ANNALS OF ECONOMICS AND FINANCE **15-2**, 433–455 (2014)

Combining Portfolio Models*

Peter Schanbacher

Department of Economics, University of Konstanz, Universitätsstraße 10, D-78464 Konstanz, Germany
E-mail: peter.schanbacher@uni-konstanz.de

The best asset allocation model is searched for. In this paper, we argue that it is unlikely to find an individual model which continuously outperforms its competitors. Rather one should consider a combined model out of a given set of asset allocation models. In a large empirical study using various standard asset allocation models, we find that (i) the best model depends strongly on the chosen data set, (ii) it is difficult to ex-ante select the best model, and (iii) the combination of models performs exceptionally well. Frequently, the combination even outperforms the ex-post best asset allocation model. The promising results are obtained by a simple combination method based on a bootstrap procedure. More advanced combination approaches are likely to achieve even better results.

Key Words: Investment Strategy; Diversification; Markowitz; Portfolio Optimization; Model Averaging; Portfolio Allocation.
JEL Classification Numbers: C52, C53, G11, G17.

1. INTRODUCTION

In many fields of research the combination of models performs well, sometimes even better than all individual models. This empricial finding has been observed for forecasts (Smith and Wallis, 2009), experts recommendations (Genre et al., 2013), estimators (Hansen, 2010), and others (for an excellent review, see Clemen, 1989). Three explanations are provided. Different models can be based on different information sets or different information processing (Bates and Granger, 1969). Combination helps to combine those information sets or information channels, resp. The second argument is that models are differently affected by structural breaks

*We wish to thank the editor and two anonymous referees for their extensive comments which significantly improved an earlier draft of the paper. Further we thank Winfried Pohlmeier and Fabian Krueger for their valuable discussions. All errors are my own.

1529-7373/2014

(Diebold and Pauly, 1987). Some models are fine tuned in calm periods, at the cost of not being robust in turbulent times. The third argument is that the true data generating process is more complex and of a higher dimension than even the most flexible models (Stock and Watson, 2004). The combination of models is robust to the misspecification of individual models.

Most of these arguments come from the forecasting literature. But they are likely to hold for asset allocation as well. Markowitz (1952) introduced a fundamental concept of portfolio optimization. But when it comes to practice, the concept is difficult to implement (Britten-Jones, 1999). The returns' means are in particular difficult to estimate Frost and Savarino (1986). Also the error in the covariance matrix can become large (Chan et al., 1999). Alternative restricted models have been suggested: the Minimum Variance portfolio (Merton, 1980), the short-selling restricted portfolio (Jagannathan and Ma, 2003), and several norm penalized portfolios (for a Lasso restriction, see e.g. Fan et al., 2012). Even the naive portfolio performs surprisingly well (DeMiguel et al., 2009). Which individual models should be selected? This question remains unanswered. Instead selecting one particular portfolio, one could also consider the combination of several portfolios. So far only few attempts to combine asset allocation models have been suggested. Tu and Zhou (2011) combine the tangency strategy and the naive portfolio. Schanbacher (2012) considers the average over several portfolios. Many shrinkage approaches can be decomposed in the combination of two portfolios, e.g. the Ledoit and Wolf (2004) portfolio can be regarded as a combination of the moments of the Minimum Variance portfolio and the naive portfolio. A general recipe to combine several given portfolios has not been given.

We present a general framework which covers a large range of standard asset allocation models. We analyze three combination methods and one selection method: the combination of portfolios, the average of portfolios, the combination of moments and the selection of the previous best model. We use a simple bootstrap method to determine the share each individual model should get in the combination. Finally, we analyze empirically the performance of the combination and selection methods, as well as the performance of the individual models. We find that (i) no individual model outperforms its competitors, (ii) ex-ante selection of the best model appears to be difficult, (iii) the combination of portfolios often outperforms each individual model. Instead trying to improve a single asset allocation model, one should rather try to make the most of the set of available asset allocation models.

The remainder of the paper is structured as follows. Section two introduces a general decision framework. Section three translates the framework to portfolio choice. We show that the framework captures a large range of common asset allocation models. Section four presents several combination and selection methods and discusses their implications. Section five shows the empirical performance of the individual models and the combination methods. Section six summarizes and concludes.

2. GENERAL DECISION PROBLEM

We apply a single-period forecasting and decision problem. The results can be also applied to multiperiod decision making. The decision is based on a vector of state variables x_{T-1} realized over time period $T-1$ to T. The set of information at time T, $\mathcal{F}_T$, can be either based on a rolling window of length h, e.g. $\mathcal{F}_T = \{x_t\}_{t=T-h}^{T}$ or expanding $\mathcal{F}_T = \{x_t\}_{t=1}^{T}$ or with discounted information $\mathcal{F}_T = \left\{\rho^{T-t}x_t\right\}_{t=1}^{T}$ with $0 < \rho < 1$. Based on the information at $\mathcal{F}_{T-1}$ the forecaster estimates a parameter θ_T by some forecasting model M,

$$\hat{\theta}_T = M(\mathcal{F}_{T-1}) \tag{1}$$

Parameter θ_T describes the relevant properties of x_T, e.g. the moments of x_T. Based on the parameter θ_T, a decision d_T for time T is made. We assume a unique time invariant loss function $l(d,\theta) : \mathcal{D} \times \Theta \to \mathbb{R}$. The optimal decision with respect to the estimated parameter $\hat{\theta}_T$ is given by the decision minimizing the loss, e.g.

$$d_T = \arg\min_{d \in \mathcal{D}} l(d, \hat{\theta}_T) \tag{2}$$

The parameter of interest can correspond to the decision $d_T \in \mathcal{D}$ itself. Mostly, this is given if the set of decision variables $\mathcal{D}$ is univariate, e.g. for point forecasts. Also, it can be a distributional parameter of the variable of interest which gives guidance for decision $d_T \in \mathcal{D}$.

The optimization of eq. 2 is very flexible. It incorporates many standard decision problems, as e.g. the expected loss (see e.g. Pesaran and Timmermann, 2005).

Suppose there is not only one model M, but m different models $M_1, \ldots, M_m$. The corresponding parameters of interest are $\hat{\theta}_T^1, \ldots, \hat{\theta}_T^m$ with $\hat{\theta}_T^i = M_i(\mathcal{F}_{T-1})$. The decision the forecaster takes, depends on the applied model, i.e. $d_T^i = \arg\min_{d\in\mathcal{D}} l(d, \hat{\theta}_T^i)$. Generally, different models $M_i \neq M_j$ lead to different input parameters $\hat{\theta}^i \neq \hat{\theta}^j$; which lead to different decisions $d^i \neq d^j$. If several models are available, the question arises which model to take. One can try to select the best model, i.e. $M_{i^\star}$ with $i^\star = \arg\min_i l(d_T^i, \theta_T)$. In the following, the strategy to pick the best individual model is denoted by *Ind*. Unfortunately, the true parameter θ_T is not known and needs to be estimated itself. The (expected) loss $l(d_T^i, \theta_T)$ of the decision i remains unobserved. It is difficult to ex-ante select the ex-post best model. Alternatively, one can combine models $M_1, \ldots, M_m$. We call $\pi = (\pi^1, \ldots, \pi^m)$ the shares, satisfying $\iota'\pi = 1$. The element $\pi^i \geq 0$ is the share of model i in the combination. There are mainly two alternatives on how to combine. One can combine the decisions, i.e. $d_T^{(comb)} = \sum_{i=1}^m \pi^i d_T^i$. This way of combining different models is the most intuitive type. We refer to it as the $Comb$. The second alternative is the combination of the input parameters, i.e. $\hat{\theta}_T^{(mom)} = \sum_{i=1}^m \pi^i \hat{\theta}_T^i$. The decision is then given by $d_T^{(mom)} = \arg\min_{d\in\mathcal{D}} l(d, \hat{\theta}_T^{(mom)})$. In many situations the parameters of interest are the estimated moments. We call this combination approach the moment combination, abbreviated by Mom. The shares π as well as the estimated parameters $\left\{\hat{\theta}_T^i\right\}_{i=1}^m$ are somehow estimated based on $\mathcal{F}_{T-1}$. Then there exists some model satisfying, $\hat{\theta}_T^{(mom)} = M_{(mom)}(\mathcal{F}_{T-1})$. This type of combination can be regarded as a combination of models $M_1, \ldots, M_m$ or as an additional super model $M_{(mom)}$.

The decision maker not only faces estimation risk with respect to θ but he might be also uncertain about his loss function l. To remain in a well-defined setting we suppose that the loss function is known to the decision maker. Additional sources of risk is the measurement of the state variables x_T. We also assume that the decision maker does not suffer of data uncertainty.

Finally, we require that no feedbacks arise. The decisions $\{d_t\}_{t\in\mathbb{N}}$ should not have an impact on the outcome of the variable of interest $\{x_t\}_{t\in\mathbb{N}}$. For portfolio optimization and a sufficiently small investor in a liquid market, the assumption is likely to be satisfied. In macroeconomic decision making, e.g. for monetary policy, feedback effects are likely to be relevant.

3. CHOICE OF PORTFOLIO WEIGHTS

In this section we transfer the decision problem of section 2 to asset allocation. We show how common models can be implemented in the decision problem. We consider n assets, one of which might be but need not to be the risk-free asset. The return of the assets at time t is given by the $n-$dimensional vector $r_t = (r_{t,1}, \ldots, r_{t,n})'$. We concentrate on portfolio optimization based on the returns' history only. Extensions with additional state variables such as macroeconomic history could be thought of. At time $T-1$ the investor's information is given by the returns $\{r_t\}_{t=1}^{T-1}$. He has to chose his portfolio for the next period represented by weights $w_T = (w_1, \ldots, w_n)'$ with w_i being the amount invested in asset i. We require that the investor is fully invested with possible short positions. The allowed weights are then given by $w_T \in \mathcal{W} = \{w \in \mathbb{R}^n : \iota' w = 1\}$. After one period the investor receives return $w_T' r_T$. How should the investor choose weights w_T? The choice depends on the investor's loss function, his selected model and the estimation risk of the parameters of interest.

3.1. Loss Function

In a seminal paper, Markowitz (1952) introduced portfolio optimization based on the first two moments of the returns' distribution. The parameter of interest are given by the returns' mean and variance, e.g. $\theta_T = (\mu_T, \Sigma_T)$. Given portfolio weights w_T, next period's returns have mean $w_T'\mu_T$ and variance $w_T'\Sigma_T w_T$. To evaluate the risk (or loss) of portfolio w_T, we use the common Certainty Equivalent risk measure. For simplification, we drop time index T when presenting the Certainty Equivalent (CE), given by

$$CE_\gamma(w, \theta) = w'\mu - \frac{\gamma}{2} w'\Sigma w$$

Parameter γ equals the risk aversion of the investor. The CE is positively orientated, i.e. the higher the better. The risk measure covers a broad range of potential investors. It includes the risk-neutral investor ($\gamma = 0$) as well as highly risk-averse investors such as the minimum variance investor ($\gamma \to \infty$). It can be shown that the investor maximizes the CE if his utility function is quadratic, or if r is normal distributed and an exponential utility function is applied, or if the investment horizon is short. The information set of the investor consists of past returns only, $\mathcal{F}_{T-1} = \{r_t\}_{t=1}^{T-1}$. Using some model M, he estimates the parameters of interest,

$$\hat{\theta}_T = \left(\hat{\mu}, \hat{\Sigma}\right) := M(\mathcal{F}_{T-1})$$

The optimal portfolio weights for the investor are then given by

$$\hat{w}_T = \arg\max_{w \in \mathcal{W}} CE(w, \hat{\theta}_T)$$

Unfortunately, θ might be rather difficult to estimate which brings us to the next point.

3.2. Estimation Risk

The parameter of interest $\theta = (\mu, \Sigma)$ consists of the first and second moment of the returns. As the CE is a rather general loss function, we concentrate our further analysis on the CE. Assume the investor wants to optimize his CE, i.e. his loss function is given by $l(w, \theta) = w'\mu - \frac{\gamma}{2} w'\Sigma w$. Of his estimates of the first two moments $\hat{\theta}$, the investor obtains his optimal weights $\hat{w} = \arg\max_{w \in \mathcal{W}} l(w, \hat{\theta})$. The intuitive approach is to replace the first moments by their sample counterpart, i.e. $\hat{\theta} = \left(\hat{\mu}_T, \hat{\Sigma}_T\right)$. Unfortunately, based on the sample counterparts, the portfolio suffers of high estimation risk. Britten-Jones (1999) shows that the sampling error of the weights is large. In particular the mean of each asset is difficult to estimate (see Merton, 1980 or Best and Grauer, 1991 for a sensitivity analysis). To estimate the mean more stable, one could assume that the mean of all assets correspond to the average mean, i.e. $\overline{\mu} = \iota \cdot \frac{1}{n \cdot (T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T-1} r_{i,t}$. Looking at our risk measures we find that $w'\mu = \frac{1}{n \cdot (T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T-1} r_{i,t}$ is independent of the weights $w \in \mathcal{W}$. In this case the CE_γ concentrates on minimizing the variance only. Intermediated approaches could be thought of. An approach that neither tries to estimate each mean return individually, nor restricts all return means to be equal. An example is the Bayes-Stein model proposed by Jorion (1986). The model shrinks the mean towards some predetermined target mean. The shrinkage intensity is selected by the Stein (1955) method. Black and Litterman (1992) show how one can incorporate own views into portfolio optimization.

Not only the estimation risk in the mean, but also the estimation risk in the covariance matrix is large (Chan et al., 1999). Several approaches to reduce estimation risk have been proposed. Similar to before, the strongest restriction one could impose is that the covariances are zero and the variances are equal, i.e. $\hat{\Sigma} = c \cdot I$ with I being the identity matrix. Shrinkage approaches to this identity matrix or to the single factor model of Sharpe (1963) have been proposed by Ledoit and Wolf (2003, 2004a,b).

A variety of different models proposes different stable estimation procedures to determine the mean and the covariance. In the following section we discuss the some common models.

3.3. Models

Different estimation procedures result in different portfolio weights. We discuss various estimation procedures and show the link to an unified asset allocation framework. Consider some estimates of the first two moments, i.e. $\hat{\theta} = (\hat{\mu}, \hat{\Sigma})$. The optimal weights are then given by

$$w^{opt} = \arg\max_{w \in \mathcal{W}} CE_\gamma(w, \hat{\theta}) \tag{3}$$

$$= \arg\max_{w \in \mathcal{W}} w'\hat{\mu} - \frac{\gamma}{2} w'\hat{\Sigma} w \tag{4}$$

Fortunately a closed form solution for equation 3 exists and is given by

$$w^{opt} = \frac{\hat{\Sigma}^{-1}\iota}{\iota'\hat{\Sigma}^{-1}\iota} + \frac{1}{\gamma}\left(\hat{\Sigma}^{-1} - \frac{\hat{\Sigma}^{-1}\iota\iota'\hat{\Sigma}^{-1}}{\iota'\hat{\Sigma}^{-1}\iota}\right)\hat{\mu}$$

There are several approaches to estimate the first two moments. Different estimation procedures correspond to different asset allocation models. Let the estimation be based on the sample counterpart of the first two moments, e.g. $\hat{\mu}_T = \frac{1}{T-1}\sum_{t=1}^{T-1} r_t$ and $\hat{\Sigma}_T = \frac{1}{T-2}\sum_{t=1}^{T-1}(r_t - \hat{\mu}_T)'(r_t - \hat{\mu}_T)$. Applying the sample estimators to optimize the CE (i.e. $\hat{\theta}^{(MV)} = (\hat{\mu}_T, \hat{\Sigma}_T)$) results in the *Mean Variance (MV)* model, i.e. $w_T^{(MV)} = \arg\max_{w \in \mathcal{W}} w'\hat{\mu}_T - \frac{\gamma}{2} w'\hat{\Sigma}_T w$. The closed form solution is then given by

$$w^{(MV)} = \frac{\hat{\Sigma}_T^{-1}\iota}{\iota'\hat{\Sigma}_T^{-1}\iota} + \frac{1}{\gamma}\left(\hat{\Sigma}_T^{-1} - \frac{\hat{\Sigma}_T^{-1}\iota\iota'\hat{\Sigma}_T^{-1}}{\iota'\hat{\Sigma}_T^{-1}\iota}\right)\hat{\mu}_T$$

As discussed in section 3.2, the estimation risk of the mean is high. Let all returns be restricted to have equal means, i.e. $\overline{\mu} = \left(\frac{1}{n(T-1)}\sum_{i,t} r_{t,i}\right)\iota$. In this case the optimization results in the *minimum variance (MinVar)* weights, i.e. $w_T^{(MinVar)} = \arg\min_{w \in \mathcal{W}} w'\hat{\Sigma}_T w$. A closed form solution is given by $w^{(MinVar)} = (\iota'\hat{\Sigma}_T^{-1}\iota)^{-1}\hat{\Sigma}_T^{-1}\iota$. The investor obtains the MinVar weights, if he optimizes with respect to $\hat{\theta}^{(MinVar)} = (\overline{\mu}, \hat{\Sigma}_T)$, i.e. $w^{(MinVar)} = \arg\max_{w \in \mathcal{W}} CE_\gamma\left(w, \hat{\theta}^{(MinVar)}\right)$.

TABLE 1.

List of asset allocation models.

Asset Allocation Model	Reference / Description	$\hat{\theta}$	Abbreviation
Mean Variance	Best and Grauer (1991a)	$(\hat{\mu}_T, \hat{\Sigma}_T)$	MV
MV without Short-selling	Jagannathan and Ma (2003)	$(\overline{\mu}, S)$	MVSR
Minimum Variance	Merton (1980)	$(\overline{\mu}, \hat{\Sigma}_T)$	MinVar
Equally weighted	DeMiguel et al. (2009)	$(\overline{\mu}, \overline{\sigma}^2 I)$	EQ
Bayes-Stein	Jorion (1986)	$(\mu^{(BS)}, \hat{\Sigma}_T)$	BS
Ledoit-Wolf	Ledoit and Wolf (2004a)	$(\overline{\mu}, \Sigma^{(LW)})$	LW
Weight combination	$w^{(comb)} = \sum \pi^{(i)} w^{(i)}$	θ of individual models	Comb
Average combination	$w^{(average)} = \frac{1}{6}\sum w^{(i)}$	θ of individual models	Average
Moment combination	$w^{(mom)} = \arg\max_w CE(w, \theta^{(mom)})$	$\theta^{(mom)} = \sum \pi^{(i)}\theta^{(i)}$	Mom
Best individual	$w^{(ind)} = w^{(i^\star)}$, $i^\star = \arg\max_i \pi^{(i)}$	$\theta^{i^\star}$	Ind

The table lists the considered asset allocation models along with its original (or prominent) reference or a brief description, resp. The last two columns denote the moment estimator of each model and the abbreviation.

Jagannathan and Ma (2003) analyze the *mean variance short-selling restricted (MVSR)* portfolio, i.e. $w_T^{(MVSR)} = \arg\min_{w \in \mathcal{W}, w_i \geq 0} w'\hat{\Sigma}_T w$. They find that the optimization is equivalent to optimize $w^{(MVSR)} = \arg\min_{w \in \mathcal{W}} w'Sw$ where $S = \hat{\Sigma}_T + \delta\iota' + \delta'\iota$ and δ the Lagrange multipliers for the nonnegativity constraints. Under the CE, the MVSR weights are obtained if the investor optimizes with respect to $\hat{\theta}^{(MVSR)} = (\overline{\mu}, S)$. The short-selling restricted portfolio is a special case of the L_1 norm regularization (DeMiguel et al., 2009).

The *Bayes-Stein (BS)* model (Jorion, 1986) is obtained by shrinking the mean towards some prior value, i.e. $\mu^{(BS)} = (1-\lambda)\hat{\mu}_T + \lambda\mu^{(Target)}$. Jorion selects the mean of the minimum variance portfolio as the target mean. The Bayes-Stein Model corresponds then to the estimated parameters $\hat{\theta}^{(BS)} = (\mu^{(BS)}, \hat{\Sigma}_T)$.

The *Ledoit and Wolf (LW)* portfolio is given by $w^{(LW)} = \arg\min_{w\in\mathcal{W}} w'\Sigma^{(LW)}w$ with the Ledoit-Wolf covariance matrix $\Sigma^{(LW)} = \delta F + (1-\delta)\hat{\Sigma}_T$. The shrinkage target can be a single factor model or a constant correlation matrix (for further information see Ledoit and Wolf, 2003, 2004a,b). We apply the constant correlation approach. The Ledoit and Wolf model is then given by $\hat{\theta}^{(LW)} = (\overline{\mu}, \Sigma^{(LW)})$.

The *equally weighted (EQ)* portfolio ($w^{(EQ)} = \frac{1}{n}\iota$) performs surprisingly well as it not suffers of estimation risk (DeMiguel et al., 2009). The equally weighted portfolio corresponds to the investor optimizing $\hat{\theta}^{(EQ)} = (\overline{\mu}, \overline{\sigma}^2 I)$ with the average variance being $\overline{\sigma}^2 = \frac{1}{n(T-2)}\Sigma_{t,i}(r_{t,i} - \hat{\mu}_{T,i})^2$.

We find that common asset allocation models can be incorporated into the framework of eq. 3 by using different estimators for $\hat{\theta}$. Table 1 presents an overview of the stated models, the corresponding moment estimators and their abbreviations.

3.4. Which model is best?

Merton (1980) shows that the mean is difficult to estimate. As the mean estimate contains high estimation risk, these days most models relay on the estimation of the covariance matrix only. The estimation risk of the covariance matrix was encountered by various shrinkage approaches. The shrinkage of the covariance matrix is related to the shrinkage of the norm of the weights (Fan et al., 2012). Step by step literature moved forward, characterized by the quest for the best model. Recent horse races, however, showed that there is no generally best model. DeMiguel et al. (2009) conduct a large horse race of many asset allocation models using various data sets. Their main finding is that it is hard to significantly beat the equally weighted portfolio. Their study also reveals that the optimal portfolio depends on the applied data set. For some data sets the MinVar model performs best, for others it is the MVSR or the EQ. In one case (the SMB and HML portfolio) even the unstable MV portfolio performs best. This finding is not surprising. In turbulent periods estimation risk is high. High regularized asset allocation model as the equally weighted portfolio will perform well. In calm and stable periods estimation risk is low. Despite its sensitivity to estimation risk (Best and Grauer, 1991b), in these periods the standard MV portfolio can perform well.

We conclude that it is unlikely to find a generally best model. An attractive alternative is to let data select or combine the optimal model from a set of asset allocation models. But why should the combination work well?

Reality is usually much more complex than reflect by low parameterized models. High dimensional models suffer of high estimation risk. A combination can deliver a good trade-off between capturing complex reality and reducing estimation noise. The combination of several misspecified models might better reflect reality than an individual model. The same holds true if models are biased. If some models are upward biased and others are downward biased, the combination can be unbiased. Even if the best model is available in a large pool of models, it can be unlikely that the forecaster selects this model ex-ante. A combination of different models can give an insurance of against choosing the wrong model. The idea of combination is pursued in the following section.

4. COMBINATION AND SELECTION OF MODELS

Consider a set of m asset allocation models. These models correspond to m different estimation procedures of the parameter of interest, i.e. $\Theta_T = \left(\hat{\theta}_T^1, \ldots, \hat{\theta}_T^m\right)$ with $\hat{\theta}_T^i$ being the moments estimated by the ith estimation procedure. The corresponding portfolio weights are given by $\mathbf{W}_T = \left(w_T^1, \ldots, w_T^m\right)$. Each element represents the optimal weight with respect to the considered asset allocation model, i.e. $w_T^i = \arg\max_{w \in \mathcal{W}} CE\left(w, \hat{\theta}_T^i\right)$.

There are mainly three alternatives how to make use of the m asset allocation models. The combination of the portfolio weights, the combination of the parameter of interest or selection of individual models. The combination / selection methods are summarized in table 1.

4.1. Combination of Weights

The first method is to combine portfolio weights. Consider the shares $\pi = (\pi^1, \ldots, \pi^m)$ with $\iota'\pi = 1$. The element $\pi^i \geq 0$ represents the share of the ith model in the combination. Then the combined portfolio weight is given by

$$w^{(comb)} = \sum_{i=1}^{m} \pi^i \cdot w_T^i \tag{5}$$

The question on how to select the shares π, we tackle in section 4.5.

4.2. Combination of Moments

The second method refers to the combination of the parameters of interest. In our case the parameters of interest are the moments $\theta = (\mu, \Sigma)$. Instead of combining the weights, one could combine moments instead. Using some shares π, the combined moments are then given by $\hat{\theta}^{(mom)} = (\hat{\mu}^{(mom)}, \hat{\Sigma}^{(mom)})$ with $\hat{\mu}^{(mom)} = \sum_{i=1}^{m} \pi^{(i)} \hat{\mu}^{(i)}$ and $\hat{\Sigma}^{(mom)} = \sum_{i=1}^{m} \pi^{(i)} \hat{\Sigma}^{(i)}$.

The optimal weights of the moment combining approach are then given by

$$w^{(mom)} = \arg\max_{w \in \mathcal{W}} CE\left(w, \hat{\theta}^{(mom)}\right) \tag{6}$$

4.3. Selection

Finally, we approach the third method: the selection of the best model. As before, let the shares be $\pi = (\pi^1, \ldots, \pi^m)$. The share should be high for good models, and low for bad performing models. The best individual model is given by the asset allocation model which obtains the highest share. Formally, the best model equals $w^{(ind)} = w^{(i^\star)}$ with $i^\star = \arg\max_{i \in \{1,\ldots,m\}} \pi^i$.

Two main points shall be highlighted here. The combination can be superior the best individual model. A short and simple example shall highlight this fact. Let there be two assets ($n = 2$) and two allocation models ($m = 2$). Let the two assets have the same characteristic without perfect correlation, i.e. $\mu^{(1)} = \mu^{(2)}$, $\sigma^{(1)} = \sigma^{(2)}$ and $\rho \neq 1$. Let the weights of the asset allocation model be $w_T^1 = (\omega, 1 - \omega)$ and $w_T^2 = (1 - \omega, \omega)$ for some $\omega \in (0, \frac{1}{2})$. As both weights lead to the same performance. But *any* strict combination of the weights $w^{(comb)} = \pi^1 w_T^1 + \pi^2 w_T^2$ outperforms the best individual model (see A.1). The other problem to be mentioned is the instability of solution. The asset allocation shares are either estimated or determined by the investor. What happens if the investor changes the shares slightly? Consider the combination first. The shares are given by $\pi = (\pi^1, \ldots, \pi^j, \ldots, \pi^k, \ldots, \pi^m)$. Assume, the investor applieds the shares $\tilde{\pi} = (\pi^1, \ldots, \pi^j - \varepsilon, \ldots, \pi^k + \varepsilon, \ldots, \pi^m)$. The change of the weights is then given by

$$||\pi' \mathbf{W}_T - \tilde{\pi}' \mathbf{W}_T||_\infty = \varepsilon ||w_T^j - w_T^k||_\infty \leq 2\varepsilon ||\mathbf{W}_T||_\infty$$

with $||.||_\infty$ being the maximum norm. The change of the weights is bounded if the shares changes slightly. In case of selection we find a different pattern. Let e.g. $\pi^k = \max \pi < \pi^j + \varepsilon$. Consider the same change as above. Then the best individual model is $w_T^{(ind)} = w_T^k$, while the individual model used of the investor is given by $\tilde{w}_T^{(ind)} = w_T^j$. The difference can be bounded as follows $||w_T^{(ind)} - \tilde{w}_T^{(ind)}||_\infty \leq 2 \cdot ||\mathbf{W}_T||_\infty$. In case of model selection, a small change of the shares can induce a large shift of the weights. We regard these instabilities as problematic with regard to potential estimation risk.

4.4. Difference in the Combination Approaches

We discussed before, a selection of the (seemingly) best model can lead to unstable models and is therefore not recommendable. In this section we discuss the differences between both combination procedures: the combination of weights (*Comb*) and the combination of moments (*Mom*). The *Comb* can be regarded as combining the outcomes (or generally decisions) of a system. The *Mom* refers to the combination of input parameters the decision is based on. Both procedures are appropriate to counteract estimation risk. The question arises if one approach is better than the other.

Assume that the available asset allocation models have different mean estimates but the correct covariance estimate. Then both combination approaches result are equivalent.

PROPOSITION 1. *Let there be n asset allocation models $(\hat{\mu}_i, \Sigma)$ $i = 1, \ldots, n$ with μ, Σ being the moments of the returns and π the model shares. Then the combination of the weights as in eq. 5 and the combination of the moments as in eq. 6 lead to the same results, i.e. $w^{(comb)} = w^{(mom)}$.*

Proof. See A.2 ▮

For difference in the estimates of the covariances, a similar result to proposition 1 cannot be given. It also cannot be said which combination method is better. For different covariance estimates, the combination of the weights can but need not to be superior to the combination of moments. This fact is highlighted by the following simple example.

EXAMPLE 4.1. Let the covariance matrix be $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Consider the average of the true model and an alternative model, i.e. $\pi = \left(\frac{1}{2}, \frac{1}{2}\right)$. The covariance matrix of the alternative model is denoted by $\tilde{\Sigma}$ and the corresponding weights by $\tilde{w} = (\iota'\tilde{\Sigma}^{-1}\iota)^{-1}\tilde{\Sigma}^{-1}\iota$. The applied weights can be determined by (i) averaging the weights, i.e. $w^{(comb)} = \frac{1}{2}(w + \tilde{w})$ (see eq. 5) or (ii) averaging the moments, i.e. $w^{(mom)} = (\iota'\Sigma_c^{-1}\iota)^{-1}\Sigma_c^{-1}\iota$ with $\Sigma_c = \frac{1}{2}(\Sigma + \tilde{\Sigma})$ (see eq. 6). It depends on the alternative model which method is better. Table 2 presents two different covariances $\tilde{\Sigma}$. For one it is better to combine the weights, for the other it is better to combine moments.

4.5. Model Shares

How should the shares $\pi = (\pi^1, \ldots, \pi^n)$ of the models be determined? Model averaging is often performed by defining the shares on some information criterion, e.g. AIC, BIC (see e.g. Hjort and Claeskens, 2003). In the case of portfolio optimization these information criteria cannot be applied as the likelihood is unknown. Alternatively, one could determine the shares by the corresponding loss l. As the loss is often negative, some

TABLE 2.

Example 1:

Average $w, \tilde{w}$ / $\Sigma, \tilde{\Sigma}$	$w^{(comb)\prime}\Sigma w^{(comb)}$	$w^{(mom)\prime}\Sigma w^{(mom)}$
$\tilde{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1.5 \end{pmatrix}$	0.5050	0.5062
$\tilde{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}$	0.5139	0.5102

Table contains variance if weights are determined by (i) weights averaging (first column, i.e. $\frac{1}{2}(w + \tilde{w})$) or (ii) moment averaging (second column, i.e. $\frac{1}{2}(\Sigma + \tilde{\Sigma})$). The minimum variance of 0.5 is given for $w = (0.5, 0.5)'$.

transformation is needed to obtain positive shares. Often the exponential weighting is used, e.g.

$$\pi^m = \frac{\exp\left(-\lambda l(w^m, \theta)\right)}{\sum_{i=1}^n \exp\left(-\lambda l(w^i, \theta)\right)}$$

The function is sensitive to scaling in form of $\lambda \geq 0$. If the loss depends on basis points of the return rather than percentage points, the loss is scaled by a factor $\lambda = 100$. Everything but λ being constant, the solution can range from model selection (i.e. $\lambda \to \infty$) to naive weighting (i.e. $\lambda \to 0$). An intuitive idea is to set π^i equal to the probability that model i is best. Then the combination of the portfolios is given by

$$w^{(comb)} = \sum_{i=1}^{n} \pi^i \cdot w^i \tag{7}$$

with $\pi^{(i)}$ being the probability that current model dominates all other models, i.e. $l(w^i, \theta) \leq l(w^j, \theta)$ for all $j \neq i$[1]. We apply a bootstrap method to estimate the probabilities. For returns $r_1, \ldots, r_T$ we generate a random sample with replacement of T returns, $r_1^\star, \ldots, r_T^\star$. We apply all m asset allocation model to these bootstrapped returns. The procedure is repeated B times. Let $s_{i,b} = 1$ if model i is the best model in the bth bootstrapped sample, otherwise $s_{i,b} = 0$. The probability of model i being best, is estimated by

$$\hat{\pi}^i = \frac{1}{B} \sum_{b=1}^{B} s_{i,b} \tag{8}$$

[1]We assume that it holds that $l(w^{(i)}, \theta) \neq l(w^{(j)}, \theta)$ for any $j \neq i$. Otherwise the probabilities $\pi^{(i)}$ do not sum up to one and a marginal correction is needed in equation 7.

TABLE 3.

List of Data Sets.

No.	Data Set	n	Time period	Source	Abbreviation
1	10 industry portfolio	10	07/1963-12/2012	French	Ind10
2	30 industry portfolio	30	07/1963-12/2012	French	Ind30
3	48 industry portfolio	48	07/1969-12/2012	French	Ind48
4	6 Fama-French portfolios sorted by book-to-market	6	07/1927-12/2012	French	6BookMarket
5	25 Fama-French portfolios sorted by size and momentum	25	07/1927-12/2012	French	25SizeMom
6	Dow Jones Industrial	30	02/1973-12/2012	Datastream	Dow

The table lists the considered data sets of monthly returns. The number of assets n, the time period spanned by the data set, the source of the data, and the abbreviation used to refer to each data set.

The method is related to bagging Breiman (1996a,b). In our following analysis, the model shares $\pi = (\hat{\pi}^1, \ldots, \hat{\pi}^m)$ are estimated by eq. 8. There is no reason to believe that our bagging strategy is the best choice to obtain shares π. Nevertheless, the simple strategy appears to work well. More sophisticated methods to determine the shares are likely to provide even better results. An additional point is the autocorrelation structure of the time series data. One might consider block bagging (Politis et al., 1999). In our application we find that the improvement of using block bagging is small (not stated), given the uncertainty of selecting the block size. Therefore we rely on the common bagging procedure, e.g. the block size is equal to one.

5. EMPIRICAL STUDY

To compare empirically the out-of-sample performance of the combined strategy to the individual strategies, we apply six different data sets. The data sets considered are listed in table 3. We include the Fama French Industry portfolios for various sizes to analyze the models' performances when the number of assets increases. We include alternative sorting methods, namely book-to-market, size and momentum. Finally, we consider the most common equity index, the Dow Jones Industrial. The data sets are common to literature and (apart from the Dow Jones Industrial) are freely available[2]. The applied models were introduced in section 3.3 and are summarized in table 1.

5.1. Methodology for Out-of-Sample Evaluation

We calculate the CE performance of the individual models and the combination methods. We apply the following rolling-window procedure. The

[2]See French's Database on
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

estimation window is denoted by $\tau < T$ with T being the total number of returns. For the empirical study we chose the length $\tau = 60$, which corresponds to 5 years. Standing at time $t-1$ and using model $i \in \{1, \ldots, m\}$, the weights w_t^i are determined based on the past τ return observations $r_{t-\tau}, \ldots, r_{t-1}$. The rolling window approach is repeated for the next month by including next month's returns and dropping the returns of the earliest month. The approach is continued to the end of the data set. At the end there are $T-\tau$ portfolio weights for each asset allocation strategy i, i.e. w_t^i with $t = \tau+1, \ldots, T$, $i = 1, \ldots, m$. Using strategy i and being at time t leads to the out-of-sample return $r_t^i = r_t' w_t^i$ with r_t being the asset returns. The time series of returns can then be used to determine the mean and variance of each strategy, i.e.

$$\left(\hat{\sigma}^i\right)^2 = \frac{1}{T-\tau-1} \sum_{t=\tau+1}^{T} (r_t^i - \hat{\mu}^i)^2$$

where

$$\hat{\mu}^i = \frac{1}{T-\tau} \sum_{t=\tau+1}^{T} r_t^i$$

The corresponding CE is then given by

$$\hat{CE}^i = \hat{\mu} - \frac{\gamma}{2}\left(\hat{\sigma}^i\right)^2$$

Consider the combination / selection methods of m asset allocation models. Denote the matrix of portfolio weights by $w_t = \left(w_t^1, \ldots, w_t^m\right)$ and the matrix of the corresponding moments by $\theta_t = \left(\theta_t^1, \ldots, \theta_t^m\right)$. For each time t, the model share π_t^i is based on bootstrapping the past returns. The bootstrapping procedure is described in section 4.5. Note that the shares $\pi_t = \left(\pi_t^1, \ldots, \pi_t^m\right)$ are also computed out-of-sample. We use four different model combination/selection methods. The weights combination (*Comb*) is given by $w_t^{(comb)} = \pi_t' w_t$ (see eq. 5). The average combination (*Average*) is given by $w_t^{(average)} = \frac{1}{m}\iota' w_t$, which is the naive combination over all asset allocation models. The moment combination (*Mom*) is given by $w_t^{(mom)} = \arg\max_w CE(w, \theta^{(mom)})$ with $\theta^{mom} = \pi_t' \theta_t$ (see eq. 6). Finally, we also give a selection method picking only the best individual model (*Ind*), i.e. $w_t^{(ind)} = w^{(i^\star)}$ with $i^\star = \arg\max_i \pi^{(i)}$.

We measure the statistical difference between the CEs of an asset allocation model to our benchmark strategy (*Comb*) by a bootstrap method. In particular the $p-$values are computed by the methodology proposed by Ledoit and Wolf (2008). The methodology accounts for fat tails, autocorrelation and volatility clustering of the asset returns. As the methodology of

Model	Ind10	Ind30	Ind48	6BookMarket	25SizeMom	Dow
Comb	0.0174	0.0079	0.0176	0.014	0.0091	0.0088
Average	0.0221	-0.0319	-2.842	0.0169	-0.0003	-0.0126
	(0.8407)	(0.0005)	(0.0105)	(0.8307)	(0.1688)	(0.0005)
Mom	0.012	0.0049	0.0099	0.008	0.0076	-0.0077
	(0.0844)	(0.1179)	(0.0734)	(0.002)	(0.0854)	(0.0015)
Ind	0.0098	0.0074	0.0167	0.0109	0.0086	0.0088
	(0.0105)	(0.3432)	(0.2138)	(0.0115)	(0.1678)	(0.478)
MV	-0.1276	-0.9799	-79.4689	-0.0119	-0.9192	-0.3326
	(0.001)	(0.0005)	(0.0045)	(0.001)	(0.002)	(0.0005)
MVSR	0.0126	0.008	0.0172	0.0083	0.0088	0.0102
	(0.0085)	(0.533)	(0.4091)	(0.0754)	(0.4301)	(0.8616)
MinVar	0.0091	0.0058	0.0115	0.0087	0.0096	0.0018
	(0.011)	(0.1449)	(0.0884)	(0.1004)	(0.6693)	(0.0175)
EQ	0.0131	0.0072	0.0166	0.0055	0.0058	0.0077
	(0.044)	(0.3132)	(0.2428)	(0.0205)	(0.007)	(0.3037)
BS	-0.0223	-0.1272	-3.0267	0.0123	-0.1466	-0.0472
	(0.014)	(0.0005)	(0.046)	(0.2817)	(0.014)	(0.0005)
LW	0.0115	0.0082	0.0124	0.008	0.0084	0.005
	(0.0405)	(0.5854)	(0.0385)	(0.0689)	(0.2433)	(0.0649)

For each of the considered datasets, the table reports the CE for the combination ($Comb$), the average combination ($Average$), the moment combination (Mom), the past best individual model (Ind) and the asset allocation models presented in section 3.3. In parentheses is the $p-$value of the difference between the CE of each strategy from that of the combination benchmark. The $p-$values are computed as proposed by Ledoit and Wolf (2008).

Ledoit and Wolf (2008) is specified for the Sharpe ratio a minor adjustment to CE has to be made (for the adjustment details see Schanbacher, 2012). The $p-$value is computed on the two-sided test with the null hypothesis $H_0 : CE^{(comb)} = CE^{(i)}$.

5.2. Discussion of Performance

Table 3 shows the out-of-sample CE for various asset allocation models and combination / selection methods. The $p-$values shows if a strategy is different to the combination ($Comb$). In the following discussion we say a difference is significant if the $p-$value is smaller than 10%.

Consider the individual asset allocation models first. In line with literature, the MV model is the only model performing continuously bad. The short-selling restricted portfolio MVSR and the minimum variance portfolio MinVar perform continuously well. Both portfolios are even best for some data sets (Ind48 and 25SizeMom, resp.). The equally weighted portfolio EQ is not exceptionally good but shows a stable performance for all data sets. This result supports the finding of DeMiguel et al. (2009) that the naive portfolio is hard to beat significantly, although it is not the best portfolio strategy. An interesting strategy is the Bayes-Stein BS strategy. The BS can perform well, but also rather poor. It confirms the empirical finding that a portfolio strategy can work well in a certain setting but completely fail in another environment. One of the best portfolio strategies is the Ledoit and Wolf (2004a) portfolio. It can also perform best (Ind30) and performs never poor. We find that no asset allocation strategy dominates over all data sets. It depends on the data set which model should be chosen. The finding supports the idea to use a combination or selection method, instead of hunting for a single best portfolio strategy.

Now we analyze the combination methods. We find that the combination *Comb* performs exceptionally well. The combination is always among the two best strategies out of all individual models and the alternative combination methods[3]. In half of the analyzed data sets (Ind10, Ind48, 6BookMarket) the combination is even better than the ex-post best individual model. Even if you knew the best model ex-ante, you could not beat the combination. The combination significantly outperforms each individual model in at least two of the considered six data sets, while it is never significantly outperformed. We find that the *Average* over all models performs very well in some cases (Ind10, 6BookMarket) but completely fails in others (Ind48). In cases when one asset strategy fails (here the MV portfolio for high dimensional asset allocation), the *Average* suffers strongly. The moment combination *Mom* turns out not to work well. The *Mom* is never as good as the *Comb*. Instead of combining the input parameters θ, it is better to combine the decisions w. Finally, we come to the selection method *Ind* picking the best individual asset allocation model. Selecting the best individual model does not work. First, in all cases it is worse than *Comb*. Second, in all cases it is worse than the best individual model. The selection method cannot accurately detect the best individual model. We conclude that ex-ante the investor should select the combination method *Comb*. The combination provides an insurance against choosing the wrong individual model. Even ex-post, the *Comb* strategy has a good chance to be superior to the ex-post best individual model.

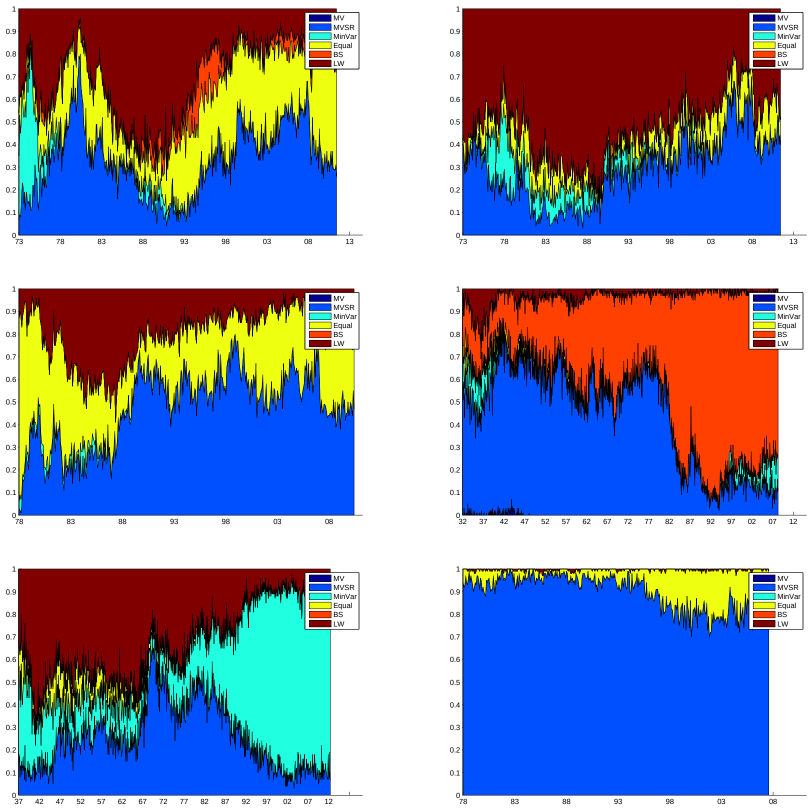
We summarize the findings over all data sets and asset allocation models (6 data sets $\times$ 6 individual models = 36 cases). The combination is never significantly beaten. In almost 9 out of 10 cases the combination outperforms the individual model. In 64% of all direct comparisons, the combination significantly beats the individual model.

[3]The only exception is Ind30 where *Comb* is third best with 0.01% CE difference to the second best.

5.3. Behavior of the Combination

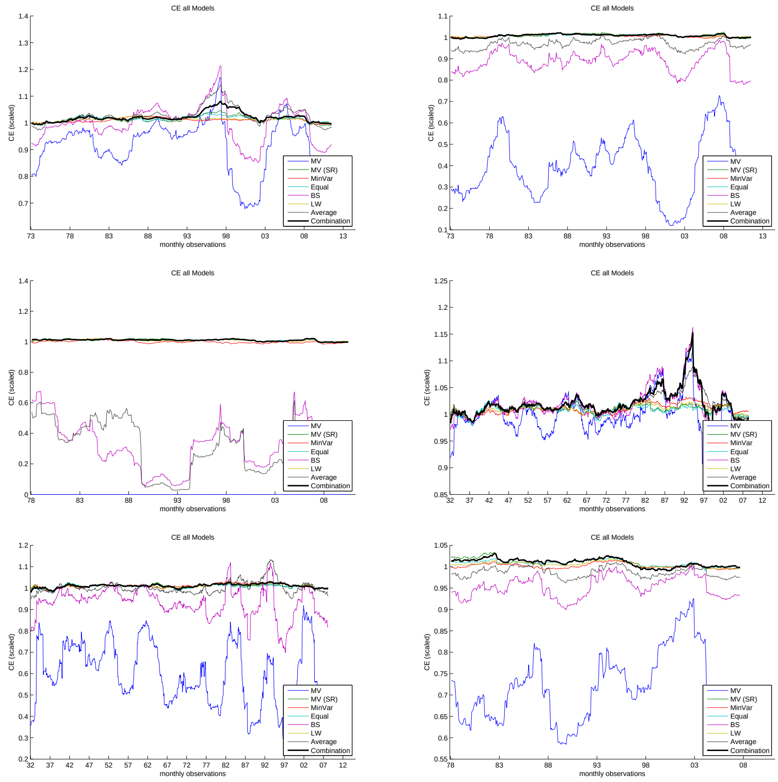
As the combination is the favorite method, we analyze its behavior. In the figures for the shares are provided. We find that there are no sudden jumps in the shares. Usually one or two models are dominating with minor parts of the remaining strategies, e.g. the LW strategy for the Ind30 data set or the MVSR strategy in the Dow data set. It can happen that the share of a model shifts slowly over time. Over a 20 years period, in the 25SizeMom data set, the share of the MinVar model increased from about 15% to about 80%. Similar results are found for the 6BookMarket data set and the BS strategy.

FIG. 1. Monthly shares for Ind10/Ind30, Ind48/6BookMarket, 25SizeMom/Dow (top down). The shares are estimated as stated in section



We find that the share of the ex-post best model is usually highest. For the Ind30 data set, the LW portfolio is strongly favored. For the Ind48 data set, the MVSR is mainly selected. Using the 6BookMarket data set,

FIG. 2. Monthly CE for Ind10/Ind30, Ind48/6BookMarket, 25SizeMom/Dow (top down). The CE of each asset allocation strategy is estimated based on the past 100 returns.



the largest share goes to the BS portfolio. And for the Dow data set, the MVSR dominates. In each case the corresponding strategy turns out to be ex-post best.

In Appendix B.3 the CE over time is given. As the CE depends on the mean and the variance, these are estimated based on the previous τ returns. We find that the combination stays always among the top best strategies. For the Ind30 data set, we find that the *Average* is strongly effect by the worse performance of the MV and the BS portfolio, while the combination is not. Looking at the period 1990-2000 of the 6BookMarket data set, we find that the performance of the combination is strongly increasing while most other strategies stay low.

We conclude that our simple proposed method leads to sensible shares of the models. The corresponding combination proves to work well over all data sets and over time.

5.4. Robustness

We find that the results are robust to various changes of the setting. The analysis is based on various data sets commonly used for horse-races in literature (see e.g. DeMiguel et al., 2009). Although the performance evaluation relied on the CE measure, the results are similar for other risk measures. In the appendix (B.1) we provide the results for using the Sharpe ratio. The chosen risk aversion equals $\gamma = 2$. We also consider alternative values, but as the insights are similar these results are not reported. As an alternative for the length of the estimation window, we also conduct the study with $\tau = 120$. These results (not stated) remain similar.

6. CONCLUSION

Literature continues to search for the single best asset allocation model. We analyze the performance of commonly used asset allocation models for standard data sets. Our results indicate that it is unlikely that there exists one individual model which continuously dominates its competitors. Instead relying on one single model, one could combine or select from a set of different asset allocation models. We contribute to literature by proposing a general setup to combine the weights or moments for a wide range of different asset allocation models. We also propose a bootstrap method to determine the share of each individual asset allocation model. We find that the combination of asset allocation models appears to perform exceptionally well. The combination significantly outperforms each individual model at least once. But it is never significantly outperformed by any model. For half of our data sets, the combination even outperforms *all* individual models. For the other data sets, it is unlikely that the investor would have ex-ante chosen the ex-post best model. We find that it is difficult to select the best model. The combination approach provides an insurance of selecting the wrong model. We conclude that combination of models is not only interesting in the context of forecasting (see e.g. Timmermann, 2006 and the references therein), but also for portfolio choice. Future research should analyze more advanced methods to combine asset allocation models.

APPENDIX A
Proofs

A.1. COMMENT

Proof. There are two assets $n = 2$ and two asset allocation models $m = 2$. Both assets have the same characteristic without perfect correlation, i.e. $\mu_1 = \mu_2$, $\sigma_1 = \sigma_2$, $\rho < 1$. The weights of the asset allocation models are given by $w_1 = (\omega, 1-\omega)$ and $w_2 = (1-\omega, \omega)$ for some $\omega \in (0, \frac{1}{2})$. Consider the combination $w_{(comb)} = \pi w_1 + (1-\pi) w_2$. The mean of the combination is the same as the mean of the individual model, i.e. $\mu' w_{(comb)} = \mu' w_{(ind)}$. Hence the difference in CE is only driven by the variance. It holds that $w_1' \Sigma w_1 = w_2' \Sigma w_2$. By the concavity property of the variance for any combination $w_{(comb)}$, it holds that $w_{(comb)}' \Sigma w_{(comb)} \leq w_1' \Sigma w_{(1)} = w_2' \Sigma w_2$. ▮

A.2. PROPOSITION 2

Proof. The optimal weights are given by

$$w^i = \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} + \frac{1}{\gamma}\left(\Sigma^{-1} - \frac{\Sigma^{-1}\iota\iota'\Sigma^{-1}}{\iota'\Sigma^{-1}\iota}\right)\hat{\mu}_i$$

then the combination of the weights is given by

$$\begin{aligned} w^{(comb)} &= \sum_{i=1}^{n} \pi^i w^i \\ &= \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} + \frac{1}{\gamma}\left(\Sigma^{-1} - \frac{\Sigma^{-1}\iota\iota'\Sigma^{-1}}{\iota'\Sigma^{-1}\iota}\right)\sum_{i=1}^{n} \pi^i \hat{\mu}_i \end{aligned}$$

The combination of the moments is given by

$$\begin{aligned} w^{(mom)} &= \frac{\Sigma^{(mom)-1}\iota}{\iota'\Sigma^{(mom)-1}\iota} + \frac{1}{\gamma}\left(\Sigma^{(mom)-1} - \frac{\Sigma^{(mom)-1}\iota\iota'\Sigma^{(mom)-1}}{\iota'\Sigma^{(mom)-1}\iota}\right)\mu^{(mom)} \\ &= \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} + \frac{1}{\gamma}\left(\Sigma^{-1} - \frac{\Sigma^{-1}\iota\iota'\Sigma^{-1}}{\iota'\Sigma^{-1}\iota}\right)\sum_{i=1}^{n} \pi^i \hat{\mu}_i \end{aligned}$$

Hence $w^{(mom)} = w^{(comb)}$ ▮

APPENDIX B

Data and Plots

B.1. SHARPE RATIO

Model	Ind10	Ind30	Ind48	6BookMarket	25SizeMom	Dow
Comb	0.3532	0.2429	0.3841	0.2988	0.2419	0.2358
Average	0.3643	0.143	0.2198	0.3388	0.3309	0.0789
	(0.6014)	(0.0569)	(0.02)	(0.9765)	(0.9905)	(0.0005)
Mom	0.3053	0.1734	0.2624	0.2497	0.2146	0.0136
	(0.2488)	(0.1239)	(0.1159)	(0.0005)	(0.0754)	(0.002)
Ind	0.2453	0.2213	0.3441	0.2854	0.2317	0.234
	(0.0045)	(0.1763)	(0.009)	(0.0649)	(0.1454)	(0.3641)
MV	0.2854	0.0934	0.1934	0.2517	0.2703	0.0086
	(0.1299)	(0.0095)	(0.007)	(0.0395)	(0.7542)	(0.001)
MVSR	0.2821	0.22	0.3381	0.2278	0.2339	0.2507
	(0.0015)	(0.1918)	(0.017)	(0.019)	(0.3796)	(0.8057)
MinVar	0.3011	0.1948	0.2666	0.2511	0.259	0.1099
	(0.1738)	(0.1404)	(0.0514)	(0.1084)	(0.7887)	(0.0125)
EQ	0.305	0.2165	0.3639	0.185	0.1878	0.2295
	(0.0819)	(0.2018)	(0.1868)	(0.0015)	(0.016)	(0.4346)
BS	0.3463	0.1426	0.2719	0.3085	0.3356	0.0615
	(0.4436)	(0.046)	(0.0524)	(0.6898)	(0.9945)	(0.0025)
LW	0.3702	0.2774	0.3957	0.2318	0.256	0.1838
	(0.6508)	(0.8147)	(0.5769)	(0.041)	(0.7807)	(0.1144)

For each of the considered datasets, the table reports the Sharpe ratio for the combination (*Comb*), the average combination (*Average*), the moment combination (*Mom*), the past best individual model (*Ind*) and the asset allocation models presented in section 3.3. In parentheses is the $p-$value of the difference between the Sharpe ratio of each strategy from that of the combination benchmark. The $p-$values are computed as proposed by Ledoit and Wolf (2008).

REFERENCES

Bates, J. and C. Granger, 1969. The combination of forecasts. *Operational Research Quarterly* **20**, 451-468.

Best, M. and R. Grauer, 1991a. On the sensitivity of mean variance efficient portfolios to changes in asset means: some analytical and computational results. *Review of Financial Studies* **4 (2)**, 315-342.

Best, M. and R. Grauer, 1991b. Sensitivity analysis for mean-variance portfolio problems. *Management Science* **37 (8)**, 980-989.

Black, F. and R. Litterman, 1992. Global portfolio optimization. *Financial Analysts Journal* **48 (5)**, 28-43.

Breiman, L., 1996a. Bagging predictors. *Machine Learning* **26 (2)**, 123-140.

Breiman, L., 1996b. Heuristics of instability and stabilization in model selection. *The Annals of Statistics* **24**, 2350-2383.

Britten-Jones, M., 1999. The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance* **54(2)**, 655-671.

Chan, L. K. C., J. Karceski, and J. Lakonishok, 1999. On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies* **12**, 937-974.

Clemen, R., 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* **5**, 559-581.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal, 2009. Generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science* **55(5)**, 798-812.

DeMiguel, V., L. Garlappi, and R. Uppal, 2009. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies* **22 (5)**, 1915-1953.

Diebold, F. and P. Pauly, 1987. Structural change and the combination of forecasts. *Journal of Forecasting* **6**, 21-40.

Fan, J., J. Zhang, and K. Yu, 2012. Asset allocation and risk assessment with gross exposure constraints for vast portfolios. *Journal of American Statistical Association* **107**, 592-606.

Frost, P. A. and J. E. Savarino, 1986. An empirical bayes approach to portfolio selection. *Journal of Financial and Quantitative Analysis* **21**, 293-305.

Genre, V., G. Kenny, A. Meyler, and A. Timmermann, 2013. Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting.*

Hansen, B. E., 2010. Averaging estimators for autoregressions with a near unit root. *Journal of Econometrics* **159 (1)**, 142-155.

Hjort, N. and G. Claeskens, 2003. Frequentist model average estimators. *Journal of the American Statistical Association* **98 (464)**, 879-899.

Jagannathan, R. and T. Ma, 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance* **54(4)**, 1651-1684.

Jorion, P., 1986. Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* **21(3)**, 2279-292.

Ledoit, O. and M. Wolf, 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* **10(5)**, 603-621.

Ledoit, O. and M. Wolf, 2004a. Honey, i shrunk the sample covariance matrix. *Journal of Portfolio Management* **21(1)**, 110-119.

Ledoit, O. and M. Wolf, 2004b. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* **88**, 365-411.

Ledoit, O. and M. Wolf, 2008. Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance* **15**, 850-859.

Markowitz, H. M., 1952. Portfolio selection. *Journal of Finance* **7**, 77-91.

Merton, R. C., 1980. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* **8**, 323-361.

Pesaran, H. and A. Timmermann, 2005. Real time econometrics. *Econometric Theory* **21 (10)**, 212-231.

Politis, D. N., J. P. Romano, and M. Wolf, 1999. Subsampling. Springer-Verlag, New York.

Schanbacher, P., 2014. Averaging across asset allocation models. *Journal of Economics and Statiscs*.

Sharpe, W. F., 1963. A simplifed model for portfolio analysis. *Management Science* **9**, 277-293.

Smith, J. and K. Wallis, 2009. A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics* **71 (3)**, 331-355.

Stein, C., 1955. Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. Berkeley Symposium on Probability and Statistics.

Stock, J. and M. Watson, 2004. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* **23**, 405-430.

Timmermann, A., 2006. Handbook of economic forecasting. Elsevier, Ch. Forecast Combinations, pp. 135-196.

Tu, J. and G. Zhou, 2011. Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics* **99**, 204-215.