Mathematical Foundation for Ensemble Machine Learning and Ensemble Portfolio Analysis

Eugene Pinsky
Department of Computer Science
Metropolitan College, Boston University
Boston, MA 02215
email: epinsky@bu.edu

August 21, 2018

1 Introduction

In many areas, combination of models often perform better than individual models (for a survey see [1]). This paper focuses on application of this approach to construct large-cap portfolios from individual large-cap mutual funds. The main performance benchmark for these large-cap managers is the S&P-500 index. Managers have their own criteria and constraints in choosing their portfolio of stocks. The weights of stocks will be heavily dependent on the optimization function (e.g. mean-variance) and other constraints that could be specific for individual managers. In constructing his/her portfolio, we can think of such a manager as "voting" that the chosen stocks in his/her sub-portfolio will be enough to beat the market index. We can combine such individual voting decisions and consider constructing an ensemble portfolio that, in some sense, represents the combined voting of individual managers. The motivation for such an approach is to improve the accuracy of predictions when the decision is made by several experts. In addition,

We can consider a number of ways to construct such an ensemble portfolio:

(1) The simplest approach is to take K managers portfolios and allocate a fraction $1/K$ of the money in each of K portfolio at each re-balancing date. Throughout this paper, we will refer to this as $1/K$ portfolio. Such a construction is easy to implement as it does not involve any computation of returns or volatility. The main motivation for such a construction is the reduction in volatility of the resulting portfolio. The extend of this decrease in volatility would depend on the correlation between portfolios. Despite its simplicity, this naive $1/K$ model behaves quite well. A comparison of this model with 14 other standard models evaluated in [2] showed that "that none of the other models was considerable better the naive model in terms of Sharpe

ratio, certainty equivalent return, or turnover". As pointed in [7], we do not have a definitive answer as to what is the best way to combine portfolios in practice.

(2) We could identify stocks in portfolios that are assigned weights higher then the corresponding weight in the SP-500 index. Managers may decide to overweight these "concentrated" stocks because they believe that a sub-portfolio of these stocks will allow them to generate returns in excess of the market return (the so-called α) with acceptable risk. We can then construct portfolios from these a sub-portfolios (with the same relative weights of concentrated stocks) and construct a portfolio as equal-weighted average of these portfolios.

In this paper, we consider the ensemble portfolio constructed using the method (2) described above, namely a portfolio constructed by averaging sub-portfolios concentrated stocks. As indicated above, if we think of each sub-portfolio as analogous to a vote that such a sub-portfolio will beat the market index, then constructing an ensemble portfolio is analogous to ensemble voting by independent experts to make a decision to invest in a subset of stocks.

Accordingly, we divide this paper in two parts. In the first part, we focus on ensemble voting. Specifically, we will show that if we have a set of managers (with probability of making a wrong decision $p < 0.5$) then increasing the ensemble size (number of managers) increases the accuracy. If managers are independent, then we can decrease the probability of an error (i.e. probability that the sub-portfolio will not beat the market) as we keep increasing the number of managers for the construction of the portfolio.

In the second part of the paper, we focus on performance characteristics of ensemble portfolio itself. Specifically, we will show that such a portfolio provides higher return than the corresponding stocks in the index and its risk is less than the average risk across managers' sub-portfolios. When compared to the naive $1/K$ strategy, the ensemble portfolio provides much higher return with a modest increase in volatility. As a result, such a portfolio exhibits a much higher Sharpe'e ratio.

Finally we should note that

2 Part I: Analysis of Ensemble Voting

In this part we focus on analyzing probability in ensemble voting. Our general setting is the following. We assume that we have fund managers with each fund manager making an independent decision to invest or not to invest. We make an investment decision based on the majority decision of these managers.

Formally, we describe the model as follows: we have $N = 2n + 1$ managers. We need an odd number of managers to use for the majority voting. For each manager i, let $p_i < 1$ denote the probability that this manager makes a mistake in his/her prediction. We can then model this by a simple Bernouilli distribution [3].

Let random variable X_i denote the decision of manager i as follows:

$$X_i = \begin{cases} 1 & \text{if manager } i \text{ makes a mistake} \\ 0 & \text{otherwise} \end{cases}$$

The probability distribution $P()$ for each manager is given by a simple rule

$$P(X_i) = \begin{cases} p_i & \text{if } X_i = 1 \\ q_i = 1 - p_i & \text{otherwise} \end{cases}$$

Define random variable $X = X_1 + X_2 + \cdots + X_N$. This random variable takes values from 0 (when all managers make the correct decision) to N (when all managers make the wrong decision). The expected value of this variable is $E[X] = np$ and its variance $\sigma^2(X) = np(1-p)$. In particular, $X \leq n$ denotes the event when the majority of the managers make the right decision.

We would like to investigate when voting with ensemble of managers gives us a higher probability of success than using just one manager (even the best one with the lowest p_i). In other words, we would like to investigate when

$$P(X \leq n) > \max_i (1 - p_i)$$

Moreover, we will establish bound on error probability of ensemble voting and show that this probability can be made arbitrarily small by taking enough managers for the ensemble.

To that end, we proceed as follows. We start with a model when all managers are independent and make decision with the same probability. We will then extend this to a more general case when these probabilities are different.

2.1 Independent Fund Managers with Same Prediction Accuracy

In this case, the model of ensemble voting is a binomial distribution model [3]. We have $N = 2n+1$ independent managers. The probability that exactly k managers make incorrect decision is

$$P(X = k) = \binom{N}{k} p^k q^{N-k}$$

We will find it convenient to introduce the following notation for the cumulative distribution function

$$F(k,N,x) = P(X \leq k) = \sum_{i=0}^{k} \binom{N}{i} x^i (1-x)^{N-i}$$

With this definition, the probability of correct decision is $F(n, 2n+1, p)$. In other words, it is the probability that at least $n+1$ fund managers made the correct decision.

Example 1: we have $N = 3$ fund managers. The probabilities and cumulative function can be written explicitly and are summarized in the table below:

Table 1: Voting with 3 Fund Managers

k	$P(X = k)$	$F(k, 3, p)$
0	q^3	q^3
1	$3pq^2$	$q^3 + 3pq^2$
2	$3p^2q$	$q^3 + 3pq^2 + 3p^2q$
3	p^3	1

For example, the probability that exactly one manager makes a wrong decision is

$$P(X = 1) = 3p(1-p)^2$$

With majority voting, the probability of correct prediction is the probability that no more than one manager makes a wrong decision. Therefore, the probability of a right decision by ensemble is:

$$F(1,3,p) = P(X = 0) + P(X = 1) = (1-p)^3 + 3p(1-p)^2$$

How does the error by ensemble compare with that of an individual manager? Specifically, is there any range of p for which it is better to use ensemble voting?

We will show that as long as error probability of individual managers $p < 0.5$, using ensemble voting reduces the error. First, we show it graphically in the following plot below:

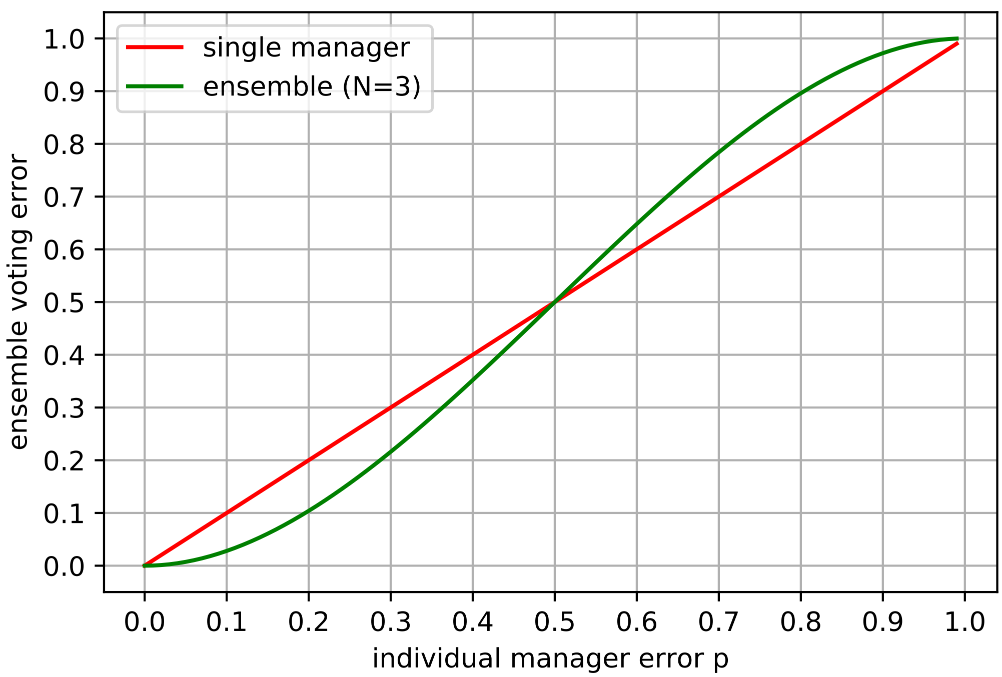

Figure 1: Comparison of 3-manager ensemble with Individual manager

We see that if the probability of error $p < 0.5$, then using ensemble gives you better result than using an individual manager.

We can also show it algebraically. If we were to use a single manager, then the probability of the right decision by such a manager is $(1-p)$. Using ensemble voting, the probability of correct decision is $F(1,3,p)$. Therefore, we are looking for p for which,

$$(1-p)^3 + 3p(1-p)^2 > 1-p$$

Or, equivalently,

$$(1-p)^2 + 3p(1-p) > 1$$

After some elementary algebra, we obtain $p < 0.5$. This result means that for the case of 3 managers, if the probability of error is less than 0.5 then using ensemble voting would result in lower error rate. This result generalizes to the general case of any N.

Example 2: Let us show how accuracy increases with the number of managers. In the plot below, we compare probabilities of voting errors for ensemble sizes $N = 3, 5, 11$.

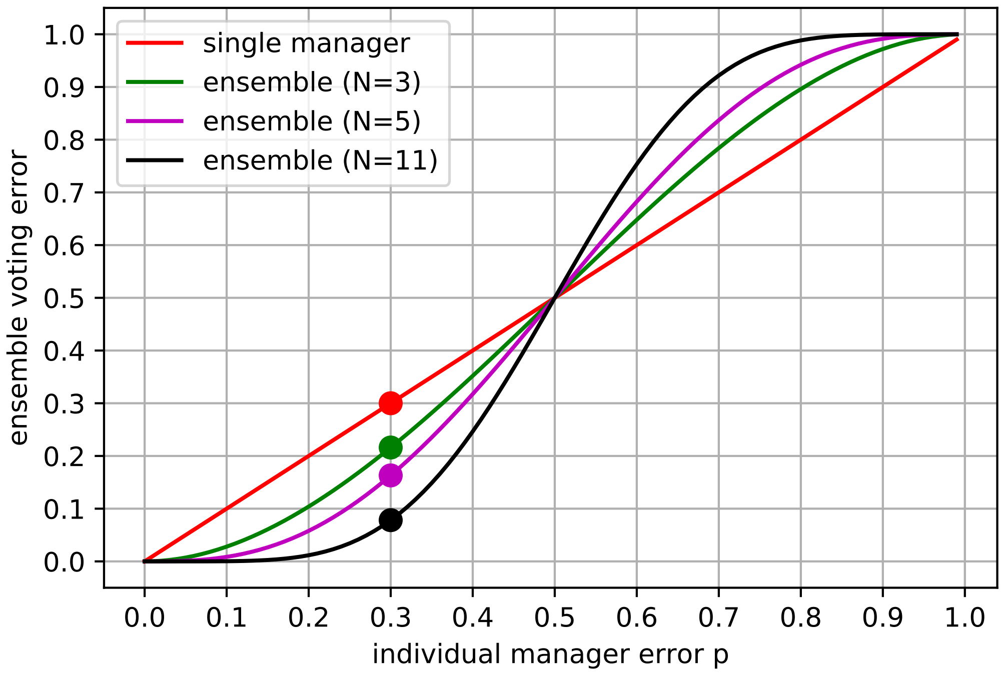

Figure 2: Effect of increasing the number of managers in the ensemble

Just as in the previous case, voting by ensemble gives lower error than using an individual manager as long as the probability of error $p < 0.5$. In this range, we can see that for any probability value, the ensemble voting error decreases as we increase the number of managers.

As an illustration, consider probabilities of voting errors for $p = 0.2$, $p = 0.3$ and $p = 0.4$ The voting errors are by the corresponding ensembles are summarized in the table below:

From this table, we can see that for any number of managers, as we increase p our accuracy decreases. And in fact, once $p.0.5$ using ensemble

Table 2: Comparison of Error probabilities for Different Ensemble Sizes

N	$p = 0.2$	$p = 0.3$	$p = 0.4$
1	0.2	0.3	0.4
3	0.10	0.23	0.35
5	0.06	0.16	0.32
11	0.01	0.08	0.25

voting gives us worse results than using a single manager.

On the other hand, for any p the accuracy increases as we increase the number of managers. For example, take $p = 0.3$. the probability of error drops from 0.23 (in case of 3 managers) to 0.08 in case of 11 managers

2.2 Some General Results

We now establish two general results. As before, let

$$F(k, N, x) = P(X \leq k) = \sum_{i=0}^{k} \binom{N}{i} x^i (1-x)^{N-i}$$

Then since $\binom{N}{k} = \binom{N}{N-k}$, it is easy to show that

$$\begin{aligned} 1 &= \sum_{i=0}^{n} \binom{2n+1}{i} p^i q^{2n+1-i} + \sum_{i=n+1}^{2n+1} \binom{2n+1}{i} p^i q^{2n+1-i} \\ &= \sum_{i=0}^{n} \binom{2n+1}{i} p^i q^{2n+1-i} + \sum_{i=0}^{n} \binom{2n+1}{i} q^i p^{2n+1-i} \\ &= F(n, 2n+1, p) + F(n, 2n+1, 1-p) \end{aligned} \tag{1}$$

We will find it convenient to define

$$A_i(n,x) = \binom{2n+1}{i} x^i (1-x)^{2n+1-i}$$

With this definition, we have

$$F(n,2n+1,p) = \sum_{i=0}^{n} A_i(n,p)\ ,\quad F(n,2n+1,1-p) = \sum_{i=0}^{n} A_i(n,1-p)$$

If $p < q$ then $(1-p)/p > 1$ and ,therefore, for each i, the ratio

$$\frac{A_i(n,p)}{A_i(n,1-p)} = \left(\frac{1-p}{p}\right)^{2n+1-2i} > 1$$

.

Therefore, $A_i(n,p) > A_i(n,1-p)$ and this implies that $F(n,2n+1,p) > F(n,2n+1,1-p)$. Since, $F(n,2n+1,p) + F(n,2n+1,1-p) = 1$ we immediately obtain:

$$F(n,2n+1,p) > 0.5$$

We will now strengthen this statement by showing that if $p < 0.5$ than as we increase the size of the ensemble, the accuracy of ensemble voting increases.

Proof: The decision to invest is made by taking the majority decision of weak learners. In the ensemble we must have an odd number $2n+1$ of managers. We therefore proceed by induction on n:

Basic step: This was proven in example 1 where we showed that $F(1,3,p) = (1-p)^3 + 3p(1-p)^2 > 1-p$ for $p < 0.5$

Inductive Step: we need to establish the following:

$$F(n+1, 2(n+1)+1, p) > F(n, 2n+1, p)$$

We use the relation $\binom{n}{i} = \binom{n-1}{i-1} + \binom{n-1}{i-1}$ to obtain

$$\begin{aligned}\binom{2n+3}{i} &= \binom{2n+2}{i} + \binom{2n+2}{i-1} \\ &= \binom{2n+1}{i} + 2\binom{2n+1}{i-1} + \binom{2n+1}{i-2}\end{aligned} \tag{2}$$

From this after some algebraic manipulations we obtain:

$$\begin{aligned}F(n+1, 2n+3, p) = &\binom{2n+1}{n+1} p^{n+1} q^n (p^2 + q^2) \\ &+ F(n, 2n+1, p)\left[q^2 + \frac{2p}{q} + p^2\right] \\ &> F(n, 2n+1, p)\left[q^2 + \frac{2p}{q} + p^2\right]\end{aligned} \tag{3}$$

Since $0 < q < 1$ we have $1/q > 1 > q$ and, in particular, $p/q > pq$. Therefore,

$$q^2 + \frac{2p}{q} + p^2 > q^2 + 2pq + p^2 = 1$$

This immediately gives us the desired result, namely

$$F(n+1, 2n+3, p) > F(n, 2n+1, p)$$

The above result provides the mathematical foundation for using ensemble: if we have a set of weak learners with $p < 0.5$, then increasing the size of the ensemble increases the accuracy.

2.3 Bounds on Ensemble Error Probabilities

Let us now return to the general case of N managers with independent voting and the resulting binomial distribution. The mean of the distribution is $\mu = Np$ and the variance $\sigma^2 = Np(1-p)$. The probability of correct decision by the ensemble is the following cumulative distribution function $F(n, 2n+1, p)$. What we need to do is to establish some bounds on $F(n, 2n+1, p)$.

Intuitively, when we average a set of random variables, we should get something close to the expected value. For a general distribution, the justification for this is the Chebyshev's inequality: if X is a random variable with expectation μ and variance σ^2 then for any $\epsilon > 0$ we have:

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{\epsilon^2}$$

When applied to the binomial distribution with identically distributed random variables X_i (and remembering that $E[X_i) = p$, the Chebyshev's inequality gives us

$$P\left(\Big| \frac{1}{N} \sum_{i=1}^{n} X_i - p \Big| \geq \epsilon \right) \leq \frac{\sigma^2}{N\epsilon^2}$$

In case of a sum of independent random variables, we can obtain even sharper bounds as shown below. Specifically, we propose to use Hoeffding's inequality [4]. Let $X_1, \ldots, X_N$ be random, identically distributed random variables with $0 \leq X_i \leq 1$. Then, the Hoeffding's inequality states that

$$P\left(\Big| \frac{1}{N} \sum_{i=1}^{N} X_i - E[X] \Big| > \epsilon \right) \leq 2\exp(-2N\epsilon^2)$$

When applied to binomial distribution (recall that $E[X] = np$), we have the following

$$P\left(X \geq (p+\epsilon)N\right) \leq \exp(-2N\epsilon^2)$$

Let us see how we can use this for our ensemble voting. The probability of voting error is $P(X \geq n+1)$. If we compute ϵ (and remembering that $N = 2n+1$) as follows

$$(p+\epsilon)(2n+1) = n+1$$

Solving for ϵ, we have

$$\epsilon = \frac{n+1}{2n+1} - p$$

From this immediately follows that the error probability of ensemble is bounded above as follows:

$$P(X \geq n+1) \leq \exp(-2\epsilon^2 N)$$

For any $n > 0$ it is trivial to show that for any N we have $0.5 - p < \epsilon < 1 - p$ and, therefore, ϵ is bounded. It immediately follows that as $N \to \infty$, the term $\exp(-2\epsilon^2 N) \to 0$. Therefore, for fixed $p < 1/2$, as we increase the number of managers we obtain

$$P(X \geq n) \to 0$$

This provides a mathematical justification for using the ensemble voting. This is illustrated in the figure below:

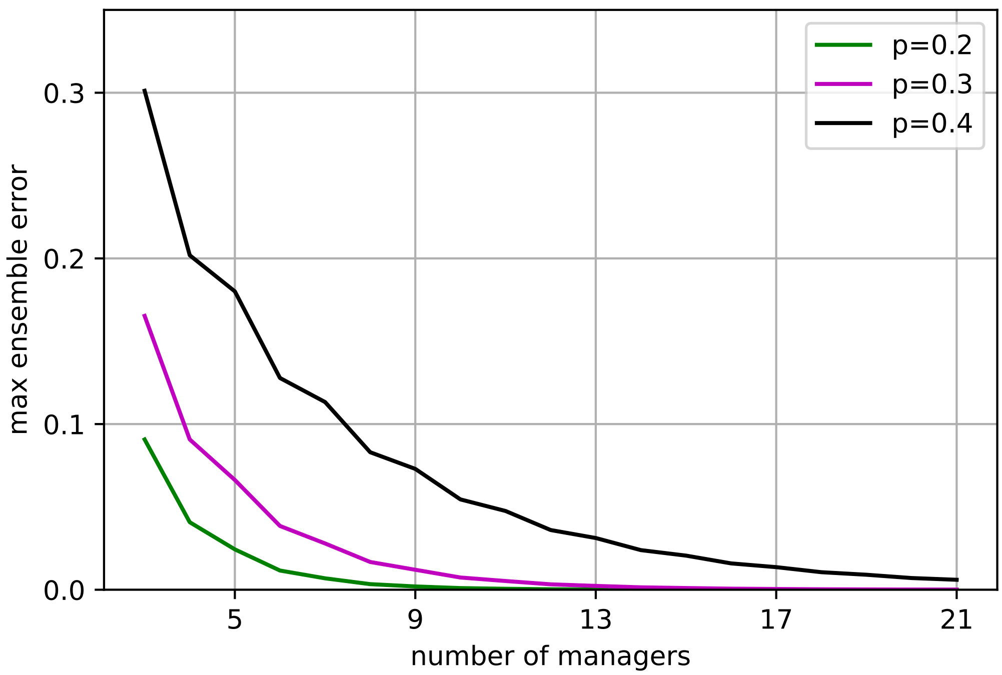

Figure 3: Improvement of Accuracy with Ensemble Size

Examining the figure, we see a sharp decrease in error probability as we increase the number of managers.

2.4 Independent Managers with Non-identical Probabilities

In the previous section, we assumed that manager probabilities are independent and identical. Generalization to independent managers with non-identical probabilities is straightforward.

As before, assume that each manager i makes a decision with probability $p_i < 0.5$. Let $X = X_1 + X_2 + \cdots + X_N$. Let $p = (p_1 + p_2 + \cdots + p_N)/N$. Then, Hoeffding's lemma can be written as follows:

$$P\left(X - Np \geq t\right) \leq \exp\left(-\frac{2t^2}{N}\right)$$

Define $t = N\epsilon$. Then, the Hoeffding's lemma can be re-written as follows:

$$P\left(X \geq (p+\epsilon)N\right) \leq \exp(-2\epsilon^2 N)$$

As before, we compute ϵ from $(p+\epsilon)(2n+1) = n+1$ to obtain

$$\epsilon = \frac{n+1}{2n+1} - p$$

and

$$P(X \geq n+1) \leq \exp(-2\epsilon^2 N)$$

As before, we show that ϵ is bounded, and this would immediately imply that the error probability of ensemble voting

$$P(X \geq n) \to 0$$

as we increase the number of managers in the ensemble.

2.5 Incorporating Dependencies in Ensemble Votingy

In the discussion above, we assumed that individual manager decisions are independent of decisions of other managers. We can think of their choosing portfolios as choosing stocks from an urn with replacement. We now extend our analysis to handle dependency.

There are specific dependency assumptions for the probability distributions for which specialized inequalities can be derived [6]. We will use the most general form of the bounds, applicable to any distribution, namely the one derived from Chebyshev's inequality.

To proceed, let us consider the case where each manager makes a decision with probability p but X_i are not independent. Specifically, we assume that the correlation coefficient for any X_i and X_j is ρ^*. Then, the variance of $X = X_1 + \cdots + X_N$ is

$$\begin{aligned} \sigma^2(X) &= \sum_i \sigma^2(X_i) + \sum_{i \neq j} cov(X_i, X_j) \\ &= Np(1-p) + \sum_{i \neq j} \rho^* \sigma(X_i)\sigma(X_j) \end{aligned} \tag{4}$$

The variance of X_i is $p(1-p)$. Therefore, we can re-write the above expression for the variance as follows:

$$\begin{aligned} \sigma^2(X) &= Np(1-p) + (N^2 - N)\rho^* p(1-p) \\ &= Np(1-p)\Big[1 + (N-1)\rho^*\Big] \end{aligned} \tag{5}$$

If the correlation $\rho^* = 0$ then the above expression reduces to $Np(1-p)$ - the variance of the sum of independently distributed Bernouilli variables. The factor $(N-1)\rho^*$ shows how the variance increases as the correlation ρ is increased. In particular, if $\rho^* << 1/N$ then the correlation would have negligible effect on the variance. If the correlation is high, then we can only get small enough variance if the probability of an error p is very small.

With the above expression for variance of X we can use the Chebyshev inequality for X to obtain

$$P\left(\left|\frac{X}{N}-p\right| \geq \epsilon\right) \leq \frac{\sigma^2(X)}{N\epsilon^2} = \frac{p(1-p)}{\epsilon^2}\left[1+(N-1)\rho^*\right]$$

As before, if we let ϵ be the solution of $(p+\epsilon)(2n+1) = n+1$ and remembering that $N = 2n+1$ then we obtain

$$\epsilon = \frac{n(1-2p)+(1-p)}{2n+1}$$

This immediately gives the following upper bound on the probability of error in the ensemble:

$$P(X \geq n) \leq p(1-p)\left[1+2n\rho^*\right]\left[\frac{n(1-2p)+(1-p)}{2n+1}\right]^{-2}$$

For large n we have

$$\frac{n(1-2p)+(1-p)}{2n+1} \mapsto \frac{1-2p}{2} + O(1/n)$$

This gives us the following bound for the error probability (for large n):

$$P(X \geq n) \leq \frac{4p(1-p)}{(1-2p)^2}\left[1+2n\rho^*\right]$$

The above bound shows that to decrease the probability of an error in the ensemble, we need to make the probability of error of individual managers extremely small (to compensate for the increase in variance du to correlation ρ^*.

In the future, we hope to consider sharper bounds on the ensemble error probabilities.

3 Part II: Analysis of the Ensemble Portfolio

In this section we consider the properties of the ensemble portfolio constructed from "concentrated" sub-portfolios of individual managers. We use the word "concentrated" sub-portfolio to emphasize that we are considering a subset of stocks from portfolios which have higher weight than the corresponding weight in the SP-500 index. The ensemble portfolio is constructed by averaging concentrated sub-portfolios across managers. By contrast, a $1/K$ portfolio is constructed by simply (equal-weighting) averaging of K portfolios.

A manager may decide to overweight particular stocks in a portfolio because the manager believes that this subset of portfolio stocks will provide a necessary return to outperform the market index without taking too much additional risk. In some sense, concentrating some stocks in a portfolio represents a manager's vote that such a portfolio will outperform the market. Therefore, intuitively, the construction of ensemble portfolio represents a result of ensemble voting by the managers that a particular set of stocks will outperform the market.

Our objective is to compare the performance of an ensemble portfolio with the market and with the $1/K$ portfolio and to argue that such ensemble portfolios provide better performance both in terms of returns and risk-adjusted returns than $1/K$ portfolio.

We start with some notation. Assume that we have there are a finite set Ω of $N > 500$ large-cap stocks $\Omega = \{S_1, S_2, \ldots, S_N\}$. Without loss of generality, assume that the first 500 stocks appear in the SP-500 index. A large cap portfolio will consist of some stocks that are in the index and some stocks that are not in the index. There are K managers portfolios are K original portfolios $P_1, ..., P_K$. We use index k to denote portfolios and i to denote security S_i. Let weight of each security S_i in the SP-500 index be v_i and let the weight of each security S_i in portfolio P_k be w_{ik}.

With this notation, we can represent each portfolio as N-dimensional vector of weights. Specifically, the market portfolio M^* is the vector

$$M^* = (v_1, v_2, \ldots, v_{500}, 0, \ldots, 0)$$

whereas a portfolio P_k is the vector

$$P_k = (w_{1k}, w_{2k}, \ldots, w_{Nk})$$

In a portfolio P_k we may have a stock S_i with $w_{ki} > v_i$. In other words, a manager decides to pick a stock and assign its weight to be more than the weight of this stocks in the SP-500 index. A manager is convinced that concentrating such stocks will help the portfolio achieve its goal of outperforming the market index. Therefore, a manager portfolio P_k consists of two disjoint sub-portfolios of stocks: the concentrated sub-portfolio C_k and non-concentrated sub-portfolio U_k. The stocks in C_k have higher weight than corresponding stocks in SP-500. The stocks in U_k are either not in the index or have weight not higher than the corresponding weights in the index. Mathematically, we can write $P_k = C_k + U_k$

Example 1: Assume that we have a universe Ω of just 5 stocks $\Omega = \{S_1, S_2, S_3, S_4, S_5\}$ Assume that the that SP-500 index consists of just 4 stocks $\{S_1, S_2, S_3, S_4\}$ each with the same weight $v_1 = v_2 = v_3 = v_4 = 0.25$. Therefore, the market portfolio is $M^* = (0.25, 0.25, 0.25, 0.25, 0)$. Assume that we have two portfolios P_1 and P_2 with the following composition:

Porfolio 1: 30% of P_1 is S_1 and 30% of P_1 is S_2. The remaining 40% is invested in S_5 that is not in the index. This portfolio is given by the following vector of weights:

$$P_1 = (0.3, 0.3, 0, 0, 0.4)$$

Within this portfolio, the weights for stocks S_1 and S_2 are greater than the corresponding weights in the index. Therefore, the concentrated sub-portfolio C_1 will consist of just two stocks S_1 and S_2 with corresponding weights $w_{11} = 0.3$ and $w_{12} = 0.3$, whereas the un-concentrated sub-portfolio U_1 will consist of just one stock S_5 with weight $w_{15} = 0.4$. Therefore, we can write $P_1 = C_1 + U_1$ where

$$C_1 = (0.3, 0.3, 0, 0, 0) \text{ and } U_1 = (0, 0, 0, 0, 0.4)$$

Portfolio 2: 50% of P_2 is S_1, 30% of P_2 is S_2, 10% of P_2 is S_3 and 10% of P_2 is S_4. This portfolio is given by the following vector of weights:

$$P_2 = (0.5, 0.3, 0.1, 0.1, 0)$$

Note that in this portfolio all stocks are in the market index. However, only the first two S_1 and S_2 have weights greater than the in the index, whereas the stocks S_3 and S_4 have weights less than in the market index. Therefore, we can write $P_2 = C_2 + U_2$ where

$$C_2 = (0.5, 0.3, 0, 0, 0) \text{ and } U_2 = (0, 0, 0, 0.1, 0.1)$$

3.1 Construction of Ensemble and $1/K$ Portfolios

Now we construct our "ensemble" portfolio P^* as follows: From each portfolio P_k, we consider consider the concentrated sub-portfolio C_k. This sub-portfolio will have weights that do not add up to 1 (unless $P_k = C_k$). Therefore, we need to normalize the weights first. To that end, let I denote the N-dimensional unit vector $(1, 1, \ldots, 1)$, and let (X, Y) denote the dot product of two vectors X and Y. We computed the normalization constant $\lambda_k = (C_k, I)^{-1}$. Then, the normalized concentrated sub-portfolio is the vector $\lambda_k C_k$, then it is easy to show that its weights now now add to 1. It is easy to show that the normalization constants $\lambda_k \geq 1$. By construction, for each security S_i in the concentrated sub-portfolio C_k its weight $w_{ik} > v_i$. we have since $\lambda_k > 1$ and the weights in the original

Our ensemble portfolio P^* is constructed by averaging the normalized concentrated sub-portfolios:

$$P^* = \frac{1}{K}(\lambda_1 C_1 + \cdots + \lambda_K C_K)$$

By contrast, a $1/K$ portfolio Q^* is constructed by averaging the original portfolios:

$$Q^* = \frac{1}{K}(P_1 + \cdots + P_K)$$

It is easy to show that the normalization constants $\lambda_k \geq 1$. By construction, for each security S_i in the concentrated sub-portfolio C_k its weight $w_{ik} > v_i$. Therefore, for such security in the ensemble portfolio its weight $\lambda_k w_{ik} \geq w_{ik} > v_i$.

Since for each portfolio $P_k = C_k + U_k$, we can rewrite the $1/K$ portfolio Q^* as follows:

$$\begin{aligned} Q^* &= \frac{1}{K}(P_1 + \cdots + P_K) \\ &= \frac{1}{K}((C_1 + U_1) + \cdots + (C_K + U_K)) \\ &\frac{1}{K}(C_1 + \cdots + C_K) + \frac{1}{K}(U_1 + \cdots + U_K) \\ &= C^* + U^* \end{aligned} \tag{6}$$

Therefore, the $1/K$ portfolio can be represented as a sum of two disjoint portfolios: the average C^* of K concentrated sub-portfolios $\{C_1, \ldots, C_K\}$ and the average U^* of K non-concentrated sub-portfolios

$\{U_1, \ldots, U_K\}$. On the other hand, by construction, the ensemble portfolio P^* will contain the same stocks as the C^* but with higher weights since $\lambda_k > 1$ for all k.

Let us illustrate the construction of such portfolios using the examples presented in the previous section.

Example: Let us consider the same portfolios two portfolios P_1 and P_2 from our previous example. Recall that these were defined as follows:

$$\begin{aligned} M^* &= (0.25, 0.25, 0.25, 0.25, 0) \\ P_1 &= (0.3, 0.3, 0, 0, 0.4) \\ P_2 &= (0.5, 0.3, 0.1, 0.1, 0) \end{aligned} \tag{7}$$

We have $P_1 = C_1 + U_1$ and $P_2 = C_2 + U_2$ where

$$\begin{aligned} C_1 &= (0.3, 0.3, 0, 0, 0) \quad \text{and} \quad U_1 = (0, 0, 0, 0, 0.4) \\ C_2 &= (0.5, 0.3, 0, 0, 0) \quad \text{and} \quad U_2 = (0, 0, 0.1, 0.1, 0) \end{aligned} \tag{8}$$

To construct the ensemble portfolio, we need to re-normalize the weights in the concentrated sub-portfolios C_1 and C_2.

For C_1 we compute $\lambda_1 = (C_1, I)^{-1} = (0.3 + 0.3)^{-1} = 1\frac{2}{3}$. The first normalized concentrated sub-portfolio is then

$$\lambda_1 C_1 = (0.5, 0.5, 0, 0, 0, 0)$$

For C_2 we compute $\lambda_2 = (C_2, I)^{-1} = (0.5 + 0.3)^{-1} = 1.25$. The second normalized concentrated sub-portfolio is then

$$\lambda_2 C_2 = (0.625, 0.375, 0, 0, 0, 0)$$

The ensemble portfolio P^* is constructed by averaging the normalized concentrated sub-portfolios $\lambda_1 C_1$ and $\lambda_2 C_2$:

$$P^* = \frac{1}{2}(\lambda_1 C_1 + \lambda_2 C_2) = (0.5625, 0.4375, 0, 0, 0)$$

By contrast, the $1/K$ portfolio Q^* is computed by averaging the weights in the original portfolios:

$$Q^* = \frac{1}{2}(P_1 + P_2) = (0.4, 0.3, 0.05, 0.05, 0.2)$$

We can also write the $1/K$ portfolio $Q^* = C^* + U^*$ where

$$\begin{aligned} C^* &= \frac{1}{2}(C_1 + C_2) = (0.4, 0.3, 0, 0, 0) \\ U^* &= \frac{1}{2}(U_1 + U_2) = (0, 0, 0.05, 0.05, 0.2) \end{aligned} \tag{9}$$

In the ensemble portfolio, by construction the (positive) weight of each security is greater than the corresponding weight of this security in the index (securities S_1 and S_2 in the above example). The ensemble portfolio will not contain any stocks that are not in the market index. For the $1/K$ portfolio this is not the case. If we look at our example above, the weights of the first two securities S_1 and S_2 are greater than their weights in the market portfolio but the weights of the other two securities S_3 and S_4 are lower than the corresponding weights in the market index. In addition, note that the $1/K$ portfolio contains security S_5 with weight 0.2 that is not in the index.

3.2 Risk/Return Analysis of Ensemble Portfolios and $1/K$ Portfolios

We are now ready to establish some general results on the performance of the ensemble portfolio P^*. To that end, we need some additional notation. Let $R()$ denote the return. In particular, $R(S_i)$ denotes the return of security S_i, whereas for any portfolio P let $R(P)$ denote the return of portfolio specified by the vector P. Similarly, $\sigma(P)$ denote the volatility of the returns in the portfolio specified by the vector of weights P. Let M^* denote the portfolio obtained by taking the same stocks as in P^* but with weights v_i in the SP-500 index. We will refer to M^* as market portfolio.

Recall, that after normalization all weights in the normalized concentrated sub-portfolios $\lambda_k C_k$ add to 1, and therefore the weights in the ensemble portfolio add to 1 as well

$$(P^*, I) = \frac{1}{K}(\sum_k \lambda_k C_k, I) = \frac{1}{K}\sum_k (\lambda_k C_k, I) = \frac{1}{K}\sum_k 1 = 1$$

Let w_{ki}^* denote the normalized weight of security S_i in the concentrated sub-portfolio C_k. And let w_i^* denote the weight of the security S_i in the ensemble portfolio.

We are now ready to analyze the properties of the ensemble portfolio P^*. Our first result is the following:

$$R(P^*) > R(M^*)$$

Proof: We assume that by concentrating a portfolio P_k in some stocks, a manager takes a bet that the concentrated portion (after normalization) will outperform the market. In other words, $R(\lambda_k C_k) >$

$R(M^*)$. If this is the case for all managers, then for the ensemble portfolio

$$R(P^*) = R(\frac{1}{K}\sum_k \lambda_k C_k) = \frac{1}{K}\sum_k R(\lambda_k C_k) > R(M^*) \qquad (10)$$

Our next results relate to the risk (volatility) $\sigma(P^*)$ of the ensemble portfolio P^* - we use standard deviation of returns as our risk measure. We obtain two different bounds for the risk of our ensemble portfolio:

The first result gives a upper bound in terms of volatility of underlying securities. Specifically, the volatility (risk) of ensemble portfolio is bounded by a weighted average of volatilities (risks) of the underlying securities where the weights are those used in the construction of P^*.

The second result gives an upper bound in terms of volatility of underlying sub-portfolios. Specifically, the volatility of ensemble portfolio is no more than the average of volatilities $\sigma(P_1^*), \ldots, P_K^*)$ of underlying sub-portfolios.

To proceed, we need some additional notation. Let $cov(S_i, S_j)$ denote the covariance of return distributions for stocks S_i and S_j and let $\sigma(S_i)$ denote the standard deviation of returns for S_i. Then, using the Cauchy-Schwartz inequality

$$|\text{cov}(S_i, S_j)| \leq \sigma(S_i)\sigma(S_j)$$

we obtain the following for the variance of returns for the ensemble portfolio P^*

$$\begin{aligned} \sigma^2(P^*) &= \sum_{i,j=1}^{N} w_i^* w_j^* \text{cov}(S_i, S_j) \leq \sum_{i,j=1}^{N} w_i^* w_j^* \sigma(S_i)\sigma(S_j) \\ &= \left[\sum_{i=1}^{N} w_i^* \sigma(S_i)\right] \left[\sum_{j=1}^{N} w_j^* \sigma(S_j)\right] = \left[\sum_{i=1}^{N} w_i^* \sigma(S_i)\right]^2 \end{aligned} \qquad (11)$$

And, therefore,

$$\sigma(P^*) \leq \sum_{i=1}^{N} w_i^* \sigma(S_i) \tag{12}$$

This means that the risk (volatility) of the ensemble portfolio P^* is less than the weighted average risk (volatility) of individual stocks where the weights correspond to the weights of securities in the ensemble portfolio.

Let us now relate the risk of the ensemble portfolio to the risk of underlying concentrated sub-portfolios. By construction, we can write $P^* = (\lambda_1 C_1 + \cdots + \lambda_K C_K^*)/K$. Therefore, for the risk of the ensemble portfolio we obtain:

$$\begin{aligned}
\sigma^2(P^*) &= \sum_{k,m=1}^{K} \frac{1}{K^2} \text{cov}(\lambda_k C_k, \lambda_m C_m) \leq \frac{1}{K^2} \sum_{k,m=1}^{K} \sigma(\lambda_k C_k) \sigma(\lambda_m C_m \\
&= \frac{1}{K^2} \left[\sum_{k=1}^{K} \sigma(\lambda_k C_k) \right] \left[\sum_{m=1}^{K} \sigma(\lambda_m C_m) \right] \\
&= \left[\frac{\sigma(\lambda_1 C_1) + \cdots + \sigma(\lambda_K C_K)}{K} \right]^2
\end{aligned} \tag{13}$$

Therefore,

$$\sigma(P^*) \leq \frac{\sigma(\lambda_1 C_1) + \cdots + \sigma(\lambda_K C_K)}{K}$$

This shows that the risk of the ensemble portfolio P^* is not worse than the average risk across (normalized) concentrated sub-portfolios.

4 Analysis of Sharpe's Ratio for the Ensemble Portfolio

In this section, we will focus on the Sharpe's ratio of the ensemble portfolio. Formally, the Sharpe's ratio is defined as the average return in excess of the risk-free rate per unit of volatility. To simplify the computation, we will assume that the risk-free rate is 0. With this assumption, the Sharpe's ratio of the ensemble portfolio is

$$\text{Sharpe}(P^*) = \frac{R(P^*)}{\sigma(P^*)}$$

We will show that this Sharpe's ratio is greater than the Sharpe's ratio of the SP-500 index. We will show this by arguing that for each sub-portfolio C_k of concentrated stocks we have

$$\text{S}harpe(C_k) \geq \text{Sharpe}(M^*)$$

Consider the concentrated sub-portfolio C_k of manager k. Recall that in this portfolio, we have N securities $S_1, \ldots, S_N$ with weights w_{ik} for each security i satisfying $w_{ik} > v_i$ where v_i is the weight of S_i in the SP-500 index. In choosing these stocks, the manager hopes to outperform the market without taking too much additional risk. In particular, we will assume that this concentrated sub-portfolio of selected stocks has a Sharpe's ratio greater than the market. With this assumption, our task is to examine the Sharpe's ratio of the ensemble portfolio.

We argue as follows. When we construct the ensemble portfolio, we take the concentrated sub-portfolios $C_1, \ldots, C_K$ and normalize the weights in each sub-portfolio C_k by the normalization constant λ_k so that the weights so that they add to 1. Mathematically, the normalized concentrated sub-portfolio is $\lambda_k C_k$. Our first result is the following:

$$\text{S}harpe(C_k) = \text{Sharpe}(\lambda_k C_k)$$

Proof: The return of the "normalized" sub-portfolio $\lambda_k C_k$ is

$$\begin{aligned} R(\lambda_k C_k) &= \sum_i w^*_{ik} R(S_i) = \sum_i \lambda_k w_{ik} R(S_i) \\ &= \lambda_k \sum_i w_{ik} R(S_i) = \lambda_k R(C_k) \end{aligned} \tag{14}$$

On the other hand, for the variance of the normalized portfolio

$$\begin{aligned} \sigma^2(\lambda_k C_k) &= \sum_{i,j} [\lambda_k w_{ik}][\lambda_k w_{jk}] \mathrm{cov}(R_i, R_j) \\ &= \lambda_k^2 \sum_{i,j} w_{ik} w_{jk} \mathrm{cov}(R_i, R_j) = \lambda_k^2 \sigma^2(C_k) \end{aligned} \tag{15}$$

From this we immediately obtain

$$\mathrm{S}harpe(\lambda_k C^*_k) = \frac{R(\lambda_k C_k)}{\sigma(\lambda_k C_k)} = \frac{\lambda_k R(C_k)}{\lambda_k \sigma(C_k)} = \mathrm{Sharpe}(C_k)$$

We made an assumption that when choosing the concentrated sub-portfolio C_k of stocks with higher weights than those in SP-500, the manager picks such stocks so that the Sharpe's ratio of any sub-portfolio is superior to that of the market, i.e. for all k:

$$\mathrm{S}harpe(C_k) \geq \mathrm{Sharpe}(M^*)$$

With this assumption, the above result implies that for the normalized portfolio $\lambda_k C_k$ we have

$$\mathrm{S}harpe(\lambda_k C_k) \geq \mathrm{Sharpe}(M^*)$$

We are now ready to derive some bounds for the Sharpe's ratio of the ensemble portfolio P^*. We will use the following inequality:

$$\min_i \left(\frac{1}{a_i}\right) \leq \frac{K}{a_1 + \cdots + a_K}$$

Let $\lambda_h C_h$ be the normalized concentrated sub-portfolio with the highest volatility. Then, for the Sharpe's Ratio of the ensemble portfolio we have

$$\begin{aligned} Sharpe(P^*) &= \frac{R(P^*)}{\sigma(P^*)} \geq \frac{K \cdot R(P^*)}{\sigma(\lambda_1 C_1) + \cdots + \sigma(\lambda_k C_K)} \\ &\geq \frac{K \cdot R(P^*)}{\sigma(\lambda_h C_h)} \geq \frac{R(\lambda_h C_h)}{\sigma(\lambda_h C_h)} \\ &= \text{Sharpe}(\lambda_h C_h) \\ &= \text{Sharpe}(C_h) \end{aligned} \tag{16}$$

The Sharpe's ratio of the ensemble portfolio is greater than the Sharpe's ratio for the most volatile concentrated sub-portfolio. Since each concentrated component C_k for any portfolio is assumed to have the Sharpe's ratio greater than the market, it follows that the Sharpe's ratio of the ensemble portfolio is higher than that of the market:

$$Sharpe(P^*) \geq Sharpe(M^*)$$

Let us now summarize our main results for the ensemble portfolio P^*:

(1) the return of the ensemble portfolio is higher than the return of the corresponding stocks in the SP-500 index

(2) the volatility of the ensemble portfolio is less than the volatility of the corresponding sub-portfolios

(3) The Sharpe's ratio of the ensemble portfolio is higher than that of the market

5 A Detailed Numerical Example

We consider 10 large-cap mutual funds {BCSIX, CAFCX, CFNBX, CWMBX, DODGX, FMIHX, HCAIX, MMDEX, RYSEX,VPCCX}. From these ten mutual funds, we constracted two portfolios: the ensemble portfolio P^* and the $1/K$ portfolio Q^*. In addition, we have the SP-500 market portfolio M^*.

First, we computed rolling annual returns for the three portfolios: P^*, Q^* and M^* summarized below

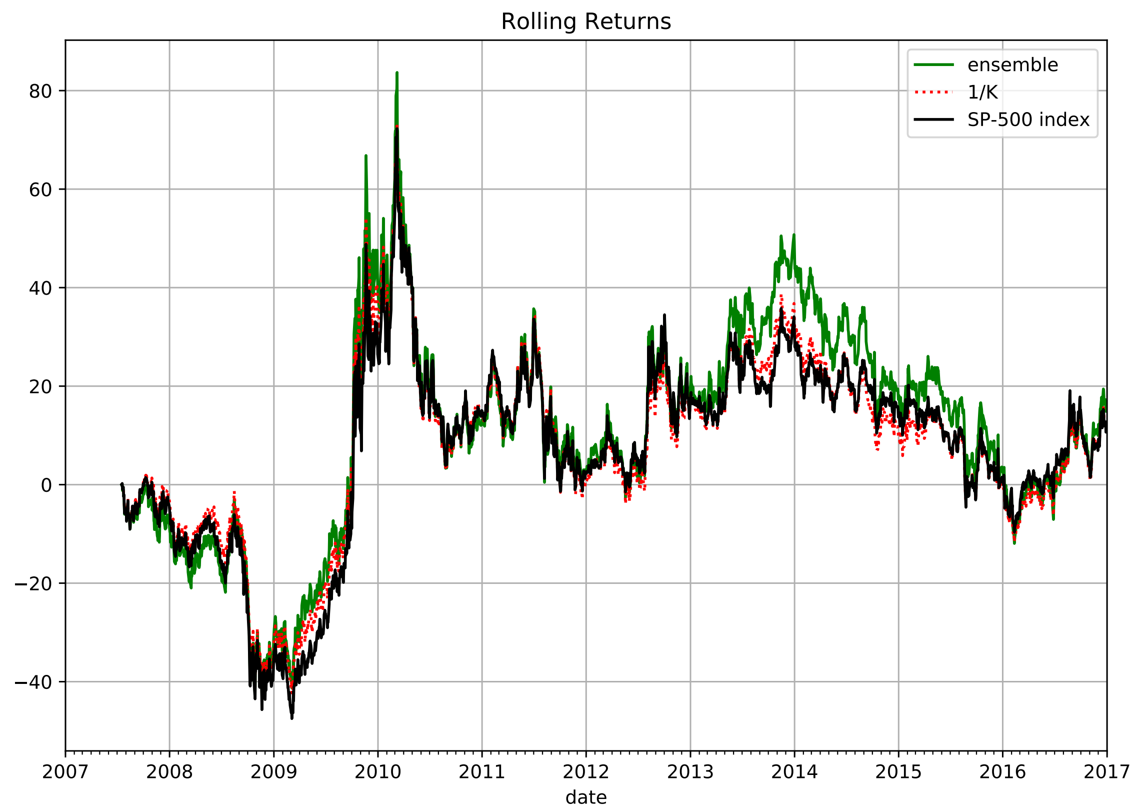

Figure 4: Rolling Annual returns

Next, let us examine the tracking error for the rolling returns. We use Δ to indicate tracking errors. Tracking errors for both $1/K$ portofolio Q^* and ensemble portfolio P^* are computed relative to the market portfolio. For example, tracking error for returns of the ensemble portfolio is $\Delta R(P^*) = R(P^*) - R(M^*)$.

Table 3: Annual Returns for Portfolios

year	SP-500	$1/K$	ensemble
2008	-36.09	-32.98	-31.59
2009	26.46	33.62	39.66
2010	15.06	14.58	13.17
2011	2.11	0.25	2.42
2012	16.75	16.82	19.42
2013	32.39	35.15	49.23
2014	13.69	9.90	17.28
2015	1.38	0.55	3.18
2016	11.96	12.47	15.73

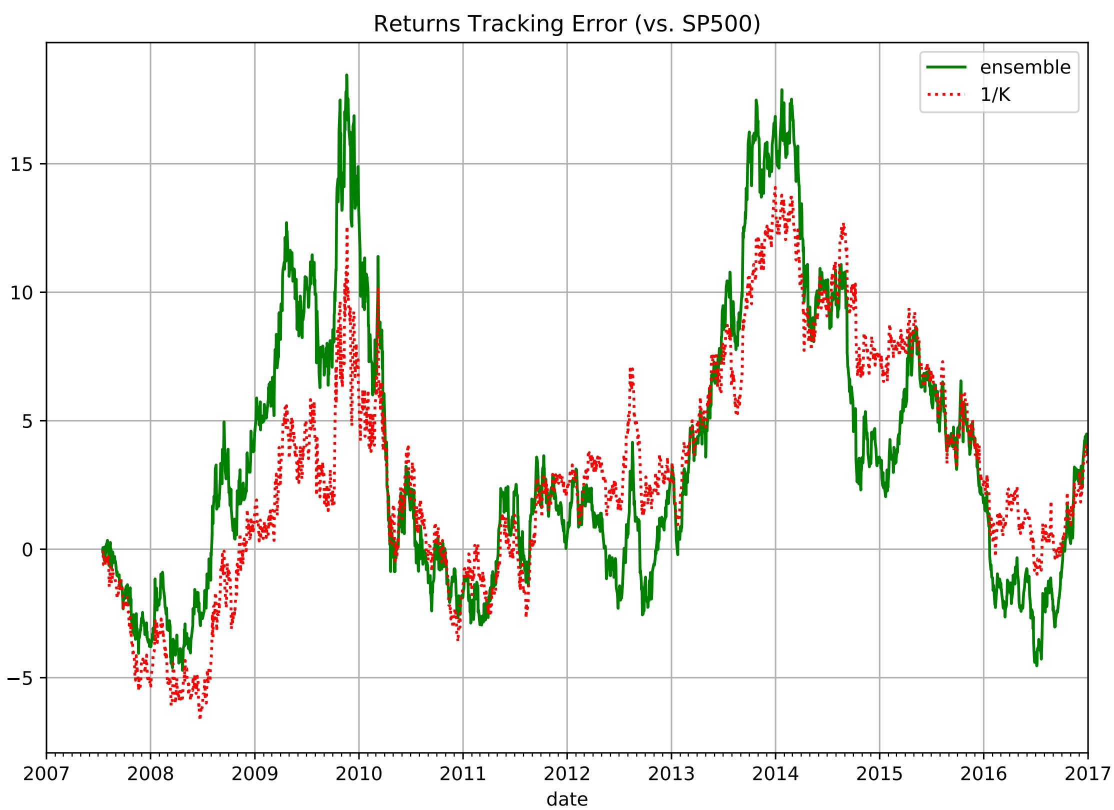

Figure 5: Rolling Annual returns

Let us summarize the annual tracking errors in the table below:

Table 4: Tracking Error for Annual Returns

year	SP-500	$1/K$	ensemble
2008	-36.09	3.11	4.50
2009	26.46	7.16	13.19
2010	15.06	-0.48	-1.89
2011	2.11	-1.86	0.31
2012	16.75	0.07	2.67
2013	32.39	2.76	16.85
2014	13.69	-3.79	3.59
2015	1.38	-0.83	1.80
2016	11.96	0.51	3.77
average	9.30	0.74	4.98

The ensemble portfolio outperform the market in 8 years out of 9 except for 2010. In that year, the ensemble portfolio under-performed the market by about 150 basis points. The average tracking error for ensemble portfolio is about 500 basis points. In some years, this tracking error is quite large: for example in 2009 and 2013, the ensemble portfolio over-performed the market by more than 1,000 basis points. By contrast, the $1/K$ portfolio has a tracking error of about 75 basis points and most of the time follows the market index quite closely.

Let us now examine the rolling volatility

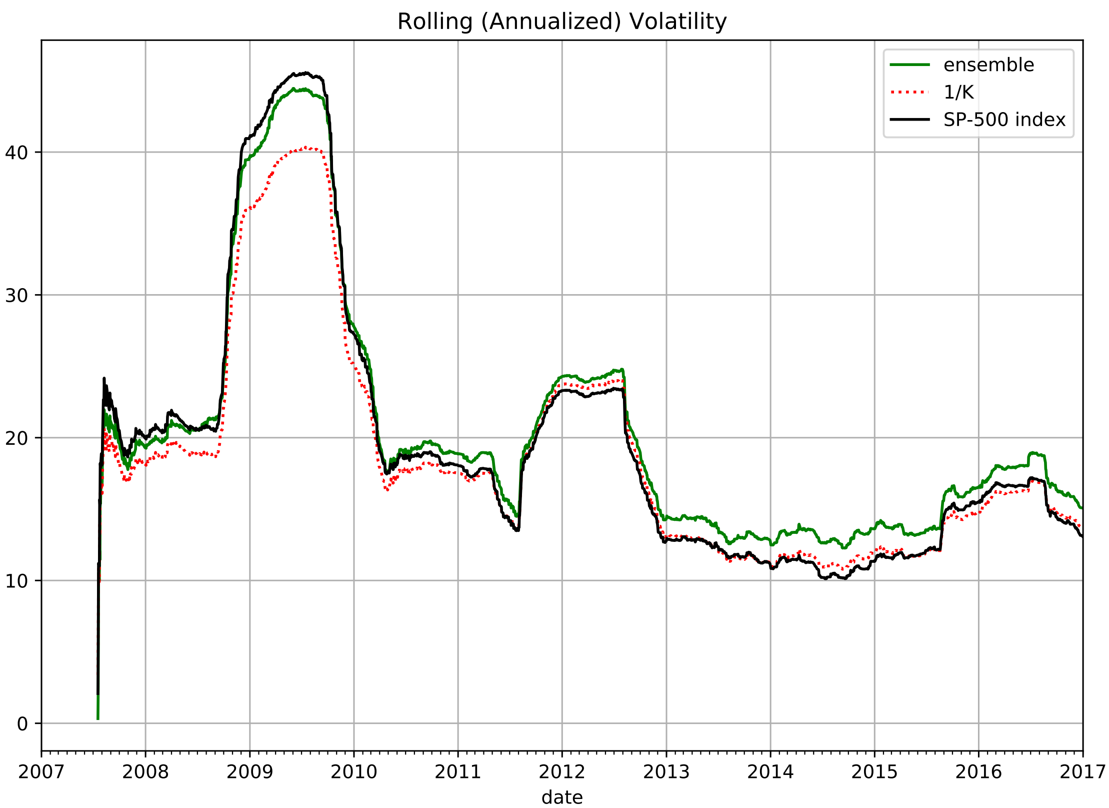

Figure 6: Rolling Annualized Volstility

Let us summarize the annualized portfolio volatility in the table below:

The annualized volatility is practically identical to that of the market. On the other hand, the annualized volatility of the ensemble portfolio is about 10% higher than the market. In fact, if we examine the tracking error for the annualized volatility, we do not see a significant increase in volatility as compared to market. To see the increase in volatility by the ensemble portfolio, consider the tracking error for the volatility Let us now examine the tracking error of this volatility

Table 5: Rolling Annualized Volatility

year	SP-500	$1/K$	ensemble
2008	41.03	36.01	39.60
2009	27.27	25.10	27.82
2010	18.06	17.50	18.87
2011	23.29	23.73	24.27
2012	12.75	13.08	14.28
2013	11.07	11.15	12.75
2014	11.38	11.94	13.65
2015	15.49	14.78	16.54
2016	13.10	13.69	15.07

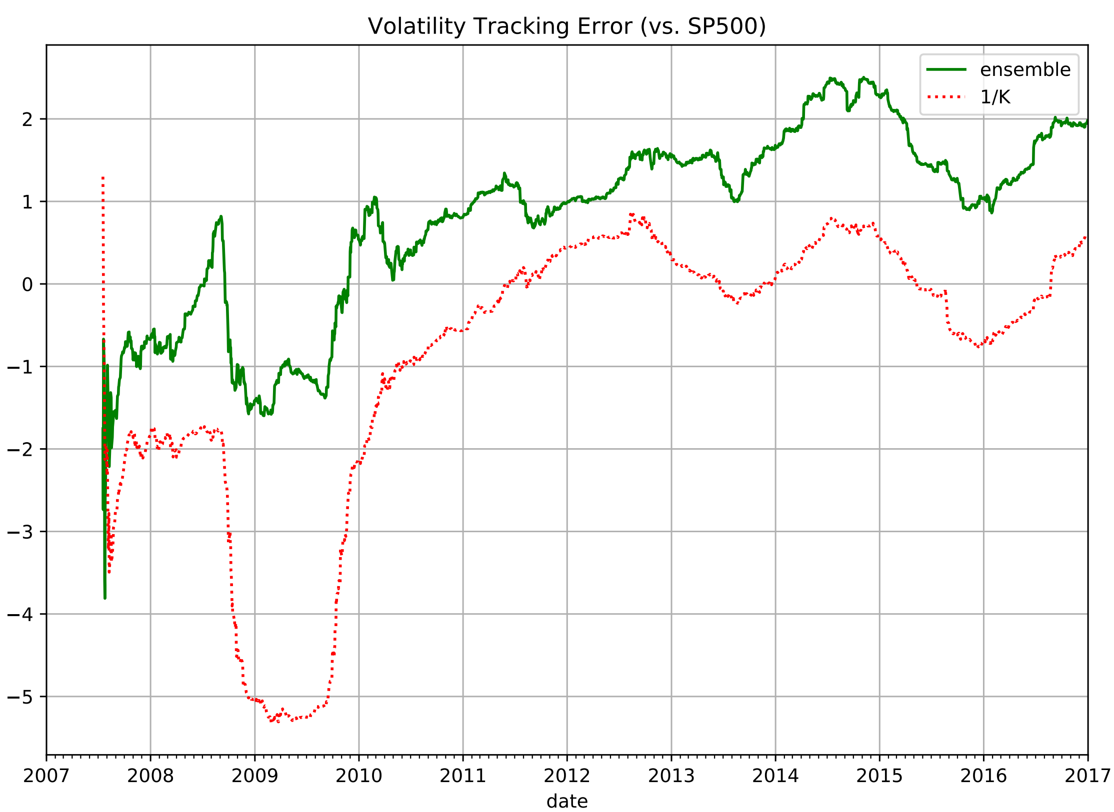

Figure 7: Volatility Tracking Error

Let us illustrate the above point by examining tracking errors in annual return and annualized volatility for both the $1/K$ and ensemble portfolios relative to that of the market. Our results are summarized below:

Table 6: Tracking Errors for Annual Returns and Volatility

Date	Returns			Volatility		
	SP-500	$1/K$	ensemble	SP-500	$1/K$	ensemble
2008	-36.09	3.11	4.50	41.03	-5.02	-1.43
2009	26.46	7.16	13.19	27.27	-2.18	0.55
2010	15.06	-0.48	-1.89	18.06	-0.56	0.80
2011	2.11	-1.86	0.31	23.29	0.44	0.98
2012	16.75	0.07	2.67	12.75	0.33	1.53
2013	32.39	2.76	16.85	11.07	0.07	1.68
2014	13.69	-3.79	3.59	11.38	0.56	2.27
2015	1.38	-0.83	1.80	15.49	-0.71	1.04
2016	11.96	0.51	3.77	13.10	0.60	1.98
average	9.30	0.74	4.98	19.27	-0.72	1.05

What we observe from the above is the significant increase in annualized returns and a modest increase in volatility. Therefore, we would expect the Sharpe's ratio of the ensemble portfolio to be higher than that of the market. In the figure below, we plot the rolling Sharpe's ratio for the three portfolios:

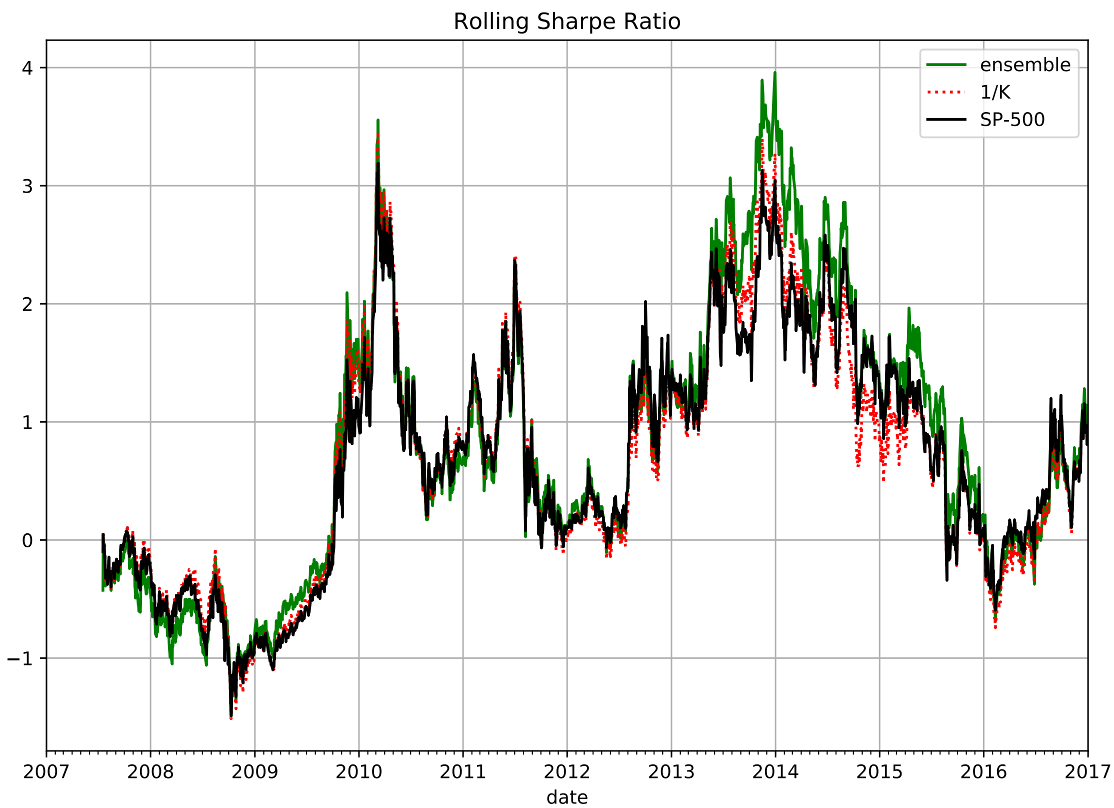

Figure 8: Rolling Annual returns

Let us summarize annual Sharpe's ratio for our portfolios in the table below:

We can see that in every year except for 2010, we get better Sharpe's ratio for the ensemble than for the $1/K$ portfolio.

To see the increase in the Sharpe ratio, let us examine the difference between ensemble and $1/K$ portfolio Sharpe's ratio with that of the market Turning now to the Sharpe's Ratio

Table 7: Rolling Annual Sharpe's Ratios

year	SP-500	$1/K$	ensemble
2008	-0.88	-0.92	-0.80
2009	0.97	1.34	1.43
2010	0.83	0.83	0.70
2011	0.09	0.01	0.10
2012	1.31	1.29	1.36
2013	2.92	3.15	3.86
2014	1.20	0.83	1.27
2015	0.09	0.04	0.19
2016	0.91	0.91	1.04
average	0.83	0.83	1.02

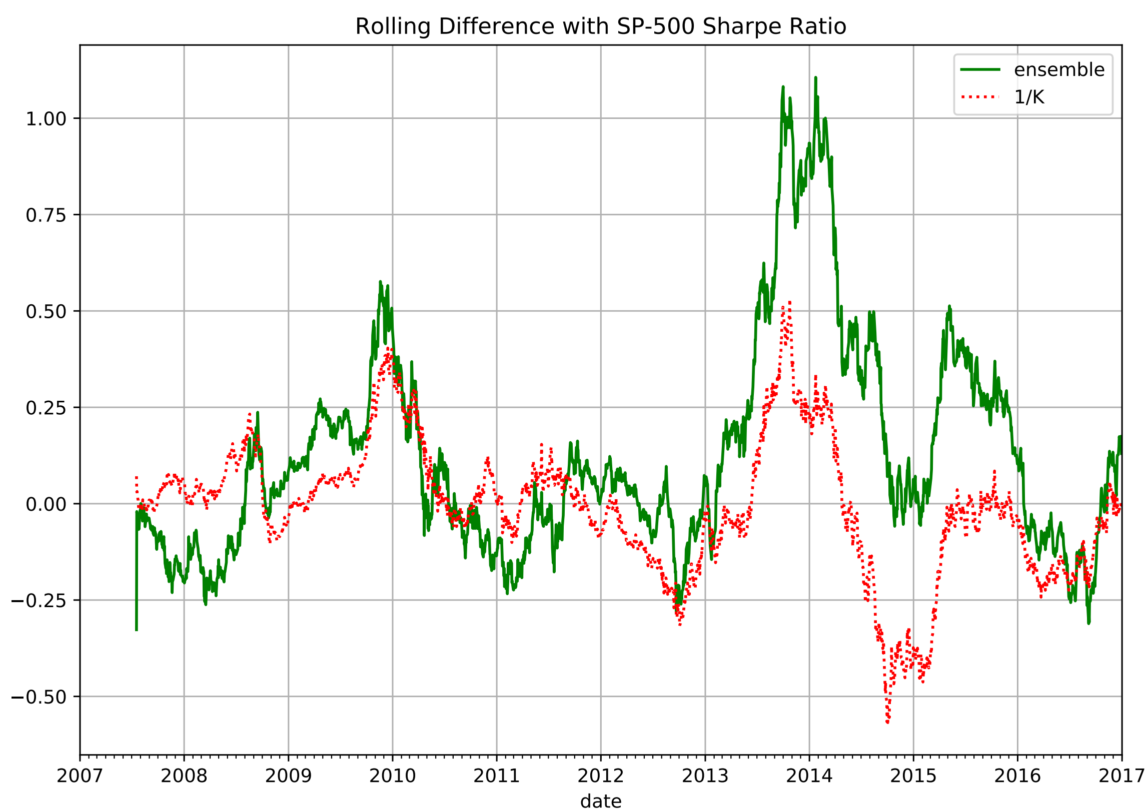

Figure 9: Rolling Sharpe Ratio Tracking Error

6 Numerical results

In the previous section we considered the comparison of two specific portfolios: ensemble portfolio and $1/K$ portfolio generated from a particular set of 10 large-cap mutual funds. In this section we provide some statistical results on a sample of such portfolios.

In particular, we consider a universe of 21 large-cap mutual funds (a list is presented in the Appendix). We generated 100 random portfolios with each portfolio consisting of 10 funds. From each portfolio of 10 mutual funds, we computed an ensemble portfolio and a $1/K$ portfolio. For these portfolios, we generated a number of histograms for annual returns, tracking volatility and Sharpe ratios.

We start with examining the distribution of annual returns for these 100 $1/K$ portfolios and corresponding ensemble portfolios for each year from 2008 till 2016.

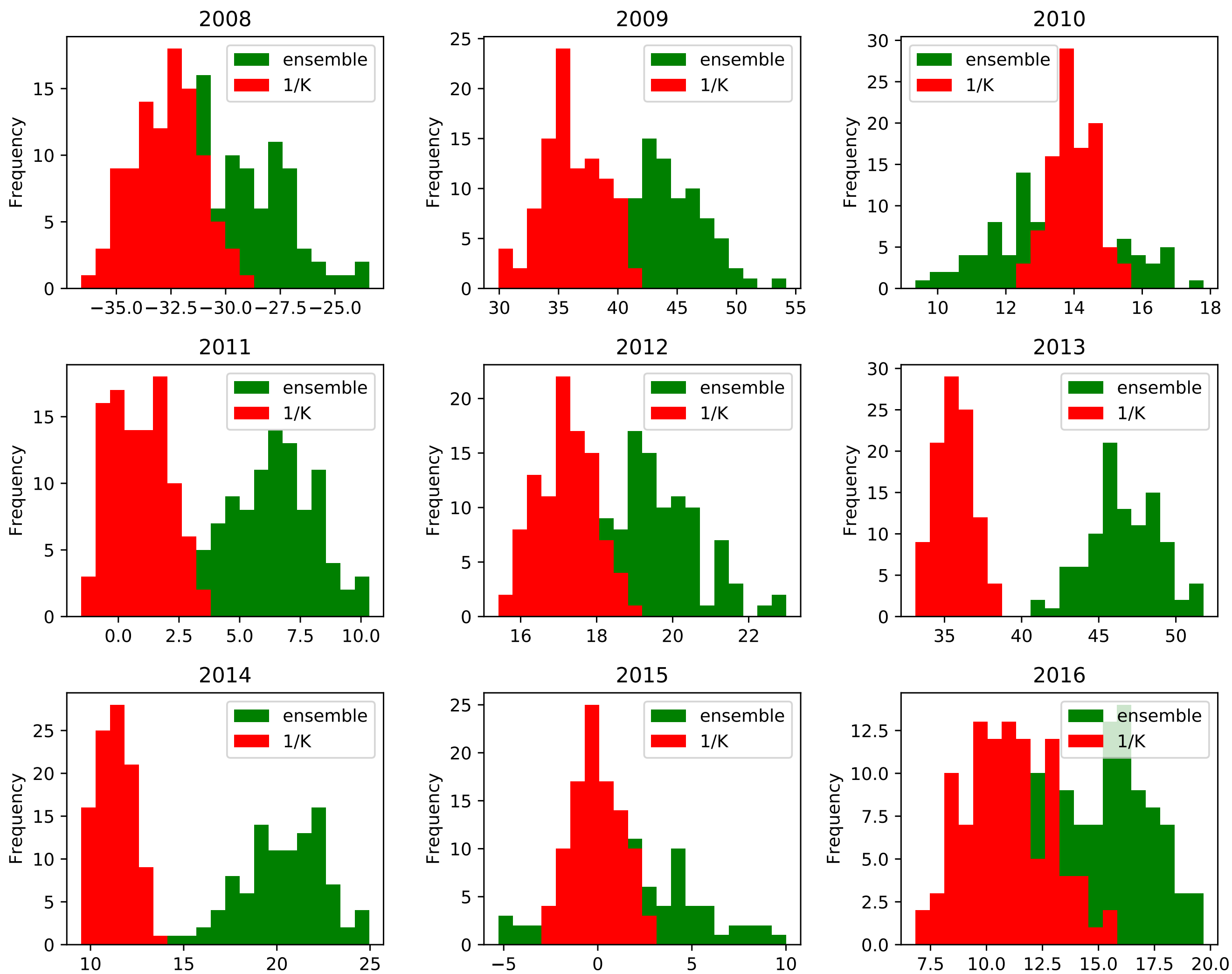

Figure 10: Distribution of Returns

We see a significant shift in performance in favor of the ensemble portfolios. For example, in 2013, 2014 and 2016 the difference in returns is about 500 to 1,000 basis points for a typical ensemble portfolio in our sample. To compare these returns to that of the market, we consider the distribution of tracking error for the returns. This is shown in the figure below:

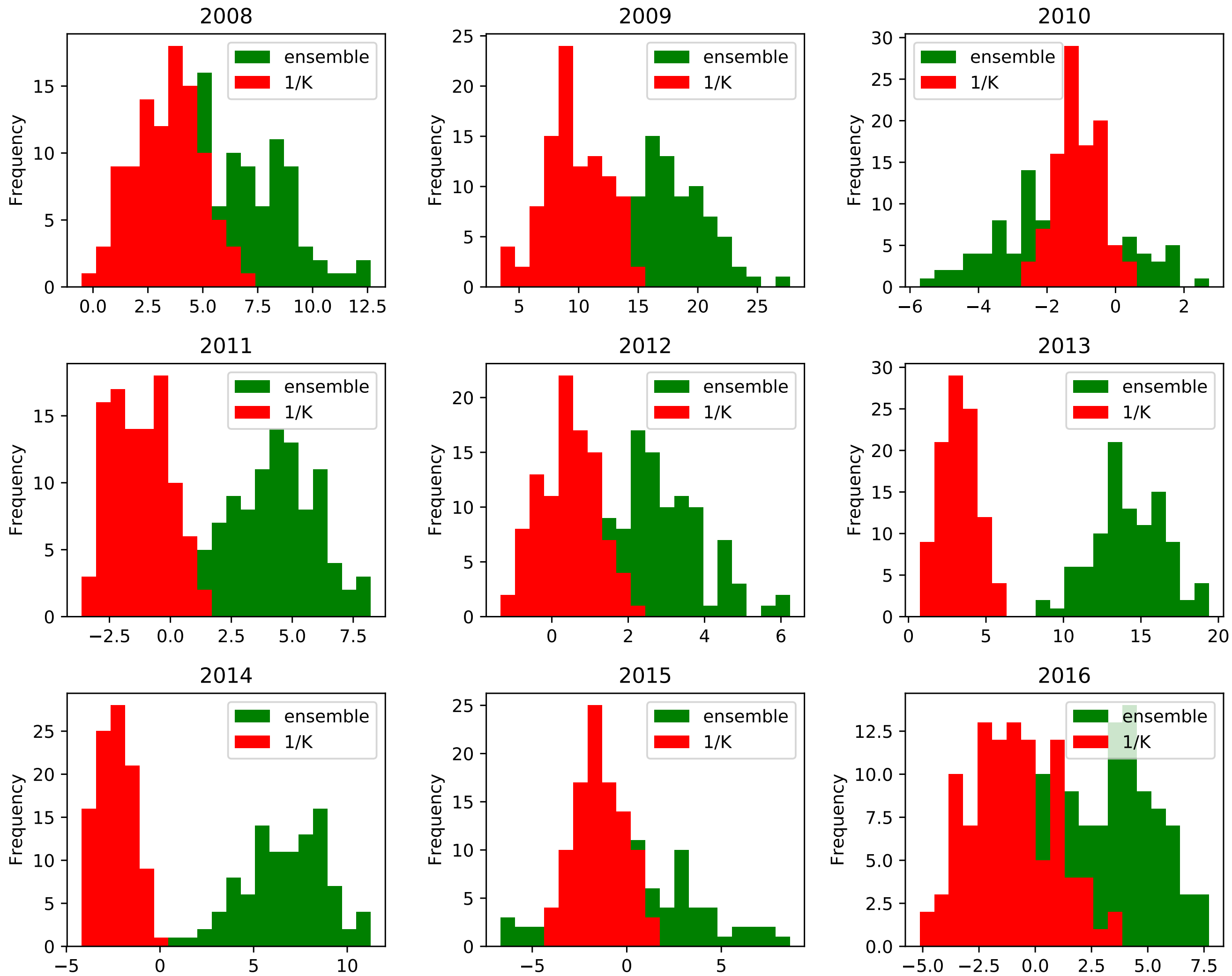

Figure 11: Distribution of Return Tracking Errors

Except possibly for 2010, the ensemble portfolios seem to outperform the market and the $1/K$ portfolios, sometimes quite significantly. On the other hand, the ensemble portfolios seem to under-perform the market. In fact, the $1/K$ portfolios in our sample under-performed the market in

On the other hand, the increase in annual returns comes at a price of increased volatility. Let us compare the distributions of tracking errors for volatility for $1/K$ and ensemble portfolios.

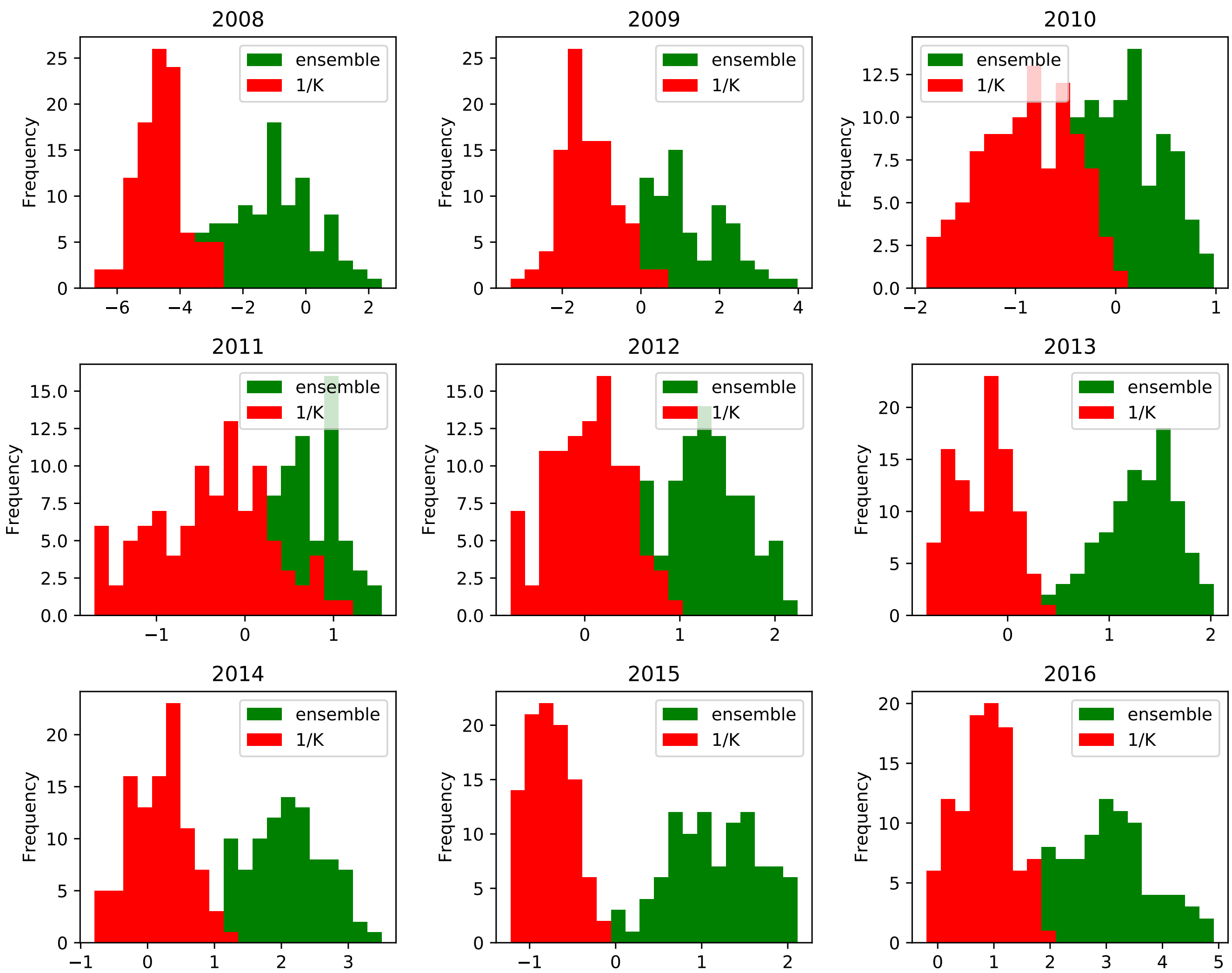

Figure 12: Distribution of Tracking Error for Volatility

Next, we compare the distribution of tracking error for Sharpe ratio. First, we note that the volatility of $1/K$ portfolios are comparable to that of the market or slightly lower. If funds were independent and exhibit low correlation with each other, we would expect a significant decrease in volatility when averaging such funds to create $1/K$ portfolios. The reason that we do not see such a drastic reduction in volatility for $1/K$ portfolios is high degree of correlation between such portfolios and between these portfolios and the market. On the other hand, the volatility of ensemble portfolios is higher than that of the market as managers concentrate in higher return stocks in hopes to beat the market index. What we wee from these histograms is that the additional increase in volatility over the $1/K$ portfolios is about 2%.

What does this mean? It means that with ensemble portfolios we are paying a price of a modest increase in risk but gain much more in returns. We can see this point more clearly if we examine the distribution of tracking error for the Sharpe's ratio.

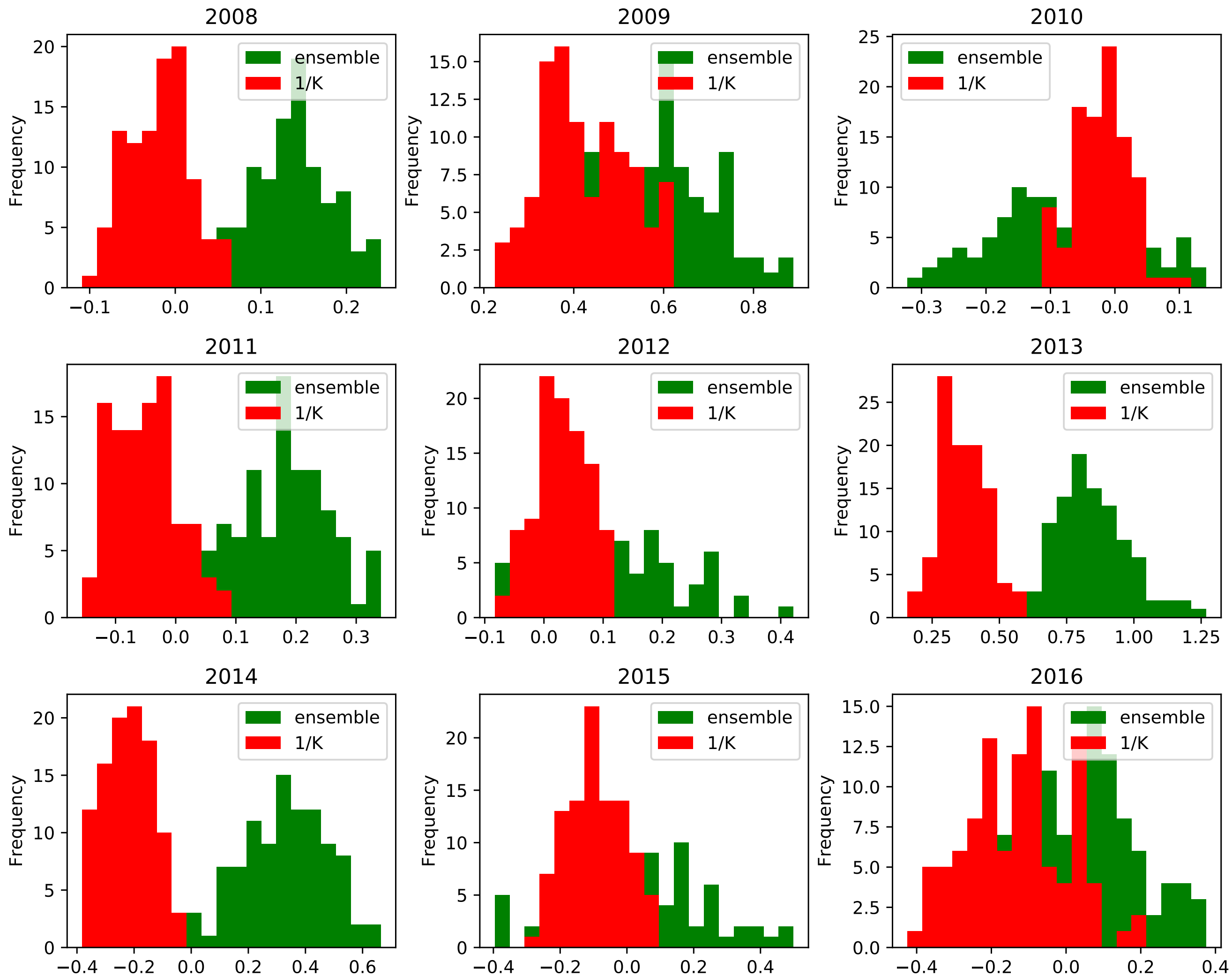

Figure 13: Distribution of Tracking Error for Sharpe Ratios

From the above histograms, we see that the Sharpe ratios for ensemble portfolios dominate those for the $1/K$ portfolios.

7 Conclusion

In this work we have provided a mathematical foundation for ensemble machine learning. We have shown that if error probabilities for each manager is less than 0.5 and manager decisions are independent, then using a majority voting to make a decision results in lower error probability. This error probability can be made smaller by adding enough managers to the ensemble. The resulting ensemble portfolio has higher return than the corresponding stocks with weights from S&P-500. At the same time, the risk of this ensemble portfolio does not exceed the average of the risks of individual sub-portfolios.

8 Appendix: Large-Cap Funds Used in Study

References

[1] R. Clemen, Combining forecasts: A Review and Annotated Bibliography, 1989, International Journal of Forecasting 5, 559-581.

[2] V. DeMiguel, L. Garlappi, R. Uppal, Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy, 2009, The Review of Financial Studies, 22, 1915-1953.

[3] W. Feller, Introduction to Probability Theory and Its Applications, 1950, J.Wiley and Sons.

[4] W. Hoeffding, Probability Inequality for Sums of Bounded Random Variables, 1963, Journal of the American Statistical Association, 56 (301), 13-30

Table 8: A Large-Cap Funds

Ticker	Fund Name
BCSIX	Brown Capital Mgmt Small Co Inv
CAFCX	American Funds AMCAP 529C
CFNBX	American Funds Fundamental Invs 529B
CNGBX	American Funds New Economy 529B
CWMBX	American Funds Washington Mutual 529B
DODGX	Dodge & Cox Stock
FMIHX	FMI Large Cap
HCAIX	Harbor Capital Appreciation Inv
MMDEX	Praxis Growth Index I
MVIIX	Praxis Value Index
OARLX	Oakmark Select II
OARMX	Oakmark II
POGRX	PRIMECAP Odyssey Growth
POSKX	PRIMECAP Odyssey Stock
RYSEX	Royce Special Equity Invmt
SEQUX	Sequoia
VDIGX	Vanguard Dividend Growth
VHCOX	Vanguard Capital Opportunity Inv
VPCCX	Vanguard PRIMECAP Core Inv
VPMCX	Vanguard PRIMECAP Inv
YACKX	AMG Yacktman Service

[5] H. Markowitz, Portfolio Selection. 1952, Journal of Finance 7, 7791

[6] Pelekis, C. Ramon, J. Mediterr., Hoeffdings Inequality for Sums of Dependent Random Variables, J. Math. (2017).

[7] P. Schaubacker, Combining Portfolio Models, 2014, Annals of Economics and Finance, 1502, 433-455.