefit is reached as n becomes large, i.e., when all idiosyncratic risk is removed.

The diversification costs are expressed in currency units. To be able to compare the benefits to the costs, the benefits have to be brought under the same numeraire to be able to determine the optimal level of diversification. Therefore the benefits are translated in units of expected returns. To do this, the risk reduction benefits of diversification in units of expected return are determined by a simple comparison of two portfolios. Let $P(n)$ denote one of the randomized portfolios with size n . Let the m -stock portfolio, $P(m)$, denote the benchmark portfolio. The benchmark portfolio constitutes the most fully diversified portfolio such that $m > n$.

Due to the randomized selection, all stocks are viewed as having the same expected return, R . This return is equal to the sum of the risk-free rate, R_f , plus the equity premium (EP), i.e., $EP = R - R_f$. If investors can borrow and lend at the risk-free rate, the m -stock portfolio can be levered, through borrowing or lending to form the levered portfolio $P(n^*)$. This linearly changes the standard deviation in accordance with the market line, see Figure 3 in Statman(2004). The standard deviation of the levered portfolio $P(n^*)$ equals the standard deviation of the less diversified n -stock portfolio, say. Then the expected return of the levered portfolio is

$$R_{n^*} = R_f + \frac{\sigma_n}{\sigma_m} (R - R_f) = R_f + \frac{\sigma_n}{\sigma_m} EP, \quad (1)$$

where R_{n^*} denotes the expected return of the levered portfolio $P(n^*)$, R_m denotes the expected return of n -stock portfolio $P(m)$, and set $R_m = R_n = R$ be

constant, due to the random stock picking in the naive portfolio composition strategy. Denote σ_n and σ_m be the standard deviation of n -stock portfolio $P(n)$ and m -stock portfolio $P(m)$, respectively.

Equation (1) defines the "Total (capital) Market Line" and all levered portfolios $P(n^*)$ lie on this line. From (1), one can derive the difference between the expected returns of the n -stock portfolio, R , and the expected return of its corresponding levered portfolio, R_{n^*} . The incremental benefit (B_n) of increased diversification from n to m stocks expressed in units of expected returns is

$$B_n^{stdv} = R_{n^*} - R = \left(\frac{\sigma_n}{\sigma_m} - 1 \right) \times EP. \quad (2)$$

On the basis of this equation, Statman (1987) estimated that the optimal level of diversification amounts to holding about 40 different stocks by balancing these benefits with his net additional cost measure. Later, on the basis of new figures and a lower estimate of net cost in Statman (2004), the optimal level of diversification estimate increased to about 300 stocks. These estimates constitute the low diversification puzzle, given that actual portfolios contain less than 10 different stocks in reality. In the following sections, we consider investors who are using alternative risk measures rather than variance to explain this puzzle.

2.2.2 Mean-VaR model

In the mean-variance context the CAPM provides an equilibrium theory of asset prices that we need not repeat here. Analogously, Roy's (1952) safety first theory

as formulated by Arzac and Bawa (1977) provides an equilibrium theory if a downside risk measure is used instead of the standard deviation. In this section we translate the benefit of portfolio diversification with respect to a downside risk measure into the money metric by means of Arzac and Bawa's equilibrium model. The portfolio choice of the safety-first investor is to maximize expected return subject to a downside risk constraint.

We develop a cost-benefit analysis of diversification effect in the framework of the safety first mean-VaR model. We derive a relation similar to (1) in the mean-VaR context. Recall that the VaR is defined as follows: $\Pr \{x \leq -VaR\} = \delta$ for some desired probability level δ . The safety-first investor maximizes expected return subject a downside risk constraint. Arzac and Bawa (1977) use the Value-at-Risk as the downside risk measure in their equilibrium analysis. In contrast to the variance, the VaR measure recognizes a different level of exposure depending on the imposed risk (probability) level. For example, a VaR level at $\delta = 0.001$ at the daily frequency refers to an 'extreme risk' since it corresponds to approximately one event per five years. Whereas the 0.01 or 0.05 levels are used as non-extreme risk levels, which correspond to one event per 100 days or 20 days.

If the m -stock portfolio $P(m)$ is levered with the risk-free asset with weight ω , then we get the levered portfolio $P(n^*)$ with the expected return of $R_{n^*} = \omega R + (1 - \omega)R_f$. As we show in the Appendix A, the value at risk of the levered

portfolio then follows as

$$VaR_{n*} = \omega VaR_m - (1 - \omega)R_f, \quad (3)$$

where VaR_{n*} and VaR_m are the value at risk of portfolio $P(n*)$ and $P(m)$, respectively. The VaR_{n*} is equal to VaR_n , the Value-at-Risk of a less diversified n -stock portfolio $P(n)$. Analogous to (1), the expected return of the levered portfolio can thus be expressed as⁴

$$R_{n*} = R_f + \frac{VaR_n + R_f}{VaR_m + R_f} (R - R_f) \quad (4)$$

by substituting ω from (3) into $R_{n*} = R_f + \omega(R - R_f)$. Note that this equation corresponds to the equation (14) from Arzac and Bawa's (1977) equilibrium analysis. The incremental benefit of increased diversification from n to m stocks, B_n , on basis of the VaR measure thus reads

$$B_n^{VaR} = R_{n*} - R = \left\{ \frac{VaR_n + R_f}{VaR_m + R_f} - 1 \right\} \times EP. \quad (5)$$

2.2.3 Mean-ES model

A similar expression can be derived if the expected shortfall (ES) is used as the measure of downside risk. If the distribution of return x is continuous, ES at

⁴In (4) the ratio of standard deviations in (1) is replaced by the ratio of the VaRs shifted by R_f . This is necessary since the VaR measure is not translation invariant, while the standard deviation measure is.

confidence level $(1 - \delta)$ is defined as

$$ES(q) = -E(x|x \leq q) = - \int_{-\infty}^q x \frac{f(x)}{F(q)} dx,$$

where $f(\cdot)$ and $F(\cdot)$ are the density and distribution function of x and $\Pr\{x \leq q\} = \delta$. With arguments similar to the mean-VaR setting, we obtain the following expression for the levered portfolio $P(n^*)$

$$R_{n^*} = R_f + \frac{ES_n + R_f}{ES_m + R_f} (R - R_f). \quad (6)$$

Note that $R_{n^*} = \omega R + (1 - \omega)R_f$. As we show in the Appendix A

$$ES_{n^*} = \omega ES_m - (1 - \omega)R_f,$$

where ES_{n^*} and ES_m are the expected shortfall of portfolio $P(n^*)$ and $P(m)$, respectively, at the loss probability δ . The ES_{n^*} is equal to ES_n , the expected shortfall of an n -stock portfolio $P(n)$. The incremental benefit of increased diversification from n to m stocks on basis of the ES measure then reads

$$B_n^{ES} = R_{n^*} - R = \left\{ \frac{ES_n + R_f}{ES_m + R_f} - 1 \right\} \times EP. \quad (7)$$

To conclude, it is pretty straightforward to adapt Statman's incremental benefit of diversification measure (2) to the case of downside risk measures, as (5) and (7) show.

3 Results from the Empirical Dataset

In this section, we investigate the empirical relevance of the different risk measures for resolving the diversification puzzle. The historical simulations are based on a dataset with actual stock returns. Thus the empirical exercise is not burdened by the whether the idiosyncratic terms are dependent or not, or whether there are common factors that generate the dependence between stock returns. Neither do we impose anything regarding the shape of the return distribution, whether it be normal or heavy tailed. We just let the data speak for themselves.

The results can be summarized as follows. First, at the extreme risk levels such as $\delta = 0.025$ or 0.001 , safety-first investors require fewer stocks for the diversification than mean-variance investors. As we expect from our theoretical analysis in the following sections, the historical simulation indicates that one expects the diversification to require fewer assets under safety first criterion paired with extremal risk concerns. Second, at risk level such as $\delta = 0.05$ or 0.01 there is no substantive qualitative difference between mean-variance and safety first criteria.

For the empirical analysis we randomly select equally weighted n stock portfolios. From the empirical distribution of n stock portfolio return we calculate the VaR, ES and variance measures. This means that we do not rely on any prior distributional assumption regarding the distribution of the returns, nor do we rely on a specific assumption regarding the cross-dependency between stock returns. Note that a portfolio construction by random selection assures

the assumption of identical ex ante expected returns. In the second part of this section we also consider more able investors than just the random stock picker.

We choose 888 stocks from the NYSE and 425 stocks from the NASDAQ (a total 1313 stocks). We use daily returns (close-to-close data), including cash dividends. The data were obtained from the Datastream. The data spans the period from January 1, 1985, through February 15, 2005, giving a sample size of 5251. Thus more than 20 years of daily data are considered, including the short-lived 1987 crash. These particular stock series⁵ were selected as these have a complete record span during the period. We construct equally weighted n -stock portfolio $P(n)$, $n = 1, 2, \dots, 1313$, by randomly selecting stocks from the 1313 stocks without replacement. The averages of the standard deviation, historical VaR and ES from 1000 different portfolios with n -stocks are calculated for each n .

3.1 The Naive 1/n Strategy

We calculate the corresponding incremental benefits from diversification as per formulas (2), (5) and (7) at the $\delta = 0.05, 0.01, 0.025$ and 0.001 risk levels. Note that an event with probability $\delta = 0.001$ corresponds to an extreme event that may occur about once every five years. The δ -level 0.05 reflects events that occur about every month. This is also approximately the level where the (fitted) normal distribution and the (fat tailed) empirical distribution cross. So for investors with a genuine concern for downside risk, only the δ -levels below

⁵Thus there is some selection bias towards thinner tails as the worst performing stocks are omitted; this is partly balanced by the fact that newly listed companies are also excluded.

5% should be relevant. The optimal level of diversification depends on where these incremental benefits equate with the incremental costs. To this end we use Statman's (2004) estimate of 0.06 percent additional net cost when moving from a small n -stock portfolio to the fully diversified portfolio⁶. As mentioned before, this presents some conceptual difficulty since the additional net costs should go down to zero. As a check on consistency we therefore also investigate by how much the portfolio size increases if we allow for a linear decline of these costs. Furthermore, as in the Statman (2004), we use an equity premium of 3.44% and the risk-free rate is 2.19%.

The results in Table 1 show that in the case of mean-variance model the optimal level of diversification is about 400 stocks. In the mean-VaR model, the optimal level of diversification is 250 stocks when the risk level is $\delta = 0.05$, while this declines to a mere 50 stocks at $\delta = 0.001$. In the mean-ES model, the optimal level of diversification is 250 stocks if $\delta = 0.05$, and 75 stocks when $\delta = 0.001$.⁷

Over the moderate range between the 0.05 and the 0.01 δ -levels, the op-

⁶Statman (2004) assumes that the cost of holding the fully diversified portfolio can be approximated by the expense ratio of the Vanguard Total Stock Market Index Fund, which at the time was 0.20 percent per annum (these run currently at 0.15%). Furthermore, Statman used 0.14 percent as a conservative estimate of the expected annual costs of buying and holding portfolios of individual stocks. The difference between these two estimates then yields the imputed 0.06% incremental costs, which assumed to be independent of n .

⁷A perhaps somewhat puzzling fact in Table 1 is that the excess benefits for some entries fall below -0.06 for the $\delta = 0.0025$ and 0.001 entries. This implies that the incremental benefit (5) can be negative as one increases the number of stocks from n to m . As (13) and (21) show, if $n < m$ then necessarily $VaR_n > VaR_m$, so that the benefits are always positive in theory. The phenomenon stems from the coarseness of the empirical distribution in the tail area. Due to the limited number of observations, it can easily happen that $VaR_n < VaR_m$, whereas the underlying distributions do not generate this behavior. A simple example is two draws of three loss returns (8, 5, 4) and (5, 1, 7). Consider $\Pr\{X > VaR\} \leq 1/3$. For both sets of returns the $VaR = 5$. But for the averaged portfolio $\Pr\{X > VaR\} \leq 1/3$ has a VaR of 5.5.

timal level of diversification varies from 175 to 400 different stocks. Thus at moderate risk levels, the amount of diversification is considerable under all criteria. For the more extreme risk levels with δ equal to 0.025 and 0.001, there is a large difference in the amount of diversification between the mean-variance investor and the safety first investors who rely on the mean-VaR or mean-ES criteria. The optimal levels chosen by these safety first investors approach the levels observed in practice. Since the mean-variance criterion underestimates the downside risk far into the tails of the distribution, the benefit of diversification is overestimated. The non-parametrically implemented downside risk measures do remove this bias and come closer to actual portfolio sizes.

We also implemented the case of decreasing incremental cost. As discussed in section 2.1 Statman (1987, 2004) assumes the constant additional net cost. Suppose, however, that the additional net cost declines as n increase. It turns out that this hardly increases the portfolio sizes under the extremal risk consideration. For example, we found that the optimal level of diversification increases from 300 to 350 stocks when the risk level is $\delta = 0.05$, while there are no differences when $\delta = 0.0025$ or 0.001 in the mean-VaR model. Therefore we do not repeat these results (but are available upon request).

3.2 Judicious portfolio selection

In practice, at least a subset of investors does better than the random stock picker (by implication some others must do worse). Whether this is due to skill or luck is hard to tell. According to Jacob (1974), Johnson and Shannon (1974)

# of Stocks	ST DEV	VaR5	VaR01	ES5	ES01
5	4.08	3.70	2.29	3.32	1.72
20	1.39	1.24	0.37	0.98	0.35
50	0.58	0.48	0.00	0.35	0.06
75	0.38	0.28	-0.07	0.21	0.01
100	0.26	0.18	-0.09	0.13	-0.02
250	0.06	0.01	-0.11	0.00	-0.05
400	0.01	-0.02	-0.11	-0.03	-0.06
500	-0.02	-0.04	-0.10	-0.04	-0.06
1000	-0.05	-0.06	-0.07	-0.06	-0.06
Excess benefits					

and others, an investor can reduce unsystematic risk significantly with only a few securities if he or she chooses stocks sensibly. Per contrast, Goetzmann and Kumar (2008) do not find any significant evidence of diversification improvement by active means, such as, picking less correlated stocks. Risk reduction through proper stock selection may thus reflect the investor's skills, but may also just be an artifact. Whatever the case may be, it is of interest to investigate the effects of judicious stock selection on portfolio diversification.

In our experiment, the "active and skilled" investor still constructs an "equally weighted portfolio", but supposedly is able to select stocks from a subset to attain a lower risk level than is possible under pure random selection from the universe of all stocks. A sophisticated investor is someone who randomly picks a portfolio from a collection of low beta stocks. Low beta stocks naturally have lower downside risk stemming from the market factor.

We perform an alternative experiment of judicious portfolio selection whereby the investor can choose from the subset of low beta stocks. We estimated all betas and consider the subset of stocks that have a beta in the range of $[0.5, 0.9]$. From this subset, we randomly selected equally weighted n stock portfolios. From the empirical distribution of n stock portfolio return we calculate the VaR, ES and variance risk measures. The results in Table 2 show that in the case of mean-variance model the optimal level of diversification is about 35 stocks, considerably lower than the 400 stocks in Table 1. The standard deviation entries in the first column are lower than the same entries in the first column of Table 1, due to the fact that the stocks in the subset all have a beta less than

one. In the mean-VaR model, the optimal level of diversification is 35 stocks at the $\delta = 0.05$ risk level, while this declines to 14 stocks when $\delta = 0.001$. In the mean-ES model, the optimal level of diversification is 25 stocks if $\delta = 0.05$, and just 11 stocks when $\delta = 0.001$. This alternative judicious strategy lowers the optimal portfolio sizes to a range between 11 to 17 respectively, at the more extreme risk levels of $\delta = 0.001$ and 0.0025. Such risk levels of the safety first investor are at least consistent with revealed investor preference of low portfolio diversification as are reported in the literature.

3.2.1 Summary of empirical results

At relatively high δ -levels such as $\delta = 0.05$ or 0.01 the variance and the VaR criterion imply similar portfolio sizes. An event with $\delta = 0.05$, for example, happens on average every 20 days, which is common risk, not extremal risk. Even for a safety first investor who cares about downside risk, but not extremal risk, we may conclude that there is no substantive qualitative difference between two different risk measures. This changes drastically at lower δ -levels.

The safety first investor with some skill for stock picking and with a risk appetite of $\delta = 0.001$, composes portfolios of about ten stocks for proper diversification. The safety first criterion paired with such a low δ -level may not be the only criterion in the universe that can explain the observed low diversification. Nevertheless, our analysis shows that downside risk measures are able to go a long way towards an explanation, whereas this is not possible with mean variance type utility functions. The reason is that downside risk criteria are sen-

sitive to qualitative differences in the tail of the return distributions, whereas the variance measure is not.

The next sections examine if some of these properties can be explained by theory. Our assumption is that the extremal tail area of empirical distribution is shaped like the Pareto distribution as the extreme value theory suggests, even though the normal law applies well around 1% or 5% tail area of empirical distribution.

4 Analytical Form of Diversification Benefit

In this and following sections we examine analytically some of the determinants of the empirical results identified in the previous section. Our objective is to understand why the historical simulations provide qualitatively similar results at the common level of risk regardless the way the risk is measured, and why the risk measures imply qualitative differences in the tail of the return distributions. Our main focus is on identifying the properties of downside risk measures in the tail area of the return distribution, where the heavy-tail theory applies.

4.1 Diversification benefits from mean-variance model

As a risk measure, one can estimate a standard deviation without any distributional assumption. When investors consider the variance measure, it is assumed normality assumption implicitly but not explicitly. Thus we can derive the benefits of diversification with some additional assumptions variance and covariance structure.

Consider an investor who composes an equally weighted n -stock portfolio by randomly selecting n different securities from the universe of m securities, $n < m$. From the point of view of the random stock picker, the expected returns of each security are all equal to the averages of all m securities. Let r_i denote the return of the i -th security with the expected return R , and standard deviation σ . The standard deviation of a n -stock portfolio is

$$\sqrt{\sum_{j=1}^n \sum_{i=1}^n \omega_i \omega_j Cov(r_i, r_j)},$$

where ω_i is the weight of stock i in the portfolio and $Cov(r_i, r_j)$ is the covariance between the returns of stocks i and j .

To allow for non-diversifiable market risk between the security returns, we first consider a single index model in which the idiosyncratic risk is assumed to be independent of the market risk r_{mkt}

$$r_i = \beta_i r_{mkt} + q_i, \tag{8}$$

where r_{mkt} is the (excess) return on the market portfolio, β_i is the amount of market risk and q_i is the idiosyncratic risk of the return on asset i . The idiosyncratic risk may be diversified away fully in arbitrarily large portfolios and hence is not priced. But the cross-sectional dependence induced by the common market risk factor has to be held in every portfolio. We further assume the idiosyncratic risk q_i and the market risk r_{mkt} to be distributed with variances σ_{mkt}^2 and σ_q^2 , respectively, and $E[r_{mkt}] = R$ and $E[q_i] = 0$. The idiosyncratic

risks are independent, i.e. have zero cross correlation $Cov(q_i, q_j) = 0$. Due to the random stock picking, we also assume that the beta of stock i is also a random variable β_i with $E(\beta_i) = \beta$ and $Var[\beta_i] = \sigma_\beta^2$. The standard deviation σ_n is the standard deviation of the equally weighted n -stock portfolio

$$\sigma_n = \sqrt{\beta^2 \sigma_{mkt}^2 + \frac{1}{n} \left(\sigma_{mkt}^2 \sigma_\beta^2 + R^2 \sigma_\beta^2 + \sigma_q^2 \right)}. \quad (9)$$

Note that the expected standard deviation of the portfolio declines as the number of stocks in the portfolio increases. The limit of the diversification benefit is reached as the number of stocks in the portfolio becomes large, i.e., when all idiosyncratic risk is removed. Note that if $n \rightarrow \infty$

$$\sigma_n \rightarrow \beta \sigma_{mkt}. \quad (10)$$

Thus the incremental benefit of diversification (1) for the mean-variance investor is

$$B_n^{stdv} = \left\{ \sqrt{\frac{\left(\frac{1}{n}\right) + \left(\frac{n-1}{n}\right)\rho}{\left(\frac{1}{m}\right) + \left(\frac{m-1}{m}\right)\rho}} - 1 \right\} \times EP \quad (11)$$

where

$$\rho = \frac{\beta^2 \sigma_{mkt}^2}{\sigma_{mkt}^2 \sigma_\beta^2 + R^2 \sigma_\beta^2 + \sigma_q^2}.$$

Note that the benefits from diversification come at a rate equal to the square root of the number of assets n . On the basis of this equation, Statman (1987) estimated that the optimal level of diversification amounts to holding about 40 different stocks by balancing these benefits with his net additional cost measure.

Later, on the basis of new figures of ρ , EP , m and a lower estimate of net cost in Statman (2004), the optimal level of diversification estimate increased to about 300 stocks.

As the market portfolio is large the diversification benefit can be simplified to $B_{n\infty}^{stdv} = \lim_{m \rightarrow \infty} B_n^{stdv}$,

$$B_{n\infty}^{stdv} = \left(\frac{1}{\beta\sigma_{mkt}} \sqrt{\beta^2\sigma_{mkt}^2 + \frac{1}{n} (\sigma_{mkt}^2\sigma_\beta^2 + R^2\sigma_\beta^2 + \sigma_q^2)} - 1 \right) \times EP \quad (12)$$

if $m \rightarrow \infty$ and $\beta \neq 0$.

4.2 Diversification benefits under downside risk measures

Given the concern for downside risk, it becomes important to characterize the tail risk adequately. While the normal distribution is standard fare in financial analysis and is suitable for many questions, it is by now well realized that the normal law is less appropriate in the area of risk management and downside risk. Therefore we investigate the heavy tail distribution and its implications for the risk measures under consideration.

In comparison to the normal distribution the distribution of asset returns has more returns concentrated in the very center and more returns in the tails of the distribution. This fat tail property is modelled by assuming that the distribution in the tail areas behaves like a Pareto distribution; see Jansen and de Vries (1991) for the empirical relevance. The tail of the Pareto distribution declines at a power rate, which is always slower than the exponential decline of

the normal distribution. Other distributions like the Student-t and non-normal sum-stable distributions also exhibit Pareto type tails.

4.2.1 Diversification benefits in the case of normal distribution

Under normality, the VaR level of $r(n)$ by the random stock picker is given by

$$VaR_n = -R_n + z_\delta \sigma_n, \quad (13)$$

where $\Pr\{r(n) \leq -VaR_n\} = \delta$, and $r(n) \sim N(R_n, \sigma_n^2)$, and $-z_\delta$ is the δ quantile of $N(0, 1)$, so that $z_\delta \sigma_n > 0$ for δ small. From (5) the incremental benefit of diversification to the safety first investor who uses the VaR risk measure is therefore, cf. (4),

$$B_n^{VaR} = \left\{ \frac{z_\delta \sigma_n - EP}{z_\delta \sigma_m - EP} - 1 \right\} \times EP, \quad (14)$$

where $R_n = R_m = R$ and $EP = R - R_f$. For $B_{n\infty}^{VaR} = \lim_{m \rightarrow \infty} B_n^{VaR}$, i.e. if the market portfolio is large, then using (10)

$$B_{n\infty}^{VaR} = (\sigma_n - \beta \sigma_{mkt}) \left(\frac{z_\delta EP}{z_\delta \beta \sigma_{mkt} - EP} \right). \quad (15)$$

We proceed analogously for the B_n^{ES} measure. Under normality, the ES level

of $r(n)$ for the random stock picker is given by

$$\begin{aligned}
ES_n(-VaR_n) &= -\frac{1}{\delta} \int_{-\infty}^{-VaR_n} x \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{-\frac{1}{2}\left(\frac{x-R_n}{\sigma_n}\right)^2\right\} dx \\
&= -\frac{1}{\delta} \int_{-\infty}^{-z_\delta} (R_n + \sigma_n z) \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\} dz \\
&= -R_n + ES_{z_\delta} \times \sigma_n,
\end{aligned}$$

where ES_{z_δ} is the expected shortfall of the standard normal distribution at the probability level δ . So that by (7)

$$B_n^{ES} = \left\{ \frac{ES_{z_\delta}\sigma_n - EP}{ES_{z_\delta}\sigma_m - EP} - 1 \right\} \times EP. \quad (16)$$

If the market portfolio is large (i.e., $m \rightarrow \infty$) this measure simplifies to

$$B_{n\infty}^{ES} = (\sigma_n - \beta\sigma_{mkt}) \frac{ES_{z_\delta} \times EP}{ES_{z_\delta}\beta\sigma_{mkt} - EP}.$$

4.2.2 Diversification benefits in the case of heavy tails

For the purpose of presentation we first assume that the returns of securities are identically and independently distributed. This counterfactual assumption of independence is relaxed later by allowing for common factors such as (8).

Suppose the $\{r_i\}$ are generated by a distribution with fat tails, which vary regularly at infinity. Thus, far from the origin the Pareto term dominates:

$$\Pr\{r_i \leq -s\} = As^{-\alpha}[1 + o(1)], \quad \alpha > 0, A > 0, \quad (17)$$

as $s \rightarrow \infty$. The Pareto term implies that only moments up to α are bounded and hence the terminology of fat tails. Per contrast, the normal distribution has all moments bounded because of its exponential tail shape.

An implication of the fat tail property is the simplicity of the tail probabilities for convoluted data. By Feller's Theorem (1971, VIII.8) we have that if r_i and r_j are independently distributed and adhere to (17), then

$$\Pr\{r_i + r_j \leq -s\} = 2As^{-\alpha}[1 + o(1)].$$

Feller's theorem is presented in detail in the Appendix B. Some intuition for the Feller theorem is as follows. Let losses $-X$ be iid Pareto distributed with scale $A = 1$. Then for large s the probability of one or two severe losses is

$$1 - P\{X_1 > -s, X_2 > -s\} = 1 - (1 - s^{-\alpha})^2 \approx 2s^{-\alpha},$$

since the second term $s^{-2\alpha}$ is of smaller order. The probability of one or two losses is to a first order equal to the sum of the marginal (single) loss probabilities. Thus only the marginal (univariate) probability mass along the axes counts. Similarly, the mass below the line $X_1 + X_2 = s$ is also determined by how much probability mass is aligned along the axes below this line, i.e. $2s^{-\alpha}$.⁸

Thus suppose that the $\{r_i\}$ are generated by a fat-tailed distribution satisfying (17). From the Feller's Theorem, one can derive the diversification effect

⁸For a proof of the Feller theorem by elementary integration, see Dacorogna et al.(2001). Appendix B.2 gives a more intuitive derivation.

for the equally weighted portfolio $P(n)$ at the larger loss levels. The return of an n -stock portfolio

$$r(n) = \frac{1}{n} \sum_{i=1}^n r_i$$

satisfies

$$\Pr \{r(n) \leq -s\} = n^{1-\alpha} A s^{-\alpha} [1 + o(1)] \quad (18)$$

as $s \rightarrow \infty$. Note how the weighing affects the scale. In particular observe that the loss probability is lower for larger n if $\alpha > 1$. This latter requirement boils down to requiring that the mean of the return is bounded. In an early contribution Fama and Miller (1971) already noted that if $\alpha < 1$ diversification increases the risk (for the case of sum stable distributions). But a finite mean is undisputed for most financial securities. Note that upon inversion, (17) implies that at a constant risk level p , the loss level s changes as follows

$$s \simeq \left(\frac{A}{p}\right)^{1/\alpha} n^{1-1/\alpha}.$$

Thus at a given risk level, the diversification speed is $1 - 1/\alpha$. If $\alpha > 2$, implying that the variance is finite, the diversification speed is larger than the square root (implied by e.g. the normal distribution).

We now relax the assumption of independence between the security returns and allow for non-diversifiable market risk. The market risk reduces the benefits of diversification. For the consistency of comparison, we consider a single index model in which the idiosyncratic risk is assumed to be independent of the market

risk, $r_i = \beta_i r_{mkt} + q_i$, as (8).

We apply Feller's theorem again for deriving the benefits from cross-sectional portfolio diversification in this single index model. In this single index model the q_i are cross-sectionally independent and are, moreover, independent from market risk factor r_{mkt} . In addition, suppose that the distributions of q_i and r_{mkt} are regularly varying with the same tail index but different scales A and C . Thus assume

$$\Pr\{q_i \leq -s\}/A = \Pr\{r_{mkt} \leq -s\}/C = s^{-\alpha}[1 + o(1)], \quad (19)$$

where $\alpha > 0$, $A > 0$ and $C > 0$. Since the portfolio elements are randomly chosen, we assume that the beta of stock i is also a random variable β_i . It is assumed that the β_i are distributed on the support $[0, a]$, and hence have all moments bounded. This is in contrast to the other random variables q_i and r_{mkt} that only have moments up to α .

Consider the return of an equally weighted portfolio

$$r(n) = \frac{1}{n} \sum_{i=1}^n \beta_i r_{mkt} + \frac{1}{n} \sum_{i=1}^n q_i.$$

We like to determine the probability

$$\Pr\{r(n) \leq -s\}$$

for s large. To this end we use a combination of the Feller convolution argu-

ment and the Breiman result for products of random variables presented in the Appendix B.1. We first rewrite the probability by a conditioning argument

$$\begin{aligned} \Pr\{r(n) \leq -s\} &= \Pr\left\{\frac{1}{n} \sum_{i=1}^n \beta_i r_{mkt} + \frac{1}{n} \sum_{i=1}^n q_i \leq -s\right\} \\ &= E\left[\Pr\left\{\bar{\beta} r_{mkt} + \frac{1}{n} \sum_{i=1}^n q_i \leq -s \mid \beta_i\right\}\right], \end{aligned}$$

where $\bar{\beta} = \frac{1}{n} \sum_{i=1}^n \beta_i$. Next we apply the convolution result of Feller in combination with the Breiman result to get

$$\begin{aligned} \Pr\{r(n) \leq -s\} &= E\left[(C\bar{\beta}^\alpha + An^{1-\alpha})s^{-\alpha}[1 + o(1)]\right] \\ &= (CE[\bar{\beta}^\alpha] + An^{1-\alpha})s^{-\alpha}[1 + o(1)] \end{aligned} \tag{20}$$

as $s \rightarrow \infty$. This result capitalizes on the assumption that the distribution of the β_i has all moments bounded.

In general one finds that the single index model does not hold exactly due to the fact that $\text{Cov}[q_i, q_j]$ is typically also non-zero for off diagonal elements. Thus though the q_i may be independent from the market risk factor r_{mkt} (they are uncorrelated with r_{mkt} by construction), they are typically not cross sectionally independent from each other. This case is usually referred to as the market model. Given the Feller theorem, it is not difficult to extend (20) to allow for this feature, but we leave it to the reader.

Denote δ as the fixed desired probability level such that

$$\delta = \Pr \{r(n) \leq -VaR_n\}.$$

Since the VaR measure is defined at a given probability level, rather than at a given quantile level, (20) is not the desired final result. Consider holding the probability constant but letting the VaR level change as the number of assets n increases. By first order inversion based on De Bruijn's theorem⁹, we finally obtain

$$VaR_n = (CE [\bar{\beta}^\alpha] + An^{1-\alpha})^{1/\alpha} \delta^{-1/\alpha} [1 + o(1)] \text{ as } \delta \rightarrow 0.$$

The advantage is that we can now take care of the stochastic nature of the β_i , made easy by the fact that we only need an expectation, which is not stochastic.

But it is not directly easy to see what the order of magnitude of

$$T(\alpha, n) = E \left[\left(\frac{1}{n} \sum_{i=1}^n \beta_i \right)^\alpha \right]$$

is. In the Appendix B.3 we argue that there exists an $W \in [\alpha, k(k-1)]$ and where k is the integer closest to α , such that $k \geq \alpha$.

$$E \left[\left(\frac{1}{n} \sum_{i=1}^n \beta_i \right)^\alpha \right] = \beta^\alpha \left\{ 1 + \frac{W}{2} \frac{\sigma_\beta^2}{\beta^2} \frac{1}{n} + O(n^{-2}) \right\}$$

⁹See de Bruijn's inverse in Theorem 1.5.13 of Bingham, Goldie and Teugels (1987).

where $E(\beta_i) = \beta$ and $Var[\beta_i] = \sigma_\beta^2$. Hence, combining terms, we obtain

$$\begin{aligned} VaR_n &= \left[C\beta^\alpha \left\{ 1 + \frac{W}{2} \frac{\sigma_\beta^2}{\beta^2} \frac{1}{n} + O(n^{-2}) \right\} + An^{1-\alpha} \right]^{1/\alpha} \delta^{-1/\alpha} [1 + o(1)] \\ &= [v_1 + v_2 n^{-1} + v_3 n^{1-\alpha} + O(n^{-2})]^{1/\alpha} [1 + o(1)] \end{aligned} \quad (21)$$

say, and where $v_1 = C\beta^\alpha/\delta$, $v_2 = \frac{1}{2}C\beta^{\alpha-2}W\sigma_\beta^2/\delta$, $v_3 = A/\delta$. Note that for the special case of identical beta's $\sigma_\beta^2 = 0$ and (21) reduces to

$$VaR_n = [v_1 + v_3 n^{1-\alpha}]^{1/\alpha} [1 + o(1)]. \quad (22)$$

Note that for $\alpha > 2$ the idiosyncratic part $v_3 n^{1-\alpha}$ is smaller than the market factor term $v_2 n^{-1}$ for sufficiently large n . As n increases these two factors determine the rate of the diversification effect for the Value at Risk. This can be seen by differentiation:

$$\begin{aligned} \frac{\partial VaR_n}{\partial n} &= \frac{1}{\alpha} (v_1 + v_2 n^{-1} + v_3 n^{1-\alpha} + O(n^{-2}))^{1/\alpha-1} [1 + o(1)] \\ &\quad \times [-v_2 n^{-2} + (1-\alpha)v_3 n^{-\alpha} + O(n^{-3})]. \end{aligned} \quad (23)$$

Note that differentiating $v_2 n^{-1} + v_3 n^{1-\alpha} + O(n^{-2})$ gives the same terms as between the second square brackets in (23). This shows that an increase in n changes VaR_n approximately by a constant times $v_2 n^{-1} + v_3 n^{1-\alpha} + O(n^{-2})$.

This result can now be used in (4) to obtain an explicit expression for the benefit of diversification for mean-VaR safety first investors when asset returns

are heavy tailed distributed and the acceptable risk level δ is low. Specifically, define the benefits as $B_{n*} \equiv R_{n*} - R$. Then combining (4) and (21) gives

$$B_n^{VaR} = \left\{ \frac{[v_1 + v_2 n^{-1} + v_3 n^{1-\alpha} + O(n^{-2})]^{1/\alpha} [1 + o(1)] + R_f}{[v_1 + v_2 m^{-1} + v_3 m^{1-\alpha} + O(m^{-2})]^{1/\alpha} [1 + o(1)] + R_f} - 1 \right\} \times EP, \quad (24)$$

and where EP is the equity premium $R - R_f$.

For the other downside risk measure, ES , we can derive a similar expression.

The level of ES_n follows

$$ES_n = - \int_{-\infty}^q x \frac{f_n(x)}{F_n(q)} dx$$

for the given q such that $\Pr\{r(n) \leq q\} = \delta$ and $f_n(\cdot)$ and $F_n(\cdot)$ are the probability and cumulative density function of the returns of the n -stock portfolio. From Proposition 1 of Danielsson, Jorgensen, Sarma and de Vries (2006), one obtains the following approximation for ES at the given probability level δ

$$ES_n = \frac{\alpha}{\alpha - 1} [v_1 + v_2 n^{-1} + v_3 n^{1-\alpha} + O(n^{-2})]^{1/\alpha} [1 + o(1)]. \quad (25)$$

The benefit of increased diversification from n to m stocks under the mean-ES criterion can thus be approximated by

$$B_n^{ES} = \left\{ \frac{\frac{\alpha}{\alpha-1} [v_1 + v_2 n^{-1} + v_3 n^{1-\alpha} + O(n^{-2})]^{1/\alpha} [1 + o(1)] + R_f}{\frac{\alpha}{\alpha-1} [v_1 + v_2 m^{-1} + v_3 m^{1-\alpha} + O(m^{-2})]^{1/\alpha} [1 + o(1)] + R_f} - 1 \right\} \times EP \quad (26)$$

where we used (6) and (25).

Lastly, consider the standard deviation risk measure. Even if the returns are heavy tailed, as long as $\alpha > 2$ the variance exists. Under this condition one can proceed and use (2) for mean-variance optimizing agents even if the return distribution is heavy tailed.

5 Theoretical Comparison of Diversification Benefits

We compare the benefits from diversification for different type of investors who employ different risk measures. The diversification benefits under alternative risk measures are compared as the number of assets increases. The comparison is between a mean-variance investor and a safety first investor. The comparison is made under the alternative assumptions that returns are normally distributed and that returns are heavy tailed distributed. We start the comparison by assuming that asset returns are (counterfactually) normally distributed. For the consistency of comparison, we consider a single index model, $r_i = \beta_i r_{mkt} + q_i$, as (8).

We show that the three different risk measures $B_{n\infty}^{stdv}$, $B_{n\infty}^{VaR}$ and $B_{n\infty}^{ES}$ are equivalent in terms of diversification benefits if the asset returns are normally distributed.

Proposition 1 (Equivalence under Normality) *Suppose asset returns $\{r_i\}$ follow the single index model $r_i = \beta_i r_{mkt} + q_i$. Let the q_i be i.i.d normally distributed for all i , and be independent of r_{mkt} , which is normally distributed as*

well. The variances of q_i and r_{mkt} are denoted as σ_q^2 and σ_{mkt}^2 , respectively, and $R = E[r_{mkt}]$. Suppose that the acceptable risk level δ is below 50%. Then the diversification benefits of $B_{n\infty}^{stdv}$, $B_{n\infty}^{VaR}$ and $B_{n\infty}^{ES}$ are identical to the extent that the ratios $B_{n\infty}^{stdv} / B_{n\infty}^{VaR}$ and $B_{n\infty}^{stdv} / B_{n\infty}^{ES}$ do not depend on n .

Proof. First rewrite (12) using (9) and (10) as followings

$$B_{n\infty}^{stdv} = \left(\frac{\sigma_n}{\beta\sigma_{mkt}} - 1 \right) \times EP = (\sigma_n - \beta\sigma_{mkt}) \left(\frac{EP}{\beta\sigma_{mkt}} \right).$$

Under normality, the incremental benefit of diversification to the safety first investor who uses the VaR risk measure is

$$B_{n\infty}^{VaR} = (\sigma_n - \beta\sigma_{mkt}) \left(\frac{z_\delta EP}{z_\delta \beta \sigma_{mkt} - EP} \right). \quad (27)$$

Since the second factor between brackets in (27) does not depend on n , we get

$$\frac{B_{n\infty}^{stdv}}{B_{n\infty}^{VaR}} = \left(\frac{EP}{\beta\sigma_{mkt}} \right) / \left(\frac{z_\delta EP}{z_\delta \beta \sigma_{mkt} - EP} \right) = \frac{z_\delta \beta \sigma_{mkt} - EP}{z_\delta \beta \sigma_{mkt}}$$

which is constant, i.e. independent of n .

We proceed analogously for the B_n^{ES} measure. Under normality

$$B_{n\infty}^{ES} = (\sigma_n - \beta\sigma_{mkt}) \frac{ES_{z_\delta} \times EP}{ES_{z_\delta} \beta \sigma_{mkt} - EP}.$$

Hence,

$$\frac{B_{n\infty}^{stdv}}{B_{n\infty}^{ES}} = \frac{ES_{z_\delta} \beta \sigma_{mkt} - EP}{ES_{z_\delta} \beta \sigma_{mkt}}$$

and similarly

$$\frac{B_{n\infty}^{VaR}}{B_{n\infty}^{ES}} = \frac{ES_{z_\delta} \beta \sigma_{mkt} - EP}{z_\delta \beta \sigma_{mkt} - EP} \frac{z_\delta}{ES_{z_\delta}}.$$

■

In Proposition 1, we use the specific structure for the cross-dependency from the single index model. We can also consider the cross-dependency structure of Statman (1987, 2004), where it is assumed that the standard deviation, σ , of each n stocks, all correlations, ρ , between pairs of stocks and expected returns, R , are identical. Recall the standard deviation (9) of an n stock portfolio. By increasing the portfolio size n , we get that

$$\sigma_n = \sigma \sqrt{\left(\frac{1}{n}\right) + \left(\frac{n-1}{n}\right) \rho} \rightarrow \sigma \sqrt{\rho}.$$

The benefits of diversification (12) can then be simplified for large m to

$$B_{n\infty}^{stdv} = \left(\sqrt{\frac{1}{n\rho} + \left(\frac{n-1}{n}\right)} - 1 \right) \times EP.$$

The effects of diversification are as follows:

Corollary 1 *Suppose that the acceptable risk level δ is below 50%. Under assumptions of Statman's (1987, 2004) and if the $\{r_i\}$ are normally distributed, then the diversification benefits of $B_{n\infty}^{stdv}$, $B_{n\infty}^{VaR}$ and $B_{n\infty}^{ES}$ are identical to the extent that the ratios $B_{n\infty}^{stdv} / B_{n\infty}^{VaR}$ and $B_{n\infty}^{stdv} / B_{n\infty}^{ES}$ do not depend on n .*

Proof. From (2),

$$B_{n\infty}^{stdv} = (\sigma_n - \sigma\sqrt{\rho}) \left(\frac{EP}{\sigma\sqrt{\rho}} \right)$$

Under normality,

$$B_{n\infty}^{VaR} = (\sigma_n - \sigma\sqrt{\rho}) \left(\frac{z_\delta EP}{z_\delta \sigma\sqrt{\rho} - EP} \right)$$

and

$$B_{n\infty}^{ES} = (\sigma_n - \sigma\sqrt{\rho}) \left(\frac{ES_{z_\delta} EP}{ES_{z_\delta} \sigma\sqrt{\rho} - EP} \right).$$

■

For the mean-variance based diversification benefits one does not need the normality assumption. This is not possible for the mean-VaR or mean-ES models, since one needs a distributional assumption to estimate the VaR or ES. The following Corollary makes this precise for $B_{n\infty}^{stdv}$.

Corollary 2 *Suppose asset returns $\{r_i\}$ follow the single index model $r_i = \beta_i r_{mkt} + q_i$, where $q_i \sim (Q, \sigma_q^2)$, $r_{mkt} \sim (R, \sigma_{mkt}^2)$ and where q_i and r_{mkt} are independent. Then the diversification speed of the variance of $r(n)$ is*

$$\text{Var}[r(n)] - \beta^2 \sigma_{mkt}^2 = O(n^{-1})$$

Proof. See the Appendix C. ■

If, however, the tail of the distribution of returns is fat-tailed, then the speed

of the diversification benefits is different for different types of investors. Thus while the result for $B_{n\infty}^{stdv}$ still pertains under other distributional assumptions, as long as the variance is finite, the results for the other downside risk measures $B_{n\infty}^{VaR}$ and $B_{n\infty}^{ES}$ change.

If the market portfolio is large, the benefits (24) simplify as follows

$$\begin{aligned}
B_{n\infty}^{VaR} &\equiv \lim_{m \rightarrow \infty} B_n^{VaR} \\
&= \left\{ \frac{[v_1 + v_2 n^{-1} + v_3 n^{1-\alpha} + O(n^{-2})]^{1/\alpha} [1 + o(1)] + R_f}{v_1^{1/\alpha} [1 + o(1)] + R_f} - 1 \right\} \times EP,
\end{aligned} \tag{28}$$

We will now show that this measure gives a different diversification speed than the standard deviation measure if $\alpha > 2$, that is, if the variance exists.

For the measure of ES, we can derive a similar expression as for the case of VaR under fat-tailed distribution. The benefits expressed in (26) can be simplified if the market portfolio m is large

$$\begin{aligned}
B_{n\infty}^{ES} &\equiv \lim_{m \rightarrow \infty} B_n^{ES} \\
&= \left\{ \frac{\frac{\alpha}{\alpha-1} [v_1 + v_2 n^{-1} + v_3 n^{1-\alpha} + O(n^{-2})]^{1/\alpha} [1 + o(1)] + R_f}{\frac{\alpha}{\alpha-1} v_1^{1/\alpha} [1 + o(1)] + R_f} - 1 \right\} \times EP
\end{aligned} \tag{29}$$

We do the comparison in terms of how the benefits of diversification change as the number of assets included in the portfolio increases.

Proposition 2 *Suppose that the tail of the distribution of $r(n)$ satisfies (20).*

Moreover, assume $\alpha > 2$. Then

$$\frac{\partial B_{n\infty}^{stdv}}{\partial n} = \frac{EP (\sigma_{mkt}^2 \beta^2 + O(n^{-1}))^{-1/2}}{2\beta\sigma_{mkt}} [-\sigma_\beta^2 (\sigma_{mkt}^2 + R^2) n^{-2} - \sigma_q^2 n^{-2}]$$

while

$$\frac{\partial B_{n\infty}^{VaR}}{\partial n} = \frac{EP (v_1 + O(n^{-1}))^{1/\alpha-1} [1 + o(1)]}{\alpha v_1^{1/\alpha} [1 + o(1)] + \alpha R_f} [-v_2 n^{-2} + (1 - \alpha) v_3 n^{-\alpha} + O(n^{-3})]$$

and

$$\frac{\partial B_{n\infty}^{ES}}{\partial n} = \frac{EP (v_1 + O(n^{-1}))^{1/\alpha-1} [1 + o(1)]}{v_1^{1/\alpha} [1 + o(1)] + (\alpha - 1) R_f} [-v_2 n^{-2} + (1 - \alpha) v_3 n^{-\alpha} + O(n^{-3})].$$

Proof. From (12) straightforward differentiation gives

$$\frac{\partial B_{n\infty}^{stdv}}{\partial n} = -EP \frac{1/2}{\beta\sigma_{mkt}} \frac{\sigma_{mkt}^2 \sigma_\beta^2 + R^2 \sigma_\beta^2 + \sigma_q^2}{\sqrt{\sigma_{mkt}^2 \beta^2 + \frac{1}{n} (\sigma_{mkt}^2 \sigma_\beta^2 + R^2 \sigma_\beta^2 + \sigma_q^2)}} \frac{1}{n^2}. \quad (30)$$

We proceed in a similar fashion for the other two measures. Differentiating $B_{n\infty}^{VaR}$ with respect to n , using (23), gives

$$\begin{aligned} \frac{\partial B_{n\infty}^{VaR}}{\partial n} &= \frac{EP}{v_1^{1/\alpha} [1 + o(1)] + R_f} \frac{\partial VaR_n}{\partial n} \\ &= \frac{EP \frac{1}{\alpha} (v_1 + v_2 n^{-1} + v_3 n^{1-\alpha} + O(n^{-2}))^{1/\alpha-1} [1 + o(1)]}{v_1^{1/\alpha} [1 + o(1)] + R_f} \times \\ &\quad [-v_2 n^{-2} + (1 - \alpha) v_3 n^{-\alpha} + O(n^{-3})]. \end{aligned} \quad (31)$$

Analogously,

$$\begin{aligned}
\frac{\partial B_{n\infty}^{ES}}{\partial n} &= \frac{EP}{\frac{\alpha}{\alpha-1}v_1^{1/\alpha}[1+o(1)]+R_f} \frac{\alpha}{\alpha-1} \frac{\partial VaR_n}{\partial n} \\
&= \frac{EP \frac{1}{\alpha-1} (v_1 + v_2 n^{-1} + v_3 n^{1-\alpha} + O(n^{-2}))^{1/\alpha-1} [1+o(1)]}{\frac{\alpha}{\alpha-1}v_1^{1/\alpha}[1+o(1)]+R_f} \times \\
&\quad [-v_2 n^{-2} + (1-\alpha)v_3 n^{-\alpha} + O(n^{-3})]. \tag{32}
\end{aligned}$$

Thus $B_{n\infty}^{ES}$ and $B_{n\infty}^{VaR}$ behave similarly with respect to increases in n , while $B_{n\infty}^{stdv}$ just has the diversification speed $1/n$. ■

Thus the change in the diversification benefits as n increases differs as measured by $B_{n\infty}^{stdv}$ vis-a-vis the changes in the benefits as measured through $B_{n\infty}^{VaR}$ and $B_{n\infty}^{ES}$. In particular, from the Proposition (2) the following result is immediate:

Corollary 3 *The changes in the diversification benefits as measured through the downside risk measures $B_{n\infty}^{VaR}$ and $B_{n\infty}^{ES}$ are identical up to a scalar multiple*

$$\frac{\partial B_{n\infty}^{VaR}}{\partial n} / \frac{\partial B_{n\infty}^{ES}}{\partial n} = \frac{\alpha v_1^{1/\alpha} [1+o(1)] + (\alpha-1) R_f}{\alpha v_1^{1/\alpha} [1+o(1)] + \alpha R_f}. \tag{33}$$

Proof. Divide (31) by (32). ■

Eventually when n grows large, the diversification benefits decline to zero, i.e.

$$B_{\infty\infty}^{stdv} = B_{\infty\infty}^{VaR} = B_{\infty\infty}^{ES} = 0.$$

For large but finite n , we have

Corollary 4 Suppose that the tail of the distribution of $r(n)$ satisfies (20).

Moreover, assume $\alpha > 2$. Then there exist positive constants V, U, Z , such that¹⁰

$$\frac{B_{n\infty}^{stdv}}{V} \sim n^{-1}$$

while

$$\frac{B_{n\infty}^{VaR}}{U} = \frac{B_{n\infty}^{ES}}{Z} \sim v_2 n^{-1} + v_3 n^{1-\alpha} + O(n^{-2}).$$

Hence

$$\frac{V B_{n\infty}^{VaR}}{U B_{n\infty}^{stdv}} = \frac{V B_{n\infty}^{ES}}{Z B_{n\infty}^{stdv}} \sim v_2 + v_3 n^{2-\alpha} + O(n^{-1}).$$

Proof. From the derivatives (30), (31) and (32) and noting $\frac{\partial n^{-1}}{\partial n} = -n^{-2}$

and

$$\frac{\partial}{\partial n} [v_2 n^{-1} + (1 - \alpha) v_3 n^{1-\alpha} + O(n^{-2})] = -v_2 n^{-2} + (1 - \alpha) v_3 n^{-\alpha} + O(n^{-3}),$$

one shows that $B_{n\infty}^{stdv}/n^{-1}$, $B_{n\infty}^{VaR}/[v_2 n^{-1} + (1 - \alpha) v_3 n^{1-\alpha} + O(n^{-2})]$

and $B_{n\infty}^{ES}/[v_2 n^{-1} + (1 - \alpha) v_3 n^{1-\alpha} + O(n^{-2})]$ tend to fixed limits V, U and Z

by using l'Hospital's rule. ■

In the end, the comparison between $B_{n\infty}^{stdv}$, $B_{n\infty}^{VaR}$ and $B_{n\infty}^{ES}$ in the limit or for very large n is of lesser interest than the comparison at moderate values of n , since the point of the paper is that the benefits have to be balanced with the costs. For small n , the result (33) definitely applies as well. The comparison with $B_{n\infty}^{stdv}$ is more involved, since $B_{n\infty}^{VaR}$ and $B_{n\infty}^{ES}$ have two parts changing at

¹⁰Recall that $u(n) \sim v(n)$ stands for $\lim_{n \rightarrow \infty} u(n)/v(n) = 1$.

different rates with respect to n . The idiosyncratic part vanishes at rate $n^{1-\alpha}$, while the market part, due to the influence of the different β_i , vanishes at rate n^{-1} . This is in contrast to the result for $B_{n\infty}^{stdv}$, where both parts have the same diversification speed $1/n$.

In the expressions for the variance, VaR and ES, we look at the speed of convergence due to the idiosyncratic part and market part together. As assets in the portfolio are drawn randomly, the average of β_i is random as well. The average of the β_i converges to their mean, but this speed of convergence is slower than the idiosyncratic part. It dominates in the end under heavy tails as the Proposition 2 shows. To make this distinction clear, suppose that assets are drawn from the class of assets with the same quantity of market risk, i.e. have the same beta. Then we have the following result:

Proposition 3 *Assume the conditions of Proposition 2 and $\text{Var}[\beta_i] = 0$ for all i , then the speed of diversification benefits of $B_{n\infty}^{VaR}$ and $B_{n\infty}^{ES}$ are higher than $B_{n\infty}^{stdv}$, since*

$$B_{n\infty}^{VaR} = O(n^{-\alpha+1}), \quad B_{n\infty}^{ES} = O(n^{-\alpha+1})$$

but

$$B_{n\infty}^{stdv} = O(n^{-1}).$$

Proof. From (30), (31) and (32) in case $\sigma_\beta^2 = 0$

$$\frac{\partial B_{n\infty}^{stdv}}{\partial n} = -EP \frac{1/2}{\beta \sigma_{mkt}} \frac{\sigma_q^2}{\sqrt{\sigma_{mkt}^2 \beta^2 + \frac{1}{n} \sigma_q^2}} n^{-2},$$

and since $\sigma_\beta^2 = 0$ implies that $v_2 = 0$

$$\frac{\partial B_{n\infty}^{VaR}}{\partial n} = \frac{EP(v_1 + v_3 n^{1-\alpha})^{1/\alpha-1} [1 + o(1)]}{\alpha v_1^{1/\alpha} [1 + o(1)] + \alpha R_f} (1 - \alpha) v_3 n^{-\alpha},$$

and analogously,

$$\frac{\partial B_{n\infty}^{ES}}{\partial n} = \frac{EP(v_1 + v_3 n^{1-\alpha})^{1/\alpha-1} [1 + o(1)]}{\alpha v_1^{1/\alpha} [1 + o(1)] + (\alpha - 1) R_f} (1 - \alpha) v_3 n^{-\alpha}.$$

Thus $\partial B_{n\infty}^{stdv} / \partial n = O(n^{-2})$, while $\partial B_{n\infty}^{VaR} / \partial n = \partial B_{n\infty}^{ES} / \partial n = O(n^{-\alpha})$. These results then imply the claims for the level of the diversification benefits. ■

We provide some intuition for this result. Under independence the variance of the sum is the sum of the variances, since the expectations operator is linear in the squares, regardless the distribution (as long as second moment is bounded). Since the (tail) shape of the normal distribution is entirely determined by its variance, the speed of diversification is the same for the global risk measure and the downside risk measure under normality. Holding the risk level constant, diversification in case of the normal distribution then requires that the loss level x is reduced by the square root of the number of assets, since this is how the standard deviation changes; recall that the ratio of the squared loss level x to the variance determines the tail shape of the normal density $\exp(-x^2/2\sigma^2)$.

But the tail shape of the distribution of a convolution needs not be linear in the number of the elements in the sum. If returns are heavy tailed distributed, the tail risk changes nonlinearly, even though the variance is still linear in the number of securities. In the case of heavy tails, the tail risk is shaped like the

power of the Pareto distribution $x^{-\alpha}$. An equally weighted sum of heavy tailed random variables has the same tail shape, but with a different scale: $n^{1-\alpha}x^{-\alpha}$, see (18) or (20). Holding the risk level p constant then requires that the loss level x must be **increased** by $n^{1-1/\alpha}$ (upon inversion $x \simeq n^{-1+1/\alpha}p^{-1/\alpha}$). Since for stocks and bonds it is an empirical fact that the tail shape parameter α is larger than two (consistent with a finite variance), the diversification speed is higher than the square root of the normal case (i.e. $1 - 1/\alpha > 1/2$).

This result may appear to conflict with the fact that the normal distribution has exponentially light tails, while the Pareto distribution has power like fat tails. This difference in type of tails indeed explains why by lowering the loss level sufficiently, any power rate is eventually always beaten by the exponential decline of the normal distribution. But for the diversification result, one holds the risk level constant and evaluates by how much the loss level (VaR) has to change to compensate for the increased level of diversification. For both types of distributions this compensation occurs at a power rate. In the normal case, the variance changes linearly in the number of assets.¹¹ For heavy tails, the scale changes by one minus the tail shape parameter. In the case that the variance is finite, this scale declines more rapidly than the variance. This gives the perhaps somewhat counter intuitive higher diversification speed in the case of fat tails and downside risk when assets are drawn from the class with the same quantity of market risk.

¹¹For other light tailed distributions, it is still true that diversification changes the power. But this is generally not a linear function of the variance, e.g. as in the case of the exponential distribution (where diversification would be better than in case of heavy tails, since the sum of exponentially distributed random variables is a gamma distribution).

In general for low values of diversification (small n), we can not determine a priori which measure delivers the larger speed of diversification benefits. This is determined by the scaling constants in (30), (31) and (32). It is therefore an empirical question as to which factor dominates. But the results above suggest that if the idiosyncratic risk term dominates, then the speed may be higher as $n^{1-\alpha}$ declines more rapidly than n^{-1} if $\alpha > 2$.

6 Conclusion

The debate on ‘how many stock make a diversified portfolio’ has centered primarily on the mean-variance model. Some recent evidence has been unfavourable toward this model. This paper adopts different type of risk measures and distributional assumption to explain low diversification in reality. By adopting downside risk measures with heavy tail assumption, it is shown that the diversification puzzle can well characterized despite simple rule of optimization.

We compare the benefits of portfolio diversification under alternative risk measures. In particular we consider the standard mean-variance criterion and two popular downside risk measures. The latter measures are sensitive to the qualitative behavior of the tail of the return distribution, whereas the variance measure is not. Given that returns are not normally but fat tailed distributed, our main result shows that the downside risk measures imply a higher speed of diversification than under normality or under the mean variance criterion. This may appear counterintuitive, as indeed for sufficiently low the risk levels any

power rate (of the heavy tailed distribution) is dominated by the exponentially declining tail of the normal distribution.

For the diversification result, though, one holds the risk level constant and evaluates by how much the loss level (VaR) has to change. The diversification result operates differently on light and heavy tails. For the light tails, the power is changed, for the heavy tails the scale changes. In the normal case, the power changes through the (linear) change in the variance (and hence the square root for the standard deviation). For heavy tails, the scale changes by one minus the tail shape parameter. In the case that the variance is finite, this scale declines more rapidly than the variance as the tail shape parameter exceeds two. At a given risk level, this implies that diversification reduces the loss level at a rate equal to one minus the inverse of this power. Since for stocks and bonds it is an empirical fact that this power is larger than two (consistent with a finite variance), the diversification speed is higher than the square root of the normal case.

The incremental benefits of diversification were balanced against an incremental cost measure and optimal portfolio sizes were derived, theoretically and empirically by means of a simulation study using actual return data. If investors have some skill in selecting the lower risk stocks, the apparent low level of diversification observed in practice emerges from the analysis, while this is much harder to rationalize by using the mean-variance framework. Our downside risk measures may not be the only avenues to rationalize revealed investor preference for low diversification. But in some form the downside risk criterion may

be necessary to pick up the fat tail phenomenon that leads to the increased speed of diversification to explain the empirically observed small sized portfolios. Conversely, from a normative perspective, our theory suggests that safety first investors should hold focussed portfolios given the fat tailed nature of asset return distributions.

References

- [1] Ang, A., Chen, J., Xing, Y., 2006. Downside Risk. *The Review of Financial Studies* 19, 1191-1239.
- [2] Arzac, E.R., Bawa, V.S., 1977. Portfolio choice and equilibrium in capital markets with safety-first investors. *Journal of Financial Economics* 4, 277-288.
- [3] Beck, K.L., Perfect, S.B., Peterson, P.P., 1996. The role of alternative methodology on the relation between portfolio size and diversification. *The Financial Review* 31, 381-406.
- [4] Benartzi, S., 2001. Excess extrapolation and the allocation of 401(K) accounts to company stocks. *Journal of Finance* 56, 1747-1764.
- [5] Benartzi, S., Thaler H., 2001. Naive diversification strategies in retirement saving plans. *American Economic Review* 91, 79-98.
- [6] Bingham, N.H., Goldie, C.M., Teugels, J.L., 1987. *Regular Variation*, Cambridge University Press, Cambridge.
- [7] Campbell, J., Lettau, M., Malkiel, B., Xu, Y., 2001. Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. *Journal of Finance* 56, 1-43.
- [8] Dacarogna, M.M., Müller, U.A., Pictet, O.V., de Vries, C.G., 2001. Extremal foreign exchange returns in extremely large data sets. *Extremes* 4, 105-127.

- [9] Danielsson, J., Jorgensen, B.N., Sarma, M., de Vries, C.G., 2006. Comparing downside risk measures for heavy tailed distributions. *Economics Letters* 92, 202-208.
- [10] DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy? *The Review of Financial Studies* 22, 1915-1953.
- [11] Domian, D.L., Louton, D.A., Racine, M.D., 2007. Diversification in portfolio of individual stocks: 100 stocks are not enough. *The Financial Review* 42, 557-570.
- [12] Elton, E.J., Gruber, M.J., 1978. Risk reduction and portfolio size: An analytical solution. *Journal of Business* 50, 415-437.
- [13] Elton, E.J., Gruber, M.J., 1995. *Modern Portfolio Theory and Investment Analysis*. 5th ed., Wiley, New York.
- [14] Evans, J.L., Archer, S.H., 1968. Diversification and reduction of dispersion: an analytical solution. *Journal of Finance* 23, 761-767.
- [15] Fama, E.F., Miller, M.H., 1972. *The Theory of Finance*, Dryden Press, Hinsdale.
- [16] Feller, W., 1971. *An Introduction to Probability Theory and Its Applications*. Vol. II, Wiley, New York.
- [17] Goetzmann, W., Kumar, A., 2008. Equity Portfolio Diversification. *Review of Finance* 12, 433-463.

- [18] Gouriéroux, C., Laurent, J.P., Scaillet, O., 2000. Sensitivity analysis of values at risk. *Journal of Empirical Finance* 7, 225-246.
- [19] Hyung, N., de Vries C.G., 2005. Portfolio diversification effects of downside risk. *Journal of Financial Econometrics* 3, 107-125.
- [20] Hyung, N., de Vries C.G., 2012. Simulating and calibrating diversification against black swans. *Journal of Economic Dynamics & Control*, 36, 1162-1175.
- [21] Jacob, N.L., 1974. A Limited-Diversification Portfolio Selection Model for the Small Investor. *Journal of Finance* 29, 837-859.
- [22] Jansen, D., de Vries, C.G., 1991. On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective. *Review of Economics and Statistics* 73, 18-24.
- [23] Johnson, K.H., Shannon, D.S., 1974. A Note on Diversification and the Reduction of Dispersion. *Journal of Financial Economics* 1, 365-372.
- [24] Polkovnichenko, V., 2005. Household Portfolio Diversification. *The Review of Financial Studies* 18, 1467-1502.
- [25] Roy, A.D., 1952. Safety first and the holding of assets. *Econometrica* 20, 431-449.
- [26] Shawky, B., Smith, D.R., 2005. Portfolio selection, diversification and fund-of-fund: a note. *Accounting and Finance* 45, 185-197.

- [27] Statman, M., 1987. How many stocks make a diversified portfolio? *Journal of Financial and Quantitative Analysis* 22, 353-363.
- [28] Statman, M., 2004. The Diversification Puzzle. *Financial Analysts Journal* 60, 44-53.
- [29] Tang, G.Y.N., 2004. How efficient is naive portfolio diversification? an educational note. *Omega* 32, 155-160

Appendix

In this Appendix we derive the expressions of the VaR and ES measures for the levered portfolio $P(n^*)$, and give the Breiman and Feller lemmas used in the main text. We also provide the proof of Corollary 1 and Corollary 2.

A. Value-at-Risk and ES of the $P(n^*)$

First, we show that $VaR_{n^*} = \omega VaR_m - (1 - \omega) R_f$. Let the levered portfolio $P(n^*)$ be constructed from the m -stock portfolio $P(m)$ with weight ω and the risk-free asset with weight $(1 - \omega)$. Then VaR_{n^*} for given probability δ is defined by

$$VaR_{n^*} =: \Pr \{r_{n^*} \leq -VaR_{n^*}\} = \delta,$$

where $r_{n^*} = \omega r_m + (1 - \omega) R_f$, r_m and R_f are the returns of $P(m)$ and the risk-free asset. Similarly VaR_m is defined as

$$\Pr \{r_m \leq -VaR_m\} = \delta.$$

Now

$$\Pr \{\omega r_m + (1 - \omega) R_f \leq -VaR_{n^*}\} = \Pr \left\{ r_m \leq -\frac{VaR_{n^*} + (1 - \omega) R_f}{\omega} \right\}.$$

Thus

$$VaR_m = \frac{VaR_{n^*} + (1 - \omega) R_f}{\omega},$$

which implies (3) from the main text.

For the expected shortfall, we can derive $ES_{n^*} = \omega ES_m - (1 - \omega) R_f$ by using a similar method. The ES_{n^*} for given probability δ is defined by

$$ES_{n^*} = -E[r_{n^*} | r_{n^*} \leq -VaR_{n^*}].$$

Now

$$\begin{aligned} E[r_{n^*} | r_{n^*} \leq -VaR_{n^*}] &= E[\omega r_m + (1 - \omega) R_f | r_{n^*} \leq -VaR_{n^*}] \\ &= \omega E[r_m | r_{n^*} \leq -VaR_{n^*}] + (1 - \omega) R_f \\ &= \omega E[r_m | r_m \leq -VaR_m] + (1 - \omega) R_f \\ &= -\omega ES_m + (1 - \omega) R_f, \end{aligned}$$

where

$$r_{n^*} \leq -VaR_{n^*} \iff r_m \leq -\omega VaR_m$$

since $r_{n^*} = \omega r_m + (1 - \omega) R_f$ and $VaR_{n^*} = \omega VaR_m - (1 - \omega) R_f$.

B.1 Breiman's Lemma

Let X, Y be independent non-negative random variables. Suppose X has a distribution with a regularly varying upper tail such that for $x \geq q$

$$P\{X > x\} = x^{-\alpha} L(x), \quad \alpha > 0 \tag{A1}$$

and where $L(x)$ is a slowly varying function; that is

$$\lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = 1, \text{ for any } x > 0.$$

Assume Y has a distribution $G(y)$ with support $[0, a]$; furthermore Y has continuous density $g(y)$. Then as $t \rightarrow \infty$

$$P\{XY > t\} = t^{-\alpha} L(t) E[Y^\alpha] (1 + o(1)).$$

Proof. By the conditioning argument of Breiman (1965) and since $Y \leq a$, as $t \rightarrow \infty$

$$\begin{aligned} \frac{P\{XY > t\}}{P\{X > t\}} &= \frac{E_y[P\{X > t/y\}]}{P\{X > t\}} \\ &= \int_0^1 \frac{P\{X > t/y\}}{P\{X > t\}} g(y) dy + \int_1^a \frac{P\{X > t/y\}}{P\{X > t\}} g(y) dy \\ &\leq \int_0^1 1 dG(y) + \int_1^a \frac{P\{X > t/a\}}{P\{X > t\}} g(y) dy \\ &= G(1) + \frac{P\{X > t/a\}}{P\{X > t\}} [1 - G(1)] \\ &= G(1) + \frac{(t/a)^{-\alpha} L(t/a)}{t^{-\alpha} L(t)} [1 - G(1)] \\ &= G(1) + a^\alpha \frac{L(t/a)}{L(t)} [1 - G(1)] \\ &\rightarrow G(1) + a^\alpha [1 - G(1)] < \infty. \end{aligned}$$

Thus by the Lebesgue Convergence Theorem, we may interchange the limit and

the integral in

$$\begin{aligned}
\lim_{t \rightarrow \infty} \frac{P\{XY > t\}}{P\{X > t\}} &= \lim_{t \rightarrow \infty} \int_0^a \frac{P\{X > t/y\}}{P\{X > t\}} g(y) dy \\
&= \int_0^a \lim_{t \rightarrow \infty} \frac{P\{X > t/y\}}{P\{X > t\}} g(y) dy \\
&= \int_0^a y^\alpha g(y) \lim_{t \rightarrow \infty} \frac{L(t/y)}{L(t)} dy \\
&= \int_0^a y^\alpha g(y) dy \\
&= E[Y^\alpha].
\end{aligned}$$

■

B.2 Feller's Theorem

We briefly introduce Feller's convolution theorem (1971, VIII.8). This is needed to calculate convolutions of heavy tailed random variables. The convolution result is also used to determine the downside interdependence. The Feller theorem holds that if two independent random variables A and B satisfy (A1)

$$P\{A > t\} = P\{B > t\} = t^{-\alpha} L(t),$$

then their convolution satisfies

$$\lim_{t \rightarrow \infty} \frac{P\{A + B > t\}}{2t^{-\alpha} L(t)} = 1,$$

and where $L(t)$ is slowly varying (i.e. $\lim_{t \rightarrow \infty} L(at)/L(t) = 1$, for any $a > 0$). In other words, the theorem implies that for large failure levels t , the convolution of A and B can be approximated by the sum of the marginal distributions of A and B . All that counts for the probability of the sum is the (marginal) probability mass that is located along the two axes above the points where the line $A + B = t$ cuts the axes.

To show this, first note that since A and B are independent Pareto distributed

$$1 - P\{A \leq t, B \leq t\} = 1 - [1 - t^{-\alpha}]^2 = 2t^{-\alpha} - t^{-2\alpha} \simeq 2t^{-\alpha}$$

as $\lim_{t \rightarrow \infty} (2t^{-\alpha} - t^{-2\alpha})/t^{-\alpha} = 2$. Since (for positive random variables)

$$P\{A + B > t\} > 1 - P\{A \leq t, B \leq t\},$$

we have the bound $P\{A + B > t\} > 2t^{-\alpha}$. The Feller theorem maintains that $P\{A + B > t\}$ is in fact approximately $2t^{-\alpha}$ as t becomes large. To verify this, we demonstrate that

$$P\{A + B > t\} - [1 - P\{A \leq t, B \leq t\}],$$

which comprises the probability mass in the triangle above the line $A + B = t$ (with vertices $(0, t)$, $(t, 0)$ and (t, t)), is of an order smaller than $t^{-\alpha}$. Note that

by independence, for $\lambda \in (0, 1)$

$$P\{A > \lambda t, B > (1 - \lambda)t\} = \left(\frac{1}{\lambda(1 - \lambda)}\right)^\alpha t^{-2\alpha}.$$

Thus for any slab above the line $A + B = t$ and with vertex at $(\lambda, 1 - \lambda)$ on the line $A + B = t$, the probability mass is of an order smaller than $t^{-\alpha}$ (i.e. $\lim_{t \rightarrow \infty} t^{-2\alpha}/t^{-\alpha} = 0$). Note that this slab partly covers the triangle. By varying λ , this shows that the entire triangle must carry probability mass of an order smaller than $t^{-\alpha}$.

B.3 Expansion of expectation

We make an assumption regarding the distribution of the beta's.

Condition 1 *The β_i are positive and finite. As a result for the random stock-picker the β_i are i.i.d with all moments finite.*

It is not directly easy to see what the order of magnitude of

$$E \left[\left(\frac{1}{n} \sum_{i=1}^n \beta_i \right)^\alpha \right]$$

is.¹² Denote the first moment by

$$\beta = E[\beta_i],$$

¹²We are grateful to Laurens de Haan and Zhou Chen for insightful suggestions and discussions on this part.

and note that

$$E \left[\frac{1}{n} \sum_{i=1}^n \beta_i \right] = \beta.$$

Denote the variance by

$$E \left[(\beta_i - \beta)^2 \right] = \sigma_\beta^2;$$

hence

$$E \left[\left(\frac{1}{n} \sum_{i=1}^n \beta_i - \beta \right)^2 \right] = \frac{1}{n} \sigma_\beta^2.$$

Define the rescaled random variables

$$\hat{\beta}_i = \frac{\beta_i - \beta}{\beta}$$

and

$$Q_n = \frac{1}{n} \sum_{i=1}^n \hat{\beta}_i.$$

By the assumption on the moments of β , all moments of Q_n are finite as well.

In particular $E[Q_n] = 0$ and $E[Q_n^2] = \sigma_\beta^2 / (n\beta^2)$. Further

$$\begin{aligned} E[Q_n^3] &= \frac{1}{n^3 \beta^3} \left(n E_\beta [(\beta_i - \beta)^3] + 3(n^2 - n) E[(\beta_i - \beta)^2] E[(\beta_i - \beta)] \right. \\ &\quad \left. + (n^3 - 3n^2 + 2n) (E[(\beta_i - \beta)])^3 \right) \\ &= \frac{1}{n^2} \frac{S_\beta}{\beta^3}, \end{aligned}$$

where $S_\beta = E[(\beta_i - \beta)^3] \in \mathbb{R}$ denotes the third moment. Lastly

$$\begin{aligned}
E[Q_n^4] &= \frac{1}{n^4 \beta^4} \left(nE[(\beta_i - \beta)^4] + 4(n^2 - n)E[(\beta_i - \beta)^3]E[(\beta_i - \beta)] \right. \\
&\quad + 3(n^2 - n)E[(\beta_i - \beta)^2]E[(\beta_i - \beta)^2] \\
&\quad \left. + n(n-1)(n-2)(n-3)(E[(\beta_i - \beta)])^4 \right) \\
&= \frac{1}{n^3} \frac{K_\beta}{\beta^4} + \frac{3(n-1)}{n^3} \frac{\sigma_\beta^4}{\beta^4}
\end{aligned}$$

where $K_\beta = E[(\beta_i - \beta)^4]$ is the fourth moment of β . In general, note that for any integer $j > 0$

$$E[Q_n^j] = O(1/n^{j-1}).$$

Note that by the Condition 1 the positive support implies that $1 + Q_n \geq 0$.

For the lower bound use that for $\alpha > 2$, by Jensen's inequality and Taylor's theorem for some $\varphi \in [0, 1]$

$$\begin{aligned}
E\left[\left(\frac{1}{n} \sum_{i=1}^n \beta_i\right)^\alpha\right] &= \beta^\alpha E[(1 + Q_n)^\alpha] \\
&= \beta^\alpha E\left[\left((1 + Q_n)^2\right)^{\alpha/2}\right] \\
&\geq \beta^\alpha \left(E[(1 + Q_n)^2]\right)^{\alpha/2} \\
&= \beta^\alpha \left\{1 + \frac{\sigma_\beta^2}{\beta^2} \frac{1}{n}\right\}^{\alpha/2} \\
&= \beta^\alpha \left\{1 + \frac{\alpha}{2} \frac{\sigma_\beta^2}{\beta^2} \frac{1}{n} + \frac{\frac{\alpha}{2}(\frac{\alpha}{2} - 1)}{2} \left(\frac{\sigma_\beta^2}{\beta^2} \frac{1}{n}\right)^2 \left(1 + \varphi \frac{\sigma_\beta^2}{\beta^2} \frac{1}{n}\right)^{\frac{\alpha}{2} - 2}\right\} \\
&= \beta^\alpha + \beta^\alpha \frac{\alpha}{2} \frac{\sigma_\beta^2}{\beta^2} \frac{1}{n} + O\left(\frac{1}{n^2}\right).
\end{aligned}$$

For the upper bound we proceed as follows. Let k be the integer closest to α , such that $k \geq \alpha$. By the result on the lower bound, we have that $E[(1 + Q_n)^\alpha] >$

1. From this and Jensen's inequality, we have that

$$E[(1 + Q_n)^k] = E[(1 + Q_n)^\alpha]^{k/\alpha} \geq (E[(1 + Q_n)^\alpha])^{k/\alpha} > E[(1 + Q_n)^\alpha] > 1.$$

By the binomial theorem

$$\begin{aligned} E[(1 + Q_n)^k] &= E\left[1 + kQ_n + \frac{k(k-1)}{2}Q_n^2 + \dots + \binom{k}{j}Q_n^j + \dots + Q_n^k\right] \\ &= 1 + \frac{k(k-1)}{2} \frac{\sigma_\beta^2}{\beta^2} \frac{1}{n} + \dots + \binom{k}{j} E[Q_n^j] + \dots + E[Q_n^k]. \end{aligned}$$

Since $E[Q_n^j] = O(1/n^{j-1})$,

$$E[(1 + Q_n)^k] = 1 + \frac{k(k-1)}{2} \frac{\sigma_\beta^2}{\beta^2} \frac{1}{n} + O\left(\frac{1}{n^2}\right).$$

To conclude, we can sandwich $E_\beta[(1 + Q_n)^\alpha]$ as follows

$$1 + \frac{k(k-1)}{2} \frac{\sigma_\beta^2}{\beta^2} \frac{1}{n} + O\left(\frac{1}{n^2}\right) \geq E_\beta[(1 + Q_n)^\alpha] \geq 1 + \frac{\alpha}{2} \frac{\sigma_\beta^2}{\beta^2} \frac{1}{n} + O\left(\frac{1}{n^2}\right).$$

Thus for some $W \in [\alpha, k(k-1)]$

$$E[(1 + Q_n)^\alpha] = 1 + \frac{W}{2} \frac{\sigma_\beta^2}{\beta^2} \frac{1}{n} + O\left(\frac{1}{n^2}\right).$$

This gives

$$E \left[\left(\frac{1}{n} \sum_{i=1}^n \beta_i \right)^\alpha \right] = \beta^\alpha \left\{ 1 + \frac{W}{2} \frac{\sigma_\beta^2}{\beta^2} \frac{1}{n} + O(n^{-2}) \right\}.$$

C. Proof of Corollary 2

Assume that all means and variances exist. So let

$$E[\beta_i] = \beta; E[(\beta_i - \beta)^2] = \sigma_\beta^2,$$

$$E[q_i] = Q; E[(q_i - Q)^2] = \sigma_q^2,$$

and

$$E[r_{mkt}] = R; E[(r_{mkt} - R)^2] = \sigma_{mkt}^2.$$

Then by the independence of the three random variables

$$E[r_i] = E[\beta_i r_{mkt} + q_i] = \beta R + Q.$$

And

$$Var[r_i] = \sigma_{mkt}^2 \sigma_\beta^2 + R^2 \sigma_\beta^2 + \beta^2 \sigma_{mkt}^2 + \sigma_q^2.$$

For the portfolio

$$Var \left[\frac{1}{n} \sum_{i=1}^n r_i \right] = \frac{1}{n} \sigma_{mkt}^2 \sigma_\beta^2 + \frac{1}{n} R^2 \sigma_\beta^2 + \beta^2 \sigma_{mkt}^2 + \frac{1}{n} \sigma_q^2 \rightarrow \beta^2 \sigma_{mkt}^2$$

as $n \rightarrow \infty$.

In effect we get the result that both the idiosyncratic part σ_q^2/n and the part due to the uncertainty in the betas $(\sigma_{mkt}^2\sigma_\beta^2 + R^2\sigma_\beta^2)/n$ vanish at the same speed. Note again that for this result we do not have to assume normality.