Portfolios of Estimated Portfolios: Combination Approaches to Estimating Optimal Portfolio Weights

Bradley S. Paye
Jones Graduate School of Management, Rice University

May 5, 2007
PRELIMINARY - Please do not cite without author’s consent

Abstract

In a canonical asset allocation problem with i.i.d. risky returns, the paper considers the problem of forming the optimal linear combination of a “basis set” of portfolio decision rules. This problem is shown to be formally equivalent to a transformed version of the original portfolio problem with the same risk-free asset and a new set of risky assets, each corresponding to a conceptual position in one of the basis estimators held over both the sample period and asset holding period. Using this fundamental equivalence, results regarding the optimal combination of estimators are derived in the general asset allocation setting as well as for the special case of mean-variance preferences. The optimal combining weights depend upon unknown properties of the returns distribution and are therefore infeasible. The paper discusses feasible approaches to forming portfolios of estimated portfolios and evaluates the practical merits of these approaches in a simulation environment. In parallel with the literature on forecast combination, the simulation evidence suggests that ad hoc combination schemes such as simple averaging across estimators often outperform more sophisticated combination approaches.
Introduction

When the investment opportunity set is constant over time and the distribution of risky asset returns is known to investors, forming optimal portfolio allocations is conceptually straightforward. In practice, however, the distribution of asset returns is unknown and investors must use historical data to draw inference regarding the optimal allocation. The fact that the optimal allocation must be estimated rather than simply computed introduces estimation risk into the portfolio choice problem. When the effect of estimation risk is ignored and portfolio allocation proceeds as though estimated characteristics of the returns distribution represent the “truth,” actual investment performance may be very poor. This has been well-documented in the mean-variance problem when the standard plug-in estimator is applied (Jobson and Korkie (1980), Michaud (1989)).

A number of existing papers propose approaches designed to ameliorate the estimation risk problem in the mean variance setting. The proposed approaches include imposing a factor structure on covariance matrix (Sharpe 1963), imposing a factor structure linking returns and the covariance matrix (MacKinlay and Pastor (2000)), imposing artificial constraints on portfolios (Jagannathan and Ma (2003)), Bayesian shrinkage estimates of the mean of asset returns (Jorion (1986)), empirical Bayesian shrinkage approaches (Jorion (1986) and Frost and Savarino (1986)), robust allocation rules (Garlappi, Uppal and Wang (2006) and Goldfarb and Iyengar (2003)), shrinkage estimators of the covariance matrix of returns (Ledoit and Wolf (2003, 2004)) and the two- and three-fund portfolios suggested by Kan and Zhou (2006).\footnote{Among existing papers on estimation risk and optimal portfolio allocation, this paper relates most closely to the paper by Kan and Zhou (2006; hereafter “KZ”). The two-fund estimator proposed by KZ shrinks the plug-in estimator of the mean-variance optimal portfolio toward the zero allocation rule while the three-fund estimator represents a linear combination of the plug-in estimator and the estimated global minimum variance portfolio. As such, the KZ two- and three-fund portfolios represent special cases of combining estimators. This paper generalizes the KZ portfolios in important ways by permitting “J-fund” portfolios that extend the range of combination possibilities and providing results in a nonparametric setting that are valid beyond the mean-variance setting.}

The estimation risk problem certainly extends to more general asset allocation contexts. Indeed, estimation risk may be more severe when investors are concerned about additional, higher moments such as the skewness and kurtosis of the distribution of portfolio returns. In this more general setting, Brandt, Santa-Clara and Valkanov (2004) attempt to mitigate estimation risk by parameterizing the optimal weights as a function of asset characteristics such as book-to-market ratio and market beta.

Despite an active literature, estimation risk remains a substantial barrier to optimal portfolio implementation. DeMiguel, Garlappi, and Uppal (2007) find that, among a large set of proposed estimators of the optimal mean-variance allocation and across a number of empirical datasets, none significantly outperforms an ad-hoc $1/N$ portfolio allocation rule when performance is measured using out-of-sample Sharpe ratios, certainty equivalent returns, and turnover. These authors conclude “there are still many ‘miles to go’ before the gains promised by optimal portfolio choice can actually be realized out-of-sample.”

The objective of this paper is to walk a few miles in the right direction by providing a general
framework for combining multiple estimators of the optimal portfolio allocation. The paper considers a simple asset allocation setting in which returns are independently and identically distributed according to some unspecified distribution. Since the distribution of asset returns is unknown, the optimal asset allocation is also unknown and the objective is to estimate this allocation. From the perspective of classical statistical decision theory, the performance of a given estimator of the optimal portfolio, termed a \textit{portfolio decision rule}, is evaluated based on the expectation of utility achieved under the estimator, where this expectation is taken with respect to the joint distribution of sample data and holding period returns. The fundamental question considered in the present paper is: \textit{Given a `basis set' of $J$ portfolio decision rules, what is the optimal linear combination of these estimators in terms of the expected utility of the combined estimator?}

I show that the problem of forming an optimal linear combination of estimators is formally equivalent to a transformed version of the original asset allocation problem with the same risk-free asset and a new set of risky assets, each corresponding to a conceptual position in one of the basis set estimators that is held over both the sample period and asset holding period. This equivalence is very convenient because standard results and intuition from classical portfolio theory become immediately applicable with respect to combining estimators. For example, the optimal combining weights for the basis set of estimators are implicitly determined by standard first-order necessary conditions for the asset allocation problem in the transformed asset space. More fundamentally, the result provides immediate intuition for why combining estimators can be helpful: since estimators may be viewed as assets, the benefits of diversification in portfolio theory suggest that it may be imprudent to put “all your eggs in one basket” and simply apply one of the basis estimators. Combination can generally improve estimator performance.

Under mean-variance preferences, more explicit results are available. In particular, it is possible to provide an analytic formula for the optimal combining weights as a function of the first and second moments of risky asset returns and portfolio decision rules. In addition to investor risk aversion, the optimal weights are shown to depend upon the expected return and covariance of risky asset returns (in the original asset space), the mean and covariance matrix of each portfolio decision rule, and cross-covariances between pairs of portfolio decision rules. The mean-variance setting also provides useful intuition for why combining estimators may improve performance. In the transformed asset space, the $J$ portfolio decision rules will generate a standard mean-variance efficient frontier. Since the basis set of portfolio decision rules will typically lie inside the hyperbola that characterizes this efficient frontier, the same expected return may be achieved at lower variance by forming a nontrivial portfolio of estimated portfolios.

When only a single portfolio decision rule is included in the basis set, forming a “combined” estimator reduces to applying a scaling factor to the estimator. The optimal choice of scaling factor has the interpretation of optimal shrinkage toward the ultraconservative portfolio decision rule that allocates 100\% to the risk-free asset, where the term shrinkage is used even though the scaling factor can in general be greater than one in magnitude.

The results established in the paper regarding the optimal combination of portfolio estimators
illustrate that, for a reasonably well chosen basis set of estimators, it is likely that there are benefits to diversification among estimators in the sense that a nontrivial combination of estimators exists that will yield superior performance. An important caveat, however, is that since the optimal combination depends upon features of the true returns distribution, the optimal combining weights are latent and the combined estimator based on them is therefore infeasible. This leads to consideration of strategies for estimating the optimal combining weights in order to achieve feasible portfolios of estimated portfolios.

Although the optimal weights are characterized by the solution to the transformed asset allocation problem, estimation is complicated by the fact that since returns in the transformed problem accrue over both the sample and holding period, the available data contain only a single realization of transformed returns. To circumvent this impediment, the paper suggests an estimator of the optimal weights based on a plug-in estimator of the true unknown distribution of returns in the original asset space. With this estimated distribution of returns in place of the true distribution, a pseudo-sample of transformed asset returns may be generated via a simulation approach described in Section 4 of the paper. An estimate of the optimal portfolio allocation over these pseudo-returns then serves as an estimate of the optimal combining weights.

Since the set of simulated pseudo-returns may be made arbitrarily large, estimation error in this stage of the procedure can be made negligible. Of course, since the estimated return distribution differs from the true distribution, what is recovered with high precision is the optimal combining weights under the estimated distribution of returns rather than the truly optimal weights. When limited data are available, it is difficult to estimate the optimal combining weights with precision due to the same issues that plague estimation of the optimal allocation in the original problem. In such cases, an alternative is to employ an ad hoc combining factor, such as a simple averaging of the basis estimators.

To assess the empirical merit of various approaches to combining estimators, the paper considers an extremely simple portfolio problem in which an investor with mean-variance preferences allocates wealth between a single risky stock index and a risk-free asset. The focus is on the well-known estimator of the optimal allocation that ‘plugs-in’ the sample mean and variance of excess returns. A number of estimators that shrink the plug-in estimator toward a 100% allocation in the risk-free asset are considered, some which attempt to estimate the optimal shrinkage factor and others that simply specify a shrinking factor in an ad hoc fashion. The performance of these estimators, measured in terms of certainty equivalent return, is assessed in both simulated data and an out-of-sample analysis using actual US index returns. In both simulated and actual data, ad hoc shrinkage methods outperform attempts to estimate the optimal combining (shrinkage) factor. Intuitively, at lower sample sizes the performance of shrinkage estimators that attempt to estimate the optimal shrinking factor suffer from substantial sample variation in the estimated shrinking rate. Because it entails no estimation risk, an ad hoc shrinking factor will outperform data-driven approaches so

---

\(^2\)Depending on the application, this plug-in estimator might be the empirical distribution or a parametric estimator.

\(^3\)The forecasting literature provides motivation for this approach, since equal-weighted forecast combinations often perform well relative to more sophisticated combination strategies. See the review article by Timmermann (2006).
long as the arbitrarily chosen shrinking rate is not too far from the truly optimal rate,

A more disconcerting result is that an estimator that simply sets the allocation to stock equal
to the inverse of the risk preference parameter (ignoring all available data) outperforms all other
approaches considered, both in simulated and actual data. The results illustrate that, even in the
absence of dimensionality concerns, estimation risk is a serious impediment to empirically driven
asset allocation. For empirically realistic sample sizes, it appears difficult to outperform a ‘naive’
allocation rule that is completely ignores the historical sample of returns. The surprisingly strong
relative performance of the 1/N estimator in the multi-asset context documented by DeMiguel,
Garlappi, and Uppal (2007) appears to have an analog in the simple setting with only a single risky
asset.

The remainder of the paper proceeds as follows. Section 1 describes the asset allocation problem
and discusses the evaluation of estimators of optimal portfolio weights. Section 2 considers the
problem of forming the optimal linear combination of a set of portfolio rules and presents the
fundamental equivalence result that links this problem with a transformed version of the original
asset allocation problem. Section 3 applies this equivalence to derive a number of results regarding
the optimal portfolio of estimated portfolios in the general asset allocation setting, under mean-
variance preferences and for the special case of a scaling or “shrinking” factor applied to a single
estimator. Section 4 of the paper discusses feasible approaches to estimating the truly optimal
combining weights. Section 5 presents the empirical analysis including simulation and out-of-sample
results. Section 6 summarizes and discusses directions for extension in future research.

1. Data-Driven Portfolio Allocation

This section describes the asset allocation problem considered in the paper and discusses estimators
of the optimal allocation from the viewpoint of statistical decision theory.

1.1. A Canonical Portfolio Choice Problem

We restrict attention in this paper to a canonical single-period (myopic) portfolio choice problem.
Consider an investor with current wealth $W$ who must allocate this wealth among $N + 1 \in \mathbb{N}$
financial assets where returns on $N$ of the assets are risky and the remaining asset offers a risk-free
gross return. The position is held for one period and is then liquidated. Denote the random gross
return over the holding period on risky asset $i$ as $\hat{R}^i$ for $i = 1, ..., N$ and the corresponding $N \times 1$
vector of risky returns as $\hat{R}$. The risk free gross return is denoted $R^f$. The $N \times 1$ vector $\omega$
will denote a portfolio where the fraction of wealth allocated to each risky asset is $\omega_i$, $i = 1, ..., N$
and the allocation to the risk-free asset, denoted $\omega_0$, is implicitly given by $\omega_0 = 1 - \omega'1_N$
under the standard wealth constraint where $1_N$ indicates an $N \times 1$ vector of ones.

The vector of excess returns on the risky assets may be expressed as $\left(\hat{R} - R^f1_N\right)$ and we denote
the excess return on arbitrary portfolio $\omega$ as

$$\hat{R}^\omega = \omega' \left(\hat{R} - R^f1_N\right) \quad (1)$$
End-of-period wealth is denoted $\tilde{W}$ and may be expressed as a function of the portfolio allocation using the fundamental relation

$$\tilde{W} = W(R^f + \tilde{R}^\omega). \quad (2)$$

For the bulk of the paper, we maintain the following assumption on the stochastic process governing risky stock returns:

**Assumption 1** Risky asset returns are independently and identically distributed (i.i.d.) according to the distribution $F(\tilde{R})$.

Assumption 1 is strong and rules out any temporal predictability in asset returns, including predictability in mean and variance. The assumption is mainly for convenience and allows us to develop the main results with a minimum of notational and technical fuss.

The investor’s preferences are assumed to admit an expected utility representation with a strictly increasing, concave and twice continuously differentiable von Neumann-Morgenstern utility function $U(\tilde{W})$ defined over end-of-period wealth. The investor’s portfolio choice problem may be expressed as

$$\max_{\omega} \mathbb{E}[U(\tilde{W})] = \max_{\omega} \int U(W(R^f + \tilde{R}^\omega))dF(\tilde{R}). \quad (3)$$

Let the optimal portfolio policy representing the solution to (3) be denoted as $\omega^*$. The optimal policy is characterized by the first-order necessary conditions for the investor’s optimization problem:

$$\omega^* \equiv \omega \text{ s.t. } \mathbb{E}_{F(\tilde{R})}[U(t(W[R^f + \tilde{R}^\omega]) \left( \tilde{R} - R^f 1_N \right)] = 0.$$  

Note that this formulation of the problem imposes no constraints regarding short sales or borrowing or other constraints such as restrictions on the maximum position that may be taken in a particular asset.

The portfolio allocation problem in (3) implicitly assumes that the distribution of returns $F(\tilde{R})$ is known to the investor. When the stochastic law governing future risky returns is known, the portfolio allocation problem requires no data and simply involves solving for $\omega^*$, possibly numerically. Such an assumption; however, is not realistic practice and the central focus of this paper is the realistic scenario in which the investor does not know the distribution $F(\tilde{R})$, even up to a specific parametric model. We will assume that while $F(\tilde{R})$ is unknown, a sample of historical data is available that may be used to learn about $F(\tilde{R})$ to conduct inference regarding the optimal allocation $\omega^*$.

1.2. **Portfolio Decision Rules: Estimators of the Optimal Allocation**

Suppose that a historical sample of returns of size $T$ is available and may be used to estimate the unknown optimal portfolio. We use the notation $r^T \equiv \left(r^T_1, \ldots, r^T_T \right)^\prime$ to denote a specific sample realization of the random sample of historical risky returns $R^T \equiv \left(R^T_1, \ldots, R^T_T \right)^\prime$. 

5
A portfolio decision rule $\hat{\omega}$ is defined as a (measurable) mapping $\hat{\omega} : R^T \rightarrow A$ from realizations of the sample data to the admissible set of holdings for risky assets $A$, where for the bulk of the paper we take $A = \{ \omega \in \mathbb{R}^N \}$, implying no constraints on positions in the risky assets other than the wealth constraint. This implicitly permits short sales in assets, unlimited borrowing at the risk-free rate and arbitrarily large positions in securities.\footnote{See Section X of the paper for extensions that permit additional constraints along these lines. The estimated allocation to the riskless asset in the optimal portfolio is implicitly given by $\hat{\omega}_0 = 1 - \hat{\omega}'1$.} Intuitively, a portfolio decision rule is simply a statistical estimator of the unknown optimal portfolio allocation $\omega^*$.\footnote{This approach ignores parameter estimation uncertainty, i.e., the affect of sampling variation in parameter estimates on the portfolio choice problem.}

There are many possible approaches to estimating the optimal portfolio policy. A common approach assumes a parametric model for the joint distribution of returns on risky assets. Given this parametric model, unknown parameters are estimated from historical data and then the portfolio problem in (3) is solved (possibly numerically) with the estimated set of parameters replacing the true unknown parameters. This general approach encompasses a vast number of different portfolio estimators, which vary not only over the distributional model assumed for returns (e.g., multivariate normal or Student-\(t\)) but also with respect to the estimators for the unknown model parameter (e.g., the traditional covariance matrix estimator versus the single-factor model of Sharpe (1963)).\footnote{See Berger (1985) for a formal treatment of statistical decision theory.} Additionally, nonparametric methods that make no distributional assumptions in estimating the optimal allocation are also available. Brandt (1999) develops such an approach based on the moment conditions characterizing the optimal allocation and establishes its consistency.

\subsection*{1.3. Evaluating and Comparing Portfolio Decision Rules}

In light of the multitude of available approaches to estimating the optimal portfolio allocation, it is natural to seek a means to evaluate and compare alternative estimators. Statistical decision theory provides a foundation for the evaluation of portfolio decision rules.\footnote{See Berger (1985) for a formal treatment of statistical decision theory.} From the perspective of statistical decision theory, the performance of an estimator is based on the concept of expected loss, or risk. In the context of optimal portfolio allocation, the investor’s utility function provides a natural basis for the loss function under which to compute the risk associated with a portfolio decision rule. Specifically, define the loss function associated with an arbitrary portfolio decision rule $\hat{\omega}$ as the random variable

$$L(\hat{\omega}, \hat{\omega}) = -U(WR^f + \hat{\omega})).$$

(4)

In words, (4) says that the loss function for a portfolio decision rule is simply minus the utility of end-of-period wealth achieved under the decision rule. It is important to emphasize that the randomness in the loss function arises from two distinct sources. The first source of randomness is simply the stochastic nature of risky asset returns over the holding period, governed by the distribution of asset returns $F(\hat{\omega})$. The second source of randomness stems from the fact that $\hat{\omega}$ is...
a random variable that varies in accordance with the random sample $R_T$ available for estimation. Let $F_T(\hat{\omega})$ represent the distribution of the portfolio decision rule for a given sample size $T$. This distribution is ultimately determined by the distribution of risky asset returns $F(\hat{R})$ along with the form of the estimator.\footnote{Note that under the i.i.d. assumption $F(R_T) = F(R_{t+1})^T$.} The stochastic behavior of end-of-period wealth is governed by the the joint distribution of the estimated portfolio weights and future returns, denoted $F_T(\hat{\omega}, \hat{R})$. The dependence of $F_T(\hat{\omega}, \hat{R})$ on the sample size $T$ is inherited from the dependence of $F_T(\hat{\omega})$ on $T$. In the sequel, we will frequently suppress the dependence on the sample size to simplify notation and refer to, e.g., $F(\hat{\omega})$ where the dependence on $T$ is understood. Under our maintained assumption that risky returns are i.i.d., this joint distribution factors as $F(\hat{\omega}, \hat{R}) = F(\hat{\omega})F(\hat{R}).$\footnote{Under the assumption that returns are i.i.d. (Assumption 1) we have the relation $F(R_T, R_{t+1}) = [F(R_{t+1})]^{T+1}$. Of course, the distribution $F(\hat{\omega})$ in turn depends upon $F(R_{t+1})$ given Assumption 1.}

The risk associated with a portfolio estimator $\hat{\omega}$ is defined as the expected loss of the portfolio decision rule:

$$
Risk(\hat{\omega}) = E_{F(\hat{\omega}, \hat{R})} [L(\hat{R}, \hat{\omega})] = -E_{F(\hat{\omega}, \hat{R})} [U(W(R_T^f + \hat{\omega}))].
$$

(5)

so that the risk of a portfolio decision rule is simply the opposite of the expected utility associated with the decision rule. The notation in (5) emphasizes that the expectation that defines the risk of a portfolio decision rule is taken with respect to the joint distribution $F(\hat{\omega}, \hat{R})$.

For a given sample size, the risk of a portfolio decision rule provides a complete ordering over the set of all portfolio decision rules, since for any two arbitrary decision rules $\hat{\omega}_i$ and $\hat{\omega}_j$ we have either $Risk(\hat{\omega}_i) \leq Risk(\hat{\omega}_j)$ or vice versa. It is convenient and more natural from a finance perspective to simply work with the equivalent complete ordering based on the expected utility associated with the decision rule, which we denote $EU(\hat{\omega})$ under which $EU(\hat{\omega}_i) \geq EU(\hat{\omega}_j)$ if and only if $Risk(\hat{\omega}_i) \leq Risk(\hat{\omega}_j)$. From this point onward, we will focus on this expected utility in evaluating and comparing various estimators of the optimal allocation (portfolio decision rules).

The expected utility $EU(\hat{\omega})$ is defined for any portfolio decision rule, including decision rules based on parametric or nonparametric models of the optimal portfolio policy or portfolio decision rules based on estimation of the distribution of returns. Constant decision rules, such as an equal ($1/N$) allocation to each asset, may also be considered. From (5), it is clear that the expected utility (and risk) of a portfolio decision rule depends upon both the investor’s utility function $U(\bullet)$ that defines the underlying loss function along with the distribution of returns.\footnote{The dependence on the distribution of returns is both direct through $F(R_{t+1})$ and indirect through $F(\hat{\omega})$ which in turn depends on the distribution of the random sample, which is $[F(R_{t+1})]^T$ under Assumption 1.} It is important to emphasize that the expected utility associated with a decision rule depends on the sample size. For example, in comparing a misspecified parametric decision rule with a nonparametric decision rule (e.g., the estimator suggested by Brandt (1999)) it is possible that the parametric estimator achieves higher expected utility for small sample sizes due to the relatively large sampling error associated with the nonparametric estimator; while the nonparametric estimator achieves a higher expected utility for larger sample sizes due to its consistency. A final point worth emphasizing is that since $EU(\hat{\omega})$ depends upon the unknown $F(\hat{R})$ as well as upon $F_T(\hat{\omega})$, which is typically
unknown for finite samples unless \( \hat{\omega} \) is a trivial (constant) allocation rule, then \( EU(\hat{\omega}) \) is also unknown. However, it may be possible to estimate \( EU(\hat{\omega}) \).

For a given utility function and distribution of returns, the constant portfolio decision rule \( \hat{\omega} = \omega^* \) possess an optimality property that follows immediately from the definition of \( \omega^* \):

**Lemma 1** Given \( U(\bullet), F(R_{t+1}) \) and any portfolio decision rule \( \hat{\omega} \), \( EU(\hat{\omega}) \leq EU(\omega^*) \), where \( \omega^* \) denotes the constant allocation rule that sets \( \hat{\omega} = \omega^* \).

Of course, knowledge of \( \omega^* \) is equivalent to knowledge of all aspects of \( F(\hat{R}) \) relevant for optimal portfolio allocation. Hence, while the constant portfolio decision rule that sets the allocation to the optimal allocation is clearly ‘best,’ this decision rule is infeasible in practice.

## 2. Portfolios of Estimated Portfolios: Combining Multiple Estimators

In the context of the canonical portfolio choice problem described in the preceding section, suppose that \( J \) competing estimators of the optimal portfolio (portfolio decision rules) are available, indexed \( \hat{\omega}_j \) for \( j = 1, ..., J \). Let \( EU(\hat{\omega}_j) \), \( j = 1, ..., J \), index the expected utilities associated with the estimators under consideration. For the time being, we will abstract from the latent nature of \( EU(\hat{\omega}_j) \) and take these quantities as given.

Given a set of \( J \geq 1 \) possible estimators, it may seem desirable to simply adopt the decision rule that achieves the highest expected utility, i.e., to choose \( \hat{\omega}_j \) such that \( EU(\hat{\omega}_j) \geq EU(\hat{\omega}_k) \) for \( k = 1, ..., J \).\(^{10}\) Suppose; however, that instead of simply selecting a single estimator, we consider forming linear combinations of the underlying \( J \) estimators. Formally, let \( \hat{\omega}^C \) denote a linear combination of the underlying \( J \) portfolio decision rules:

\[
\hat{\omega}^C = \sum_{j=1}^{J} \alpha_j (R^T) \hat{\omega}_j (R^T) \tag{6}
\]

The notation in (6) emphasizes the fact that both \( \alpha_j \) and \( \hat{\omega}_j \) are permitted to be functions of the random sample of returns. This permits data-driven combinations of the base set of estimators. Of course, the general specification nests the special case of constant combining weights, such as equal weights, which do not depend on the data. It is straightforward to establish that a linear combination of portfolio decision rules is itself a portfolio decision rule, justifying the notation:

**Lemma 2** Let \( \hat{\omega}_j \) be portfolio decision rules and suppose that \( \alpha_j \) are measurable functions of \( R^T \) for \( j = 1, ..., J \) for \( J \geq 1 \). Then the \( \hat{\omega}^C = \sum_{j=1}^{J} \alpha_j (R^T) \hat{\omega}_j (R^T) \) is also a portfolio decision rule.

We now discuss how forming the combined estimator \( \hat{\omega}^C \) may be interpreted as forming a “portfolio of portfolio estimators.” Note first that the combining weights \( \alpha_j \) are unrestricted. and for this reason do not directly carry the interpretation of portfolio weights. It is possible to interpret

\(^{10}\)In the case that more than one of the decision rules satisfy this criterion, an arbitrary rule could be used to select a particular estimator.
the elements \( \alpha_j \) as portfolio weights, however, once we introduce a special, additional decision rule. Consider the constant portfolio decision rule \( \hat{\omega}_0 \) that allocates all wealth to the risk-free asset, so that \( \hat{\omega}_0 \) is simply an \( N \times 1 \) vector of zeros irrespective of the sample outcome. We have the following Lemma:

**Lemma 3** Assume the same conditions as in Lemma 2. Define \( \alpha_0(R^T) \equiv 1 - 1^T_j \alpha(R^T) \). Then the estimator \( \hat{\omega}^C = \alpha_0(R^T)\hat{\omega}_0 + \hat{\omega}^C \) is a portfolio decision rule such that \( \sum_0^J \alpha_j = 1 \).

Lemma 2 shows that the combined estimator \( \hat{\omega}^C \) constitutes a “portfolio of portfolio estimators,” where the \( \alpha_j \) represent the portfolio weights in each of the \( J \) underlying portfolio decision rules and \( \alpha_0 \equiv 1 - 1^T_j \alpha \) represents the weight allocated toward a \( J + 1 \)-th ‘ultraconservative’ portfolio decision rule that simply allocates 100% of wealth to the risk-free asset. We now consider the problem of determining the optimal linear combination of the base set of estimators, i.e., the optimal \( J \times 1 \) portfolio of portfolio estimators, denoted as \( \alpha^* \).

To begin to address this problem, note that since any portfolio of the base set of portfolio decision rules is again a portfolio decision rule (by Lemma 2), the combined portfolio decision rule has a risk associated with it. The optimal combination of estimators is then the combination achieving the lowest risk, i.e., the highest expected utility. It is convenient to collect the estimators \( \hat{\omega}_j \) into the \( N \times J \) matrix \( \hat{\Omega} \) as follows:

\[
\hat{\Omega} = \begin{bmatrix} \hat{\omega}_1 & \ldots & \hat{\omega}_J \end{bmatrix},
\]

such that \( \hat{\omega}^C = \hat{\Omega} \alpha \). The problem of forming the optimal (risk minimizing) combination of portfolio decision rules may therefore be stated as:

\[
= \max_{\alpha} E_{F(\hat{\omega}_1, \ldots, \hat{\omega}_J, \hat{R})} \left[ U \left( W \left( R_f^T + \hat{\Omega} \alpha \right) \left( \hat{R} - R_f 1_N \right) \right) \right]
\]

Given this statement of the combination problem, we could immediately proceed to characterize the optimal combining weights via first-order conditions. It is more instructive, however, to first reformulate the problem. In the next subsection, we consider a reformulation that will permit us to easily derive key properties of the optimal combining weights.

### 2.1. Viewing Portfolio Decision Rules as Financial Assets: A Useful Isometry

In this subsection we show that the problem of optimally combining a set of \( J \) portfolio decision rules is equivalent to determining the optimal portfolio allocation in a transformed version of the canonical asset allocation problem. This equivalence proves extremely convenient, as all of the results from standard portfolio optimization theory then apply immediately. This will allow us to easily state key results regarding the optimal combination of portfolio decision rules, as well as develop valuable intuition.

Consider an arbitrary portfolio decision rule \( \hat{\omega} \) and suppose that an investor allocates $1 of wealth according to this rule. This allocation is viewed as occurring *prior to* obtaining a realization.
of the random variable $R^T$, consistent with the pre-sample focus of classical econometrics. Given a realization of both the sample data $r^T$ and the $N \times 1$ vector of holding period returns $\tilde{r}$, the $\$1$ invested in $\tilde{\omega}$ pays an excess return equal to $\tilde{\omega}(r^T)'(\tilde{r} - R^f 1_N)$, where the notation $\tilde{\omega}(r^T)$ indicates the realized estimate of the optimal portfolio. We denote as $\tilde{R}^\omega$ the random excess return earned by a $\$1$ buy-and-hold investment in $\tilde{\omega}$ made prior to the realization of $R^T$ and $\tilde{R}$ (hence with a holding period of length $T + 1$).

Let $F(\tilde{R}^\omega)$ represent the distribution of this random excess return. It is important to note that this distribution is implicitly determined by both the distribution $F(\tilde{\omega})$ of the portfolio decision rule $\tilde{\omega}$ as well as the distribution of basic asset returns $F(\tilde{R})$.

As a simple example of the isometry between portfolio decision rules and financial assets, consider the constant portfolio decision rule $\tilde{\omega}_0$ that allocates all wealth to the risk-free asset. Since this portfolio decision rule is equivalent to a 100% position in the risk-free asset irrespective of the sample realization, $\$1$ invested in $\tilde{\omega}_0$ earns a gross return of exactly $R^f$ with no uncertainty. Notice that in this special case, the distribution of the return on a position in the decision rule coincides precisely with the distribution of one of the underlying assets (the risk-free asset); however, this is certainly not the case in general.

We now reformulate the problem (8) using the equivalence between portfolio decision rules and financial assets with a transformed return distribution. Given $J$ portfolio decision rules, define $J$ “financial assets” where the $j$-th asset pays an excess return equal to $\tilde{R}^\omega_j \equiv \tilde{\omega}_j(\tilde{R} - R^f 1_N)$ for $j = 1, \ldots, J$. We collect these excess returns into a $J \times 1$ vector $\tilde{R}$ with joint distribution $F(\tilde{R})$. In addition, assume that a risk-free asset exists with gross return $R^f$ and consider the optimal asset allocation in this ‘transformed’ asset space. Intuitively, one might imagine that the investor has the opportunity to invest funds with $J$ different fund managers, where each manager will invest funds across the same set of $N$ assets and each manager will employ a strategy in accordance with a particular estimator $\tilde{\omega}_j$ of the optimal portfolio decision rule applied to the random sample $R^T$.

Finally, an additional ultraconservative manager is available who will simply allocate all wealth to the risk-free asset. The question of interest is: What is the optimal allocation of initial wealth among the various fund managers?

The following proposition, proved in the appendix, states that the problem of forming the optimal weighted combination of the set of portfolio decision rules is formally equivalent to the canonical asset allocation problem in the transformed space of assets:

**Proposition 4** Let $\tilde{\omega}_j$ be portfolio decision rules for $j = 1, \ldots, J$ and consider the problem described

---

11Since holding period returns in the original asset space is decorated with a tilda, the decoration emphasizes that the holding period for the transformed asset also includes the sample period, i.e. $T + 1$ periods.

12Of course, the distribution $F(\tilde{\omega})$ in turn depends upon $F(\tilde{R})$ given our i.i.d. assumption.
in (8). This problem is equivalent to the canonical asset portfolio allocation problem where the investment options include a risk free asset paying gross return \( R^f \) and \( J \) risky assets paying excess returns equal to \( \tilde{R} \) with joint distribution \( F(\tilde{R}) \).

While Proposition 1 is very simple, it is also powerful since all of the well-developed theoretical tools from financial theory may now be applied to derive insight regarding the optimal ‘portfolio of estimated portfolios.’ Indeed, the equivalence between the canonical asset allocation problem and the problem of forming a weighted combination of portfolio decision rules illustrates why combining portfolio decision rules may lead to improved portfolio performance as measured by the expected utility associated with the portfolio strategy. Combining portfolio decision rules is likely to achieve improved portfolio performance for precisely the same reason that investing in multiple assets typically improves portfolio performance in standard asset allocation problems: benefits from diversification.

Let \( e_j \) represent the \( N \times 1 \) vector with a one in the \( j \)-th position and zeros elsewhere. In general, it is possible that the optimal combination of portfolio decision rules will be such that \( \alpha^* = e_j \) for some \( j = 1, \ldots, J \) so that it is optimal to invest 100% in the \( j \)-th underlying portfolio decision rules. Such an outcome is equivalent to a portfolio problem with \( J \) risky assets in which no benefits to diversification are possible. For reasonably well-chosen decision rules this seems unlikely in practice.

Proposition 1 also suggests a pitfall in combining estimators. Since the optimal combination problem is isometric to a portfolio allocation problem, it becomes clear that learning about the optimal combining weights (estimating these weights) will be subject to the same empirical difficulties faced in the [Estimation uncertainty intuition...connect with next two sections...]

The following subsections provide various results and some discussion. Indeed, the following section of the paper is devoted to characterizing various properties of the optimal combination of portfolio decision rules and each result follows in straightforward fashion from classic results in financial economics.

3. **Optimal Portfolios of Estimated Portfolios: Results and Discussion**

This section of the paper applies Proposition 1 to characterize the optimal combination of a set of basis estimators in various portfolio allocation settings. We begin with the general problem addressed in (3) and then turn to consider situations in which investor preferences may be expressed as a function of moments of excess returns. Finally, we discuss a special case of the general combination problem which entails the construction of a shrinkage estimator for a given estimator such as the standard plug-in estimator in mean-variance portfolio theory.

3.1. **Combination of Portfolio Decision Rules in a General Setting**

In the fairly general setting characterized by equation (3) one may seek the optimal combination of \( J \) portfolio decision rules when investor preferences are represented by an arbitrary differentiable, strictly concave utility function. By Proposition 1, seeking the optimal combination of \( J \) decision
rules is equivalent to the canonical portfolio allocation problem when \( J \) risky assets are available paying excess returns \( \tilde{R} \). The optimal solution to the combination problem is implicitly determined by the first-order conditions for the equivalent allocation problem:

\[
\alpha^* \equiv \alpha \text{ s.t. } E_{F(\tilde{R})} \left[ U(W[R_f + \alpha' \tilde{R}]) \tilde{R} \right] = 0. 
\] (9)

Beyond this characterization of the optimal portfolio of estimated portfolios, few explicit results are available without additional assumptions regarding, e.g., the distributional form of asset returns. In general, the optimal portfolio of estimated portfolios will depend upon the distributional properties of excess returns and of the portfolio decision rules, wealth and upon parameters characterizing investor preferences. It is well-known that optimal portfolio weights do not depend on wealth when utility displays constant relative risk aversion (CRRA) Applying Proposition 1 gives rise to the following analogous result for the optimal combination of a set of portfolio decision rules:

\textbf{Lemma 5} Consider the problem (8) of determining the optimal combination of \( J \) portfolio decision rules. When the utility function exhibits constant relative risk aversion (CRRA), then the optimal combining weights \( \alpha^* \) are independent of initial wealth \( W \).

Except in special cases, the solution to (9) does not have a closed-form and numerical techniques must therefore be used to determine the optimal portfolio combination. Tauchen and Hussey (1991) suggest applying quadrature methods to solve this type of problem. When \( J \) is of very low dimension, this may be feasible. As noted by Jondeau and Rockinger (2004, 2005), when asset returns are fat-tailed and asymmetric the required number of quadrature points increases exponentially with the number of assets (estimators) considered. This leads to interest in moment-based approximations to expected utility. The simplest such case is the mean-variance setting, to which we turn next.

3.2. The Mean-Variance Setting

The mean-variance approach to portfolio allocation remains dominant in practice, particularly when a reasonably large number of assets is entertained. Let \( \mu \) and \( \Sigma \) represent the mean and covariance matrix of excess risky returns, respectively. In contrast to the preceding subsection, we work with centralized moments and assume that investor preferences may be represented in terms of the mean and variance of excess returns on the portfolio, i.e., as

\[
U(\omega) = \omega' \mu - \frac{\gamma}{2} \omega' \Sigma \omega 
\] (10)

where \( \gamma > 0 \) captures the risk-aversion of the investor. Conditions under which (10) is (fully) consistent with expected utility maximization include quadratic utility or multivariate normal returns distributions (see, e.g., Ingersoll (1987)). The investor’s allocation problem may be stated as

\[
\max_{\omega} \omega' \mu - \frac{\gamma}{2} \omega' \Sigma \omega, 
\]
with the well-known solution derived by Markowitz (1952, 1959):

$$
\omega^* = \frac{1}{\gamma} \mu \Sigma^{-1}. \quad (11)
$$

In practice, \(\mu\) and \(\Sigma\) are unknown and hence \(\omega^*\) must be estimated. Consider a set of \(J \geq 1\) estimators of the optimal portfolio, indexed as \(\hat{\omega}_j\). By application of Proposition 1, all of the tools of standard mean-variance portfolio analysis may be applied to understand the optimal combination of the set of portfolio estimators. In the space of transformed asset returns, where the random variation of portfolio returns includes both random variation in the estimation data and random variation in the holding period return, the \(J\) portfolio decision rules will generate a standard mean-variance efficient frontier. This again gives rise to intuition for why combining multiple portfolio estimators is likely to be beneficial, since the basis set of portfolio decision rules will typically lie inside the hyperbola that characterizes the efficient frontier so that the same expected return (in the transformed asset space) may be achieved at lower variance by forming a nontrivial portfolio of estimated portfolios.

A global minimum variance portfolio will exist, which represents the portfolio of the \(J\) portfolio estimators that yields the minimum possible variance of transformed asset returns. Similarly, a tangency portfolio of portfolio estimators will exist. Since a risk-free asset is available in the transformed portfolio problem, the minimum variance set is characterized in mean-standard deviation space as the usual pair of rays with common intercept at \(R_f\). A type of mutual-fund result obtains such that optimal combining weights for the \(J\) portfolio decision rules are proportional to the weights characterizing the tangency portfolio and hence the relative weights on each estimator do not depend on risk preferences. Finally, it is straightforward to analytically characterize the optimal linear combination of the set of portfolio decision rules as the formula is of the same form as (11), except that the moments refer to asset returns in the transformed asset space. The following Lemma summarizes:

**Lemma 6** Suppose investor preferences are represented by (10) and let \(\hat{\omega}_j\) be portfolio decision rules for \(j = 1, \ldots, J\) and assume that none of these portfolio decision rules is the ultraconservative decision rule \(\hat{\omega}_0\). Define \(\mu_R \equiv E[\bar{R}]\) and \(\Sigma_R \equiv E[(\bar{R} - \mu_R)(\bar{R} - \mu_R)^T]\). Under the assumption that \(\Sigma_R\) is nonsingular, then we have the following results:

1. (Minimum Variance Portfolio of Estimated Portfolios) The combining weights for the \(J\) portfolio decision rules that leads to the minimum variance of portfolio returns in the transformed asset space is given by

$$
\alpha^{GMV} = \frac{\Sigma_R^{-1}1_J}{1_J^T\Sigma_R^{-1}1_J}.
$$

2. (Tangency Portfolio of Estimated Portfolios) The combining weights for the \(J\) portfolio decision rules that characterizes the tangency point between the minimum variance set and the risky-asset only minimum variance set in the transformed asset space are given by

$$
\alpha^{TGCY} = \left(\frac{1}{1_J^T\Sigma_R^{-1}\mu_R}\right)\Sigma_R^{-1}\mu_R.
$$
3. (Optimal Combining Weights) The formula for the optimal linear combination of the \( J \) portfolio decision rules is given by
\[
\alpha^* = \frac{1}{\gamma} \Sigma_R^{-1} \mu_R.
\] (12)

4. (Mutual Fund Theorem) The optimal linear combination of the \( J \) portfolio decision rules is proportional to the combining weights characterizing the tangency portfolio in the transformed asset space. Thus, the relative proportions allocated to each estimator in forming the optimal portfolio of estimated portfolios are independent of risk preferences.

The condition that \( \Sigma_R \) is nonsingular amounts to requiring that no redundant assets exist among the \( J \) assets in the transformed space. An example of a violation of this condition would be if two estimators \( \hat{\omega} \) and \( c\hat{\omega} \) with \( c \neq 0 \) are included in the combination problem. Since these estimators are proportional, the inclusion of both creates a redundant asset in the transformed asset space and \( \Sigma_R \) will fail to be invertible.

The formula for the optimal combination in Lemma 4.3 is expressed in terms of the first and second moments of returns on assets in the transformed asset space. These moments implicitly involve moments of the excess returns on the original assets and of the joint distribution of the portfolio decision rules. It is possible to explicitly characterize the optimal combination \( \alpha^* \) in terms of these fundamental moments. In order to do so, we again collect the estimators \( \hat{\omega}_j \) into the \( N \times J \) matrix \( \hat{\Omega} \) of (7).

Now, define \( \mu_{\hat{\Omega}} = E[\hat{\Omega}] \) with \( j \)-th column equal to \( \mu_{\hat{\omega}_j} = E(\hat{\omega}_j) \). Let \( \sigma_{i,j} = \Sigma_{i,j} \), the covariance of excess returns on assets \( i \) and \( j \) in the original asset space. Similarly, let \( \sigma_{i,j}^{\omega_l,k} \) for \( i, j \leq N \) and \( l, k \leq J \) denote the covariance of the estimated portfolio weights on the \( i \)-th and \( j \)-th assets (in the original asset space) for the pair of portfolio decision rules \( \hat{\omega}_l \) and \( \hat{\omega}_k \). In the case where \( l = k \), \( \sigma_{i,j}^{\omega_l,k} \) is simply the \( i,j \)-th element of \( \Sigma_{\hat{\omega}_l} \), the covariance matrix of portfolio decision rule \( \hat{\omega}_l \). When \( l \neq k \), \( \sigma_{i,j}^{\omega_l,k} \) represents a cross-covariance between elements of \( \hat{\omega}_l \) and \( \hat{\omega}_k \), i.e., the covariance between \( \hat{\omega}_{l,i} \) and \( \hat{\omega}_{l,j} \). With this notation in place, we have the following proposition:

**Proposition 7** The moments mean \( \mu_{\hat{\Omega}} \) and covariance matrix \( \Sigma_{\hat{\Omega}} \) of excess returns in the transformed asset space may be expressed as:

\[
\mu_{\hat{\Omega}} = \begin{bmatrix}
\mu_{\hat{\omega}_1} \\
\vdots \\
\mu_{\hat{\omega}_J}
\end{bmatrix} = \mu_{\hat{\Omega}}^T \mu
\]

and:

\[
(\Sigma_{\hat{\Omega}})_{i,j} = \mu_{\hat{\omega}_i}^T \Sigma \mu_{\hat{\omega}_j} + \sum_{m,n<N} \sigma_{ij} \sigma_{mn} \hat{\omega}_m \hat{\omega}_n \] (13)

Proposition 2 illustrates that the optimal combination of portfolio decision rules in the mean-variance setting depends upon all of the following in addition to investor risk preferences: 1.) the
expected excess return and covariances of assets in the original asset space; 2.) the expected portfolio weights for each of the portfolio decision rules; 3.) the covariances of estimated portfolio weights among pairs of assets for each portfolio decision rule; and 4.) the covariances of estimated portfolio weights across pairs of portfolio decision rules.

3.3. Higher Moments and Optimal Combination

While the bulk of asset allocation literature considers the well-known and tractable case of mean-variance preferences, higher moments and co-moments might also be important to investors. Recent papers that consider higher order moment-based preferences include Ang and Bekaert (2002), Harvey, Liechty, Liechty and Muller (2004), Guidolin and Timmermann (2005) and Jondeau and Rockinger (2005, 2006). Proposition 1 may be applied to characterize the optimal combination of a basis set of estimators under preference functions that incorporate higher moments.

Here we briefly consider the specific case of a moment-based preference function that incorporates skewness and kurtosis. Suppose that investor preference are represented by a moment-functional of the form

\[ EU(\tilde{W}) = \omega' \mu - \frac{\kappa_1}{2} \omega' \Sigma \omega + \frac{\kappa_2}{3} \omega' S(\omega \otimes \omega) - \frac{\kappa_3}{4} \omega' K(\omega \otimes \omega \otimes \omega), \]

where \( S \) and \( K \) indicate the co-skewness and co-kurtosis matrices of excess returns, respectively, while \( \kappa_1, \kappa_2 \) and \( \kappa_3 \) are positive constants implying that the investor dislikes portfolio variance, likes positive skewness and dislikes kurtosis in portfolio returns.\(^{13}\) Given a basis set of portfolio decision rules, Proposition 1 implies that the optimal combining weights may be characterized by the first-order condition

\[ 0 = \tilde{\mu} + \kappa_1 \tilde{\Sigma} \alpha + \kappa_2 \tilde{S}(\alpha \otimes \alpha) + \kappa_3 \tilde{K}(\alpha \otimes \alpha \otimes \alpha), \quad (14) \]

where the notation emphasizes that the return moments now refer to excess returns on the transformed set of assets. As Jondeau and Rockinger (2005, 2006) point out, the equations (14) may be solved numerically with a lower difficulty relative to the quadrature-based integration of the utility function in order to determine the optimal combination of a set of basis estimators under four-moment preferences. Of course, the approach sketched here could be extended to even higher moments if desired.

3.4. Optimal Shrinkage Toward the Ultraconservative Portfolio Decision Rule

The general theory permits linear combinations of any finite number of estimators. The case of a single portfolio decision rule is of particular interest. In such a case, the linear combination simply amounts to a scaling of a given portfolio estimator \( \tilde{\omega} \) by the factor \( \alpha \) so that \( \tilde{\omega}^C = \alpha \tilde{\omega} \). An

\(^{13}\)The constants may be calibrated to approximate a particular differentiable concave utility function through application of a Taylor series expansion. See, e.g., Jondeau and Rockinger (2006) and Guidolin and Timmermann (2005) for examples of this approach.
alternative interpretation is that the econometrician forms a portfolio consisting of the estimator \( \hat{\omega} \) and the estimator \( \hat{\omega}_0 \), the estimator that simply allocates 100% of wealth to the risk-free asset. For the case where \( 0 \leq \alpha \leq 1 \), the combined estimator takes the form of a James-Stein shrinkage estimator that shrinks toward the portfolio decision rule \( \hat{\omega}_0 \). The degree of shrinkage is of course governed by \( \alpha \) with \( \alpha = 1 \) representing no shrinkage and \( \alpha = 0 \) representing complete shrinkage to a 100% position in the risk-free asset. We now present results for the special case where an arbitrary estimator is re-scaled by the factor \( \alpha \). In what follows we will refer to such an estimator as a “shrinkage” estimator and to \( \alpha \) as a “shrinkage factor” even though \( \alpha \) is not restricted to the interval \([0, 1]\).

The equivalence established in Proposition 1 illustrates that the optimal shrinkage factor for an arbitrary portfolio decision rule \( \hat{\omega} \) is equivalent to the optimal allocation to the ‘asset’ that pays of excess returns \( \hat{R}_\omega \). Let \( \mu \) represent the \( N \times 1 \) vector of excess returns on the risky assets and \( \mu_{\hat{\omega}} = E[\hat{\omega}] \) denote the expectation of the weights portfolio weights corresponding to an arbitrary portfolio decision rule. In an asset allocation problem with a risk-free asset and a single risky asset, it is well-known that the sign of the the allocation to the risky asset is determined by the sign of the risk premium on the risky asset (see, e.g., Ingersoll (1987)). The direct analog for the shrinkage problem is the following:

**Lemma 8** Let be an arbitrary portfolio decision rule and assume that \( U(\cdot) \) is strictly concave. Then the optimal shrinkage factor \( \alpha^* \) applied to \( \hat{\omega} \) is greater than, less than, or equal to zero iff \( (\mu_{\hat{\omega}})' \mu \) is greater than, less than, or equal to zero, respectively.

The lemma establishes that the, for a given vector \( \mu \) of risk-premia on the risky assets, the sign of \( \alpha^* \) is driven by the bias \( \mu_{\hat{\omega}} - \omega^* \). One implication of the lemma is that a portfolio decision rule may be sufficiently biased so that complete shrinkage is optimal. This occurs if \( (\mu_{\hat{\omega}})' \mu \) is exactly zero. By the optimality of \( \omega^* \), \( \mu^T\omega^* \geq 0 \) and will be strictly greater than zero so long as the optimal allocation includes some position in the risky assets. This yields the following corollary:

**Corollary 9** Suppose that the optimal \( \omega^* \) is nonzero so that \( \mu^T\omega^* > 0 \). If \( \hat{\omega} \) is unbiased for the optimal allocation such that \( E(\hat{\omega}) = \omega^\ast \) then \( \alpha^* > 0 \).

The corollary implies that when a portfolio decision rule is unbiased, complete shrinkage is never optimal, irrespective of both the degree of risk aversion of the investor as well as the variance, kurtosis, etc. that the estimator \( \hat{\omega} \) may exhibit in small samples. Of course, the optimal degree of shrinkage may be very high so that complete shrinkage is nearly optimal.

In addition to the above characterization of the sign of the optimal shrinkage factor, standard results from the canonical asset allocation problem with a single risky asset characterize the relation between the optimal shrinking factor and initial wealth, summarized in the following Lemma:

**Lemma 10** Let be an arbitrary portfolio decision rule and assume that \( U(\cdot) \) is twice differentiable and strictly concave. Then \( \frac{\partial \alpha^*}{\partial W} \) is, \( >, =, < 0 \) under decreasing relative risk aversion (DRRA), constant relative risk aversion (CRRA) and increasing relative risk aversion (IRRA), respectively.
3.4.1. Shrinking a Portfolio Decision Rule in the Mean-Variance Setting

Consider an arbitrary portfolio decision rule \( \hat{\omega} \) in the context of the mean-variance problem discussed previously. Let \( \hat{R} = \hat{\omega}'(\hat{R} - R_f 1_N) \) denote the excess return of $1 invested in \( \hat{\omega} \), viewed from a pre-sample perspective and consider the problem of optimal shrinkage of \( \hat{\omega} \) toward \( \hat{\omega}_0 \), i.e., shrinkage toward a 100% risk-free portfolio strategy. In the transformed asset space the problem is equivalent to a mean-variance asset allocation problem with a risk-free asset paying gross return \( R_f \) and a single risky asset with excess return \( \hat{R} \). Letting the expectation and variance of the excess return on this risky asset be \( \mu_{\hat{R}} \) and \( \sigma_{\hat{R}}^2 \), respectively, the formula for the optimal shrinkage factor follows immediately as a special case of Lemma 4.3:

$$\alpha^* = \frac{\mu_{\hat{R}}}{\gamma \sigma_{\hat{R}}^2}$$  \hspace{1cm} (15)

Several remarks on (15) are in order. First, note that \( \alpha^* \) clearly satisfies Lemma 6 and Corollary 1, as \( \alpha^* \) is positive if and only if \( \mu_{\hat{R}} \) is positive. Second, the investor’s risk aversion parameter \( \gamma \) enters the formula for \( \alpha^* \) in an intuitive way: higher investor risk aversion (higher \( \gamma \)) leads to a lower \( \alpha^* \) and consequently stronger shrinkage toward the ultraconservative portfolio decision rule.

Kan and Zhou (2007) provide analytical results for the optimal choice of shrinking factor \( \alpha \) for the plug-in estimator as a function of \( N, T \) and the mean and variance of excess returns when returns follow the multivariate normal distribution. Their analytic expression for the optimal shrinking factor is

$$\alpha^* = \left[ \frac{(T - N - 1)(T - N - 4)}{(T - 2)(T - N - 2)} \right] \left( \frac{\mu' \Sigma^{-1} \mu}{\mu' \Sigma^{-1} \mu + \frac{N}{T}} \right).$$  \hspace{1cm} (16)

Obviously, (15) is identical to (16 ) when returns are normally distributed and \( \hat{\omega} \) is the plug-in estimator. The result in (15) provides for an optimal “two-fund rule” for any choice of estimator \( \hat{\omega} \) (so long as first and second moments exist) and for potentially non-normally distributed returns.

4. Feasible Approaches to Forming Portfolios of Estimated Portfolios

The results of the preceding section demonstrate that, in general, it is possible to combine multiple portfolio decision rules, i.e., to form a portfolio of estimated portfolios, in such a manner that the combined estimator outperforms any of the individual estimators based on achieved expected utility. A significant drawback; however, is the fact that the optimal combination depends on the underlying distribution of asset returns and as such is unknown, rendering the optimal combination infeasible. In this section of the paper we discuss feasible approaches to combining portfolio decision rules and describe a general approach for consistent estimation of optimal combining weights.

4.1. The Estimation Problem and Alternative Approaches

Since the problem of finding the optimal combining weights is equivalent to the canonical asset allocation problem in a transformed asset space, estimating the optimal set of combining weights
is equivalent to the problem of estimating optimal portfolio weights (over the transformed assets). An immediate difficulty; however, arises from the fact that the transformed assets in Proposition 1 entail a conceptual position held from prior to the realization of the historical sample data through the end of the sample period as well as over the next (upcoming) period, i.e., a total holding length of \( T + 1 \) periods. In a typical application, monthly historical returns on assets of interest might be available over a period of approximately 30 years, hence 360 total historical returns. In the absence of e.g., structural breaks (ruled out under the i.i.d. assumption maintained in this paper), we would like to base estimation for each of the portfolio decision rules under consideration on as large a sample as possible. Assuming that we are willing to sacrifice the most recent historical observation as a pseudo-holding period return (implying a sample size of \( T = 359 \) allocated to each portfolio decision rule), we have available only a single realization of returns \( \hat{R} \) in the transformed asset space. This presents a serious obstacle to the notion of estimating \( \alpha^* \) using standard estimators of optimal portfolio weights, since estimation would be limited to a sample size of one.

Kan and Zhou (2007) consider this problem and derive estimators of the optimal combination for two special cases of interest: 1.) a shrinkage factor for the standard ‘plug-in’ estimator in the mean-variance problem and 2.) the problem of optimally combining the plug-in estimators of the tangency portfolio and the global minimum variance portfolio. Kan and Zhou make use of the assumption that returns are multivariate normal to derive expressions for the optimal combining weights as a function of the sample size, number of assets and mean and variance of returns. In the case of shrinking the plug-in estimator of the tangency portfolio, this expression is (16) and Kan and Zhou (2007) suggest a feasible estimator that is based on an adjusted plug-in estimate of the key term \( \mu'\Sigma^{-1}\mu \) (see Kan and Zhou (2007) for details). An estimator for the optimal combination of the plug-in estimators of the tangency portfolio and the GMV portfolio is constructed in similar fashion.

This paper suggests an alternative approach to estimating the optimal combination of a set of basis estimators that delivers consistent estimates in a general setting. Specifically, the method outlined below does not require distributional assumptions. Furthermore, and perhaps more importantly, the approach allows the combination of an arbitrary set of basis estimators. The approach hinges on the observation that, under the i.i.d. assumption, all random variation in \( \hat{R} \) is ultimately governed by the underlying joint distribution of asset returns \( F'(\hat{R}) \). Although only a single historical realization of \( \hat{R} \) is available, a reasonably long historical sample of asset returns will be available, permitting estimation of \( F'(\hat{R}) \). Now suppose that an estimate \( \hat{F}'(\hat{R}) \) of \( F'(\hat{R}) \) is available. One approach to estimating \( \alpha^* \) is to ‘plug-in’ the estimate \( \hat{F}'(\hat{R}) \) in place of \( F'(\hat{R}) \) and use this distribution to simulate a large ‘pseudo-sample’ of returns \( \hat{R}_F \), where the notation emphasizes that these simulated returns are not drawn from the true distribution \( F'(\hat{R}) \) but rather from an approximation built-up from \( \hat{F}'(\hat{R}) \).

Suppose a consistent estimator of the optimal allocation is available (e.g., the Brandt (1999) estimator that is consistent for the optimal portfolio weights without requiring a distributional assumption). For sufficiently large pseudo samples \( \hat{R}_F \), the estimator will produce very accurate
estimates of the portfolio weights $\alpha^*_F$, which are optimal when asset returns (in the original asset space) are distributed according to $\hat{F}(\bar{R})$. Of course, $\hat{F}(\bar{R}) \neq F(\bar{R})$, $\alpha^*_F \neq \alpha^*$ so that the truly optimal combining weights are not recovered. Intuitively, the hope is that if $\hat{F}(\bar{R})$ is a reasonably good estimate of $F(\bar{R})$, then $\alpha^*_F$ will provide a reasonably good estimate of $\alpha^*$.

By this point the reader has likely noticed a ‘matryoshka doll’ flavor to this approach to feasible estimation of the optimal combining weights. In particular, we arrive at a new version of the original problem, estimating optimal portfolio weights for assets. The results presented in Section 3 of the paper suggest that in such circumstances combining multiple estimators or shrinking an estimator may improve estimation performance. Should we then seek to shrink a particular estimator or combine multiple estimators in estimating the optimal combining weights?

To address this question, first note that since $\hat{R}_F$ can be simulated, estimation error with respect to $\alpha^*_F$ can be made negligible through a sufficiently large pseudo-sample size. So, to the extent that shrinking or combination is entertained, it should be entertained with respect to $\hat{F}(\bar{R})$. In principle, nothing rules out a second level of shrinking or a combination approach to estimating $F(\bar{R})$; a second layer of the problem of estimating the optimal shrinking rate or combining weights immediately arises. Clearly, this type of iteration could proceed ad infinitum. In practice, one must ultimately stop at some level and simply employ a particular estimator. In current practice this is typically done at the original estimation stage. When estimators are combined, this occurs at a second level of estimation, where combining weights or shrinkage factors are estimated.

4.2. Specific Implementations of the Simulation Approach

Having presented a general strategy for estimating optimal combining weights, we now discuss specific implementation strategies. The estimation procedure may be viewed as a series of three steps: first obtain an estimate of $\hat{F}(\bar{R})$ of $F(\bar{R})$, then simulate a pseudo sample of returns $\bar{r}^{T^*}$, where $T^*$ indicates the size of the pseudo sample, and finally estimate the optimal portfolio allocation in the transformed asset space using the pseudo sample of returns. We briefly discuss each of these steps and then discuss an alternative approach that avoids estimating the optimal weights altogether.

4.2.1. Estimating the Optimal Weights in the Transformed Asset Space

Suppose that a pseudo sample of returns $\bar{r}^{T^*}$ is available. In the mean-variance setting (12) provides an analytic formula for the optimal combining weights as a function of the first and second moments of $\bar{R}$. The optimal combining weights may therefore be estimated using plug-in estimates of the moments $\mu_{\bar{R}}$ and $\Sigma_{\bar{R}}$ based on $\bar{r}^{T^*}$. In the case of shrinking a portfolio decision rule in the mean-variance setting, for example, one simply needs to compute the mean and variance of $\bar{r}^{T^*}$ and then form $\hat{\alpha}^*$ by plugging these into (15). Note that since the size $T^*$ of the pseudo sample is arbitrary, this may be set to a sufficiently high number so that the plug-in moments are estimated with a very high degree of accuracy.

In more general cases, an analytic formula for the optimal combining weights as a function of moments of $\bar{R}$ is typically not available. Under four-moment preferences the optimal combining
weights may be estimated using (14) and the corresponding moment of simulated returns. For investor preferences are represented by a differentiable concave utility function, the first order condition in (9) characterizes the optimal combination and provides an orthogonality condition that may be used as the basis for estimation by method of moments. This is exactly an application of the nonparametric estimator of optimal portfolio weights proposed by Brandt (1999) specialized to the case of i.i.d. asset returns. The estimator takes the form:

\[ \hat{\alpha} = \left\{ \alpha : \frac{1}{T^*} \sum_{t=1}^{T^*} U_t (W[R_f + \alpha'R]) \tilde{R} = 0 \right\} \]  

(17)

This estimator carries the interpretation of an M-estimator that maximizes in-sample investor utility. In general, one might be concerned regarding the finite sample performance of (17), particularly for relatively large \( J \). In the present case, however, estimation is based on a simulated pseudo sample and \( T^* \) may be set to as large a value as desired, so that extremely precise estimates of \( \alpha_{F^*} \) are obtained. One way to view the estimator (17) in the present application is simply as a computational device. The distribution \( \hat{F}(\tilde{R}) \) is taken as the true distribution of asset returns in the transformed space and (17) is used to compute the optimal portfolio weights under this distribution. It is important to keep in mind that what is recovered with extremely high precision is \( \hat{F}(\tilde{R}) \) and not \( F(\tilde{R}) \), where the discrepancy arises from the fact that the pseudo sample \( \tilde{r}^{T^*} \) is generated using \( \hat{F}(\tilde{R}) \) rather than \( F(\tilde{R}) \).

4.2.2. Estimating \( F(\tilde{R}) \)

Generating a pseudo sample of returns \( \tilde{r}^{T^*} \) first requires an estimate of \( F(\tilde{R}) \). One candidate for \( \hat{F}(\tilde{R}) \) is the empirical distribution function (EDF) of historical asset returns. This amounts to a nonparametric approach to estimating \( F(\tilde{R}) \). Since any parametric model of the distribution of asset returns may be subject to misspecification of unknown form, an appealing aspect of basing \( \hat{F}(\tilde{R}) \) on the EDF is that no distributional model is required. Furthermore, the EDF is a consistent estimator for the true distribution function of returns under mild conditions. The primary drawback of basing \( \hat{F}(\tilde{R}) \) on the empirical distribution involves the curse of dimensionality. When \( \hat{F}(\tilde{R}) \) is based on the EDF, pseudo samples \( \tilde{r}^{T^*} \) may be easily generated using the i.i.d. bootstrap. More specifically, a bootstrap sample \((r^T, r)_b^T \) of size \( T + 1 \) (using sampling with replacement) from the historical asset returns is drawn and estimates of the optimal portfolio weights are computed using the first \( T \) observations of the bootstrap sample as \( \omega_j(r^T_b) \) for \( j = 1, \ldots, J \). Finally, these weights are applied to the \( T + 1 \)-th bootstrap observation, which is treated as a holding period return and this yields a set of pseudo returns for each of the decision rules.

As an alternative to employing the EDF of asset returns, \( \hat{F}(\tilde{R}) \) may be based on a parametric model such as the multivariate normal, multivariate student-\( t \), or multivariate skewed student-\( t \) distributions. Posing a parametric model for asset returns helps to mitigate the problem of the curse of dimensionality that plagues the EDF estimator. On the other hand, a parametric model may be misspecified in important ways. Recall that the asset returns in the transformed asset
space depend on the distribution of the underlying portfolio estimators \( \hat{\omega}_j \). Even if such estimators are asymptotically normal, finite sample distributions may deviate substantially from the normal, leading to \( \hat{R} \) that are not well-modeled by, e.g., the multivariate normal or multivariate student-\( t \).

Even if a parametric model for \( F(R) \) is correctly specified, key parameters such as the covariance matrix may be estimated with substantial imprecision when the dimensionality is large. To address this problem, a factor model for the covariance matrix may be specified, or a shrinkage estimator as proposed by Ledoit and Wolf (2003) might be employed. Once a parametric model for \( \hat{F}(\hat{R}) \) is estimated, a pseudo sample of returns is generated by simulating random samples of size \( T + 1 \) from the fitted distribution and proceeding as in the EDF case previously discussed.

4.2.3. **Ad Hoc Approaches to Forming Portfolios of Estimated Portfolios**

Proposition 1 illustrates that recovering the optimal combining weights is tantamount to estimating optimal portfolio weights. Given the well-known problems that plague estimation of portfolio weights, one might entertain ad hoc approaches to combination. The primary motivation for such approaches is the limited ability to obtain a precise estimate of \( F(\hat{R}) \) for a typical application with a relatively large set of assets and limited historical data. If the estimator \( \hat{F}(\hat{R}) \) is extremely noisy, this will result in a large degree of statistical uncertainty regarding the optimal combination of the portfolio decision rules.

Two specific types of ad hoc rules are of particular interest. The first approach simply averages the various decision rules so that \( \hat{\alpha} = (1/J)1_J \). Simply averaging the various portfolio decision rules under consideration may be viewed as an extreme type of shrinkage estimator (for the optimal combining weights) where shrinkage to the \( 1/J \) weighting is complete. Motivation for this approach may be drawn from the large literature on forecast combination methods and the empirical observation that naive equal weighting of forecasts frequently seems to outperform more sophisticated combining techniques (see, e.g., Timmermann (2006)). The second ad hoc rule is to allocate \( (1/N\gamma) \) to each risky asset. Setting \( J = N \) and defining each of the basis estimators as \( \hat{\omega}_j = (1/\gamma)e_j \), where \( e_j \) is the \( N \times 1 \) vector with a one in the \( j \)-th position and zeros elsewhere, illustrates that this rule is actually a special case of simple averaging. This estimator might be viewed as extreme shrinkage, since the portfolio weights do not depend at all on the data. Finally, note that in the case where \( \gamma = 1 \) the estimator is the \( (1/N) \) rule studied by DeMiguel, Garlappi, and Uppal (2007).

5. **Empirical and Simulation Evidence**

This section explores the empirical properties of the combination approach to estimating optimal portfolio weights. In the present paper, we focus on a particularly simple application: an asset allocation problem with a single risky equity index and a risk-free rate. The simplicity of the application allows us to provide useful intuition, via graphical plots, regarding the behavior of combined (and in particular shrinkage) estimators.
5.1. **Data and Benchmark Models**

Monthly excess returns data were collected for the MSCI US Equity Index over the sample period 1970:1 - 2007:1 for a total of 445 available observations. Table 1 presents summary statistics for these US excess returns. As has been well documented in existing literature, the US excess returns exhibit excess kurtosis relative to the normal as well as some negative skewness. Figure 1 presents density plots for the US excess returns series based on the normal and skew student-t distributional fits along with a nonparametric density estimate.

5.2. **Simulation Analysis**

In a series of Monte Carlo simulation experiments we analyze the behavior of combined approaches to estimating the optimal portfolio allocation in a portfolio problem with a single risky stock index and a risk-free asset. Excess returns on the risky stock index are simulated using parameter estimates from the normal distribution with mean and variance calibrated to match the full sample estimates reported in Table 1. Suppose that the investor exhibits mean-variance preferences as in (10). We focus on the behavior of the standard ‘plug-in’ estimator

\[ \hat{\omega}_{\text{Plug}} = \frac{\bar{\mu}}{\gamma \hat{\sigma}^2} \]

where \( \bar{\mu} \) and \( \hat{\sigma}^2 \) represent the sample mean and variance of excess returns. We also consider several combined estimators that shrink the plug-in estimator toward the ultraconservative portfolio decision rule \( \hat{\omega}_0 \) that simply allocates 100% of wealth to the risk-free asset (irrespective of historical data). These estimators vary only in the manner in which the shrinkage factor \( \alpha \) is determined. The various shrinkage estimators considered include

\[
\begin{align*}
\hat{\omega}_{\text{KZ}} &= (\hat{\alpha}_{\text{KZ}}) \hat{\omega}_{\text{Plug}} + (1 - \hat{\alpha}_{\text{KZ}})\hat{\omega}_0 \\
\hat{\omega}_{\text{BOOT}} &= (\hat{\alpha}_{\text{BOOT}}) \hat{\omega}_{\text{Plug}} + (1 - \hat{\alpha}_{\text{BOOT}})\hat{\omega}_0 \\
\hat{\omega}_{\text{AVG}} &= \frac{1}{2} \hat{\omega}_{\text{Plug}} + \frac{1}{2} \hat{\omega}_0 \\
\hat{\omega}_\gamma &= \frac{1}{\gamma} 
\end{align*}
\]

The first shrinkage estimator \( \hat{\omega}_{\text{KZ}} \) applies the estimator \( \hat{\alpha}_{\text{KZ}} \) of the optimal shrinkage factor suggested by Kan and Zhou (2007). The third estimator \( \hat{\omega}_{\text{BOOT}} \) estimates the optimal shrinkage factor using the approach outlined in Section 4 of the paper where a pseudo-sample of returns in the transformed asset space is generated and the shrinkage factor is estimated based on the plug-in estimator applied to the pseudo-sample of returns. We use the empirical distribution function as an estimate of the true distribution of excess returns to construct a pseudo-sample of size 6,000. As discussed in the previous section of the paper, the large pseudo-sample size minimizes the estimation error in the second step of the process. The final two estimators considered do not attempt to estimate the optimal shrinkage factor. The estimator \( \hat{\omega}_{\text{AVG}} \) averages the plug-in estimator with the 100% risk-free allocation. Of course, this is equivalent to an ad hoc assumption that the optimal
shrinking factor is 0.5. Finally, the estimator $\hat{\omega}_0$ is ignored the data entirely and simply allocates $(1/\gamma)$ to the risky asset. Note that this estimator obtains if an investor believes that the mean return is equal to its variance.

Given the interest in shrinking, it is natural to consider the estimator that applies the optimal degree of shrinkage, computed based on (16), to the plug-in estimator. We define this estimator as $\hat{\omega}_{\alpha^*} = \alpha^* \hat{\omega}_{plug} + (1 - \alpha^*)\hat{\omega}_0$. Figure 2 plots $\alpha^*$, the optimal degree of shrinkage, as a function of the sample size using the calibrated mean and variance of normally distributed excess returns. The plot indicates that the optimal degree of shrinkage is substantial even for relatively large sample sizes such as 500.

The design of the simulation experiment is as follows. A sample of $T$ returns is generated and the optimal portfolio weight is estimated using the different portfolio decision rules in (18). A new random observation of excess returns is then drawn, and the resulting portfolio return is computed for each of the portfolio decision rules considered. This process is repeated $B$ times, generating a series of $B$ out-of-sample returns for each estimator considered. Letting $\hat{\mu}_j$ and $\hat{\sigma}_j^2$ indicate the mean and variance of excess returns under the $j$-th portfolio decision rule, we compute estimates of the Sharpe ratio (SR) and certainty equivalent return (CER) associated with the decision rule as:

$$
\text{SR}_j = \frac{\hat{\mu}_j}{\hat{\sigma}_j},
$$

$$
\text{CER}_j = \hat{\mu}_j - \frac{\gamma}{2} \hat{\sigma}_j^2.
$$

By the law of large numbers, these estimates converge in probability to the true SR and CER values, so that with a sufficiently large simulation the population quantities may be essentially recovered. The true CER values have the economic interpretation as the maximum monthly management fee, as a percentage of wealth invested, that an investor would be willing to pay a money-manager who has exclusive access to a particular portfolio decision rule. Unless otherwise stated, the level of risk aversion $\gamma$ is set at three.

When optimal combining weights are estimated using a pseudo-sample of returns as described in the previous section of the paper, the resulting estimator can be computationally demanding, particularly since the pseudo-sample size should be set very high to achieve best results. For example, the $\hat{\omega}_{\text{BOOT}}$ estimator with a pseudo-sample size of 6,000 takes roughly 45 seconds to compute on a standard desktop machine, whereas computation of the other estimators in (18) is extremely rapid. In an applied setting, the longer computational time is not particularly problematic as the estimator need only be computed once; however, simulation experiments that include $\hat{\omega}_{\text{BOOT}}$ are quite costly. For this reason, we present results for the full set of estimators, including boot, only for sample sizes of 120 and 240.

Table 2 presents results for simulation experiments with sample sizes of $T = 120$ and $T = 240$. Four features of the results are noteworthy. First, the best performing estimator, both in terms of Sharpe ratio and CER, is the ad hoc estimator $\hat{\omega}_\gamma = 1/\gamma$ that allocates $1/3$ of wealth to the risky asset (recall that $\gamma = 3$). This is despite the fact that the truly optimal allocation given
the model parameters in Table 1 is in fact \( \omega^* = 65.2\% \). Second, the shrinkage estimators almost uniformly improve relative to the plug-in rule when performance is assessed using the economic CER which reflects investor preferences. The plug-in estimator achieves the lowest CER and for \( T = 240 \) and the second lowest for \( T = 120 \). Third, the marked improvement in the estimator \( \hat{\omega}_{AVG} \) as the sample size increases from 120 to 240 is due to the fact that the truly optimal shrinkage factor (see Figure 2) is not too far from 0.5 for a sample size of 240 but is substantially lower than this for a sample size of 120 so that the estimator is not shrinking enough in this case. Finally, note that the performance of the two alternative approaches to estimating the optimal shrinkage factor, \( \hat{\omega}_{KZ} \) and \( \hat{\omega}_{BOOT} \), perform very similarly for both sample sizes. Note that \( \hat{\omega}_{KZ} \) makes use of the assumption that returns are normally distributed while the estimator \( \hat{\omega}_{BOOT} \) does not. If returns are non-normal, the two estimators may exhibit less similar behavior; however, experiments in which returns are simulated using the skew student-\( t \) distribution lead to a similar finding.

To better understand the performance of the estimators \( \hat{\omega}_{KZ} \) and \( \hat{\omega}_{BOOT} \), it is useful to examine the empirical distribution of the estimated optimal shrinking factor \( \alpha^* \) at sample sizes of 120 and 240. Figure 3 presents a comparison (based on 2000 simulations) of the finite sample distributions of \( \hat{\alpha}_{KZ} \) and \( \hat{\alpha}_{BOOT} \) at a sample size of 120 and indicates on each plot the optimal shrinking factor, based on formula (16). Figure 4 presents a similar comparison for a sample size of 240. Three things are immediately apparent from Figures 3 and 4. First, the estimates of the optimal shrinking factor are very noisy at these sample sizes. Second, the distributions of both estimators are highly non-normal for both sample sizes considered, although the approximation to the normal does appear to improve at the higher sample size. Finally, the bootstrap estimator appears upward biased for the optimal shrinking factor.

Given that \( \hat{\omega}_{KZ} \) and \( \hat{\omega}_{BOOT} \) perform similarly, we omit the computationally intensive \( \hat{\omega}_{BOOT} \) estimator and repeat the simulation experiment for a range of sample sizes and with a much higher degree of precision (\( B = 50,000 \)) in order to obtain a more complete and accurate picture of the behavior of the various estimators. Figure 5 displays the resulting CERs as a function of sample size for each estimator. The plot not only includes the estimators in (18) but also the infeasible optimal shrinkage estimator \( \hat{\omega}_{opt} \). The results in the figure illustrate are striking: the best performing estimator in terms of CER is the inconsistent \( \hat{\omega}_{opt} \) estimator for sample sizes up to nearly 500. Interestingly, this estimator outperforms even the infeasible optimal shrinkage estimator for these sample sizes. For sample sizes greater than 500, the standard plug-in estimator achieves the highest CER among the feasible estimators and this estimator performs nearly identically to optimal shrinkage. For smaller sample sizes, the decision rule \( \hat{\omega}_{KZ} \) that attempts to estimate the optimal shrinking factor does outperform the plug-in estimator, suggesting that, despite noisy estimation of the optimal shrinking factor as evidenced in Figure 3, this decision rule is able to generate utility gains relative to the plug-in rule for sample sizes up to around 400. That said, the estimator that simply halves the plug-in estimator outperforms \( \hat{\omega}_{KZ} \) at these sample sizes. Although an ad hoc shrinking factor of 0.5 is inconsistent and biased for the truly optimal shrinking factor at virtually all sample sizes, it exhibits zero estimation error. On the other hand, while \( \hat{\omega}_{KZ} \) is a consistent
estimator for the optimal allocation, the substantial noise in estimating the associated shrinking factor severely degrades the performance of the estimator in terms of expected utility.

These results are rather discouraging from the viewpoint of shrinking the plug-in estimator since feasible attempts to shrink the plug-in estimator are never optimal among the set of estimators considered. For sample sizes less than 500, it is better to use the estimator \( \hat{\omega}_\gamma \). For sample sizes greater than 500 it is better to simply use the plug-in estimator. As a robustness check, Figure 6 displays results of the exact same exercise when for the case \( \gamma = 3 \). The qualitative findings are identical for this case suggesting that the findings are robust to the choice of risk aversion within the range typically considered in empirical studies.

5.3. Out-of-Sample Results

This section presents results based on an out-of-sample asset allocation exercise conducted using the US excess returns data. The setup is the same as the simulation experiments described above, except that returns are the actual sample returns rather than simulated returns. The out-of-sample experiment applies a rolling window of length 120 that is used to compute estimates of the optimal allocation for each estimator considered. This estimate is then applied in the subsequent month and the resulting portfolio excess return is tabulated. This sample then rolls forward one month and the process is repeated. Ultimately, the exercise yields 324 out-of-sample excess returns for each of the estimators. A set of summary statistics, including the mean, variance, Sharpe ratio and CER is then computed for the set of out-of-sample portfolio returns. The risk aversion parameter is set to three. Table 3 presents results.

Consistent with results in the simulation setting, the best performing estimator is \( \hat{\omega}_\gamma \). The second best performing estimator is \( \hat{\omega}_{AVG} \), which simply halves the plug-in estimate. The plug-in estimator as well as the shrinkage versions that attempt to estimate the optimal shrinkage factor perform poorly. Note that the economic significance of CER differences among the estimators is much larger in the experiment using actual data relative to the simulation results in Table 2. Indeed, a ballpark calculation puts the CER difference between \( \hat{\omega}_\gamma \) and the plug-in estimator at roughly 0.25% per month, or approximately 3% on an annualized basis. An interpretation is that an investor would be willing to pay up to 3% annually to switch out of the plug-in estimator into the the estimator \( \hat{\omega}_\gamma \). This compares with a corresponding CER difference of roughly .04% per month, or approximately 0.48% annualized, obtained in the simulation setting with the same sample size (see Panel A of Table 2). Figure 7 provides some insight regarding the substantially larger CER differences using the actual out-of-sample data. These larger differences, and the overall weak performance of a number of the estimators, is related to the large drop in equity prices in the late 1990s related to the so-called “internet bubble.” Because the preceding years had witness relatively large positive returns, the plug-in estimator based on the previous 10 years data is aggressively long in the risky stock index. The resulting large negative return leads to out-of-sample negative returns of nearly 40% for several of the estimators. The estimator \( \hat{\omega}_\gamma \), on the other hand, simply ignores the data and therefore does not experience the extremely large negative return. On the one hand this illustrates
that an added benefit of the ad hoc estimator is a type of robustness to events of this sort. On the other hand, this perhaps illustrates the need for complementary simulation analysis since this late 1990s episode drives much of the performance in the actual out-of-sample experiment.

5.4. **Summary**

The simulation and out-of-sample analysis described in this section of the paper illustrates that, even in an extremely simple asset allocation problem with only a single risky asset, estimation risk poses a serious problem in the portfolio choice setting. In both the simulated and actual data, we find that for empirically plausible sample sizes (less than 500 monthly observations) the best estimator in terms of an expected utility criterion is an ad hoc estimator that allocates \((1/\gamma)\) to the risky asset. This estimator not only outperforms the standard plug-in estimator of the optimal portfolio, but it also outperforms a variety of estimators that attempt to shrink the plug-in estimator toward zero.

6. **Conclusion**

In the context of a canonical asset allocation problem with i.i.d. returns, this paper considers forming linear combinations of an arbitrary basis set of estimators of the optimal allocation (portfolio decision rules) in order to create new estimators with superior performance as measured by the risk of the estimator. The problem of forming the optimal linear combination of a set of portfolio decision rules is shown to be formally equivalent to the original asset allocation problem with a transformed set of asset returns, where the returns on the \(N\) original assets are replaced by a set of \(J\) risky returns, each corresponding to one of the basis estimators. An important caveat with respect to the combination of estimators is that, since the optimal combination depends upon features of the true returns distribution, the optimal combining weights are latent and the combined estimator based on them is infeasible. The paper discusses approaches to feasible combination and these approaches deliver encouraging performance in a simulation experiment. Simple equal-weighting of a set of basis rules outperforms more sophisticated attempts to estimate optimal combining weights in the simulation exercise. This finding is in line with empirical findings in the related literature on forecast combination.

There is much additional work to be done. The empirical analysis in this paper considers only a simple allocation problem with a single risky asset. Additional empirical work with multiple assets, possibly including preferences over higher moments, would be of great interest. The present paper does not explicitly consider the important issue of the choice of basis estimators for combination. In the mean-variance setting Kan and Zhou consider the combination of the plug-in estimator of the tangency portfolio and GMV portfolio. Other estimators, such as the estimators suggested by MacKinlay and Pastor (2000) and Pastor (2000) may also be interesting candidates for combination.
7. Appendix

This appendix contains proofs of results or algebra that are too cumbersome for the main text.

Proof of Lemma 1

This follows from the fact that algebraic operations such as addition and multiplication with measurable functions preserves measurability. Note that since the allocation to the risk-free asset is determined implicitly as $1 - 1_N \hat{\omega}^C$, the wealth constraint is satisfied.

Proof of Lemma 2

Consider a (possibly random) set of combining weights $\alpha(R^T)$, where any dependence on $R^T$ is henceforth suppressed. Introduce a $J + 1$-th portfolio decision rule $\hat{\omega}_0$, the portfolio decision rule that allocates 100% of wealth to the risk-free asset irrespective of the historical data $R^T$ so that $\hat{\omega}_0$ is simply an $N \times 1$ vector of zeros. Given $\alpha$, let $\alpha_0 \equiv 1 - \alpha'1_J$ and create an augmented vector of $J + 1$ combining weights by defining the $(J + 1) \times 1$ vector $\hat{\alpha} \equiv (\alpha_0, \alpha')'$. Now make the following definition:

$$\hat{\omega}^C(R^T) = \alpha_0 \hat{\omega}_0 + \sum_{j=1}^{J} \alpha_j (R^T) \hat{\omega}_j (R^T)$$

By Lemma 1, the combined estimator $\hat{\omega}^C(R^T)$ is clearly a portfolio decision rule. Furthermore, by the definition of $\alpha_0$ it is clear that $\alpha'1_{J+1} = 1$, so that the vector $\hat{\alpha}$ carries the interpretation of a set of portfolio weights over the $J + 1$ portfolio decision rules. This establishes the Lemma.

Proof of Proposition 1

Given $J$ portfolio decision rules $\hat{\omega}_j, j = 1, ..., J$, define $\hat{\Omega} \equiv [\hat{\omega}_1 \ ... \ \hat{\omega}_J]$, an $N \times J$ matrix. Let $\hat{\omega}^C$ be defined as in (6) and note that we may write $\hat{\omega}^C = \hat{\Omega} \alpha$. Now, introduce a $J \times 1$ vector of assets with single period gross returns $R^fJ + \hat{\Omega}'(\hat{R} - R^f1_N)$ and consequently excess returns $\hat{R} \equiv \hat{\Omega}'(\hat{R} - R^f1_N)$. Consider the optimal portfolio allocation problem faced by an investor with initial wealth $W$ and preferences $U(\cdot)$ who must allocate wealth among these $J$ risky assets and the risk-free asset. Letting $\alpha$ denote the $J \times 1$ vector of asset allocations to the risky assets, the investor’s problem may be stated as:

$$\max_{\alpha} E_F(\hat{\omega}_1, ..., \hat{\omega}_J, \hat{R}) \left[ U \left( W \left( R^f + \alpha'\hat{R} \right) \right) \right].$$

Plugging in for $\hat{R}$ and rearranging yields:

$$\max_{\alpha} E_F(\hat{\omega}_1, ..., \hat{\omega}_J, \hat{R}) \left[ U \left( W \left( R^f + \alpha' \left( \hat{\Omega}'(\hat{R} - R^f1_N) \right) \right) \right) \right]$$

$$= \max_{\alpha} E_F(\hat{\omega}_1, ..., \hat{\omega}_J, \hat{R}) \left[ U \left( W \left( R^f + \left[ \hat{\Omega} \alpha \right]' \left( \hat{R} - R^f1_N \right) \right) \right) \right],$$

which is (8). This demonstrates that the two problems are indeed equivalent.

Proof of Proposition 2
That $\mu_R = \mu'_{\Omega} \mu$ follows immediately from the i.i.d. assumption. For the variance result we have (again using the independence):

$$\Sigma_R = E \left[ \hat{\Omega}' \Sigma \hat{\Omega} \right],$$

a $J \times J$ matrix. Consider the $l,k$-th element of the matrix $\hat{\Omega}' \Sigma \hat{\Omega}$, which may be written as

$$\hat{\Omega}' \Sigma \hat{\Omega}_{lk}$$

where the notation $\hat{\Omega}_l$ indicates the $l$-th column of $\hat{\Omega}$ and similarly for $\hat{\Omega}_k$. Expanding the quadratic form in $\Sigma$ gives the representation:

$$\hat{\Omega}'_l \Sigma \hat{\Omega}_k = \sum_{i,j \leq N} \sigma_{i,j} \hat{\omega}_{l,i} \hat{\omega}_{k,j}$$

where, e.g., $\hat{\omega}_{l,i}$ indicates the $i$-th element of $\hat{\omega}_l$. Now, taking the expectation and using the linearity of the expectation operator gives

$$\sum_{i,j \leq N} \sigma_{i,j} E[\hat{\omega}_{l,i} \hat{\omega}_{k,j}] = \sum_{i,j \leq N} \sigma_{i,j} \left\{ \text{cov}(\hat{\omega}_{l,i} \hat{\omega}_{k,j}) + E[\hat{\omega}_{l,i}] E[\hat{\omega}_{k,j}] \right\}$$

which is the desired result (13) up to the notational conventions described in the paper.
References


Table 1

Summary Statistics

This table reports summary statistics for MSCI US Equity Index returns sampled monthly over the period 1970:1 – 2007:1. Statistics are reported over the full sample period as well as over the first and second halves of the sample period. The mean, median and standard deviation are expressed in percent per month. The kurtosis is reported in excess of three, the kurtosis for the normal distribution.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.37862</td>
<td>0.17174</td>
<td>0.58456</td>
</tr>
<tr>
<td>Median</td>
<td>0.77797</td>
<td>0.11467</td>
<td>0.95558</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>4.39842</td>
<td>4.77475</td>
<td>3.98889</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.59061</td>
<td>-0.54468</td>
<td>-0.58337</td>
</tr>
<tr>
<td>Excess kurtosis</td>
<td>2.54331</td>
<td>2.972251</td>
<td>1.035681</td>
</tr>
<tr>
<td>Maximum</td>
<td>15.77264</td>
<td>15.77264</td>
<td>10.48592</td>
</tr>
</tbody>
</table>
Table 2
Simulation Results

This table reports results for a simulation analysis that examines the performance of several estimators of the optimal allocation in a portfolio problem with a single risky asset and a risk-free asset. The investor is assumed to exhibit mean-variance preferences with risk aversion parameter $\gamma=3$. In each simulation a sample of returns is generated and the optimal portfolio weight is estimated using a number of alternative portfolio decision rules. Random excess returns are normally distributed with mean and variance calibrated to the MSCI US Equity Index. A new random observation of excess returns is then drawn, and the resulting portfolio return is computed for each of the portfolio decision rules considered. This process is repeated 2,000 times, generating a series of 2,000 simulated out-of-sample returns for each estimator considered. The table then reports the mean and standard deviation of these returns, the corresponding Sharpe ratio and the certainty equivalent return (CER). Panel A reports results for a sample size of 120 (ten years) while Panel B reports results for a sample size of 240 (twenty years).

### Panel A: Sample Size = 120

<table>
<thead>
<tr>
<th>Estimator</th>
<th>mean</th>
<th>sd</th>
<th>SR</th>
<th>CER</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\omega}_{\text{PLUG}}$ (Plug-In Estimator)</td>
<td>0.3285</td>
<td>4.2176</td>
<td>0.0779</td>
<td>0.0617</td>
</tr>
<tr>
<td>$\hat{\omega}_{\text{KZ}}$ (Kan-Zhou Shrinkage)</td>
<td>0.2148</td>
<td>3.1066</td>
<td>0.0691</td>
<td>0.0860</td>
</tr>
<tr>
<td>$\hat{\omega}_{\text{BOOT}}$ (Bootstrap Shrinkage)</td>
<td>0.2308</td>
<td>2.9128</td>
<td>0.0792</td>
<td>0.0875</td>
</tr>
<tr>
<td>$\hat{\omega}_{\text{AVG}}$ (=1/2 * Plug-In Estimator)</td>
<td>0.1258</td>
<td>2.1046</td>
<td>0.0598</td>
<td>0.0594</td>
</tr>
<tr>
<td>$\hat{\omega}_{\gamma}$ (1/$\gamma$ Estimator)</td>
<td>0.1344</td>
<td>1.4684</td>
<td>0.0916</td>
<td>0.1021</td>
</tr>
</tbody>
</table>

### Panel A: Sample Size = 240

<table>
<thead>
<tr>
<th>Estimator</th>
<th>mean</th>
<th>sd</th>
<th>SR</th>
<th>CER</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\omega}_{\text{PLUG}}$ (Plug-In Estimator)</td>
<td>0.2957</td>
<td>3.6922</td>
<td>0.0801</td>
<td>0.0912</td>
</tr>
<tr>
<td>$\hat{\omega}_{\text{KZ}}$ (Kan-Zhou Shrinkage)</td>
<td>0.2276</td>
<td>2.8523</td>
<td>0.0798</td>
<td>0.1055</td>
</tr>
<tr>
<td>$\hat{\omega}_{\text{BOOT}}$ (Bootstrap Shrinkage)</td>
<td>0.2396</td>
<td>2.9907</td>
<td>0.0801</td>
<td>0.1055</td>
</tr>
<tr>
<td>$\hat{\omega}_{\text{AVG}}$ (=1/2 * Plug-In Estimator)</td>
<td>0.1619</td>
<td>1.8136</td>
<td>0.0893</td>
<td>0.1126</td>
</tr>
<tr>
<td>$\hat{\omega}_{\gamma}$ (1/$\gamma$ Estimator)</td>
<td>0.1488</td>
<td>1.4743</td>
<td>0.1009</td>
<td>0.1162</td>
</tr>
</tbody>
</table>
This table reports results for a simulation analysis, an out-of-sample asset allocation exercise conducted using the MSCI US Equity Index returns data. The setup is the same as the simulation experiments described above, except that returns are the actual sample returns rather than simulated returns. The out-of-sample experiment applies a rolling window of length 120 that is used to compute estimates of the optimal allocation for each estimator considered. This estimate is then applied in the subsequent month and the resulting portfolio excess return is tabulated. This sample then rolls forward one month and the process is repeated. Ultimately, the exercise yields 324 out-of-sample excess returns for each of the estimators. A set of summary statistics, including the mean, variance, Sharpe ratio and CER is then computed for the set of out-of-sample portfolio returns. The risk aversion parameter is set to three.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>mean</th>
<th>sd</th>
<th>SR</th>
<th>CER</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\omega}_{\text{PLUG}} ) (Plug-In Estimator)</td>
<td>0.3776</td>
<td>5.7493</td>
<td>0.0657</td>
<td>-0.1182</td>
</tr>
<tr>
<td>( \hat{\omega}_{\text{KZ}} ) (Kan-Zhou Shrinkage)</td>
<td>0.1837</td>
<td>4.2916</td>
<td>0.0428</td>
<td>-0.1338</td>
</tr>
<tr>
<td>( \hat{\omega}_{\text{BOOT}} ) (Bootstrap Shrinkage)</td>
<td>0.1425</td>
<td>4.4216</td>
<td>0.0322</td>
<td>-0.1096</td>
</tr>
<tr>
<td>( \hat{\omega}_{\text{AVG}} ) (=( \frac{1}{2} ) * Plug-In)</td>
<td>0.1888</td>
<td>2.8747</td>
<td>0.0657</td>
<td>0.0648</td>
</tr>
<tr>
<td>( \hat{\omega}_{\gamma} ) (( \frac{1}{\gamma} ) Estimator)</td>
<td>0.1884</td>
<td>1.4422</td>
<td>0.1306</td>
<td>0.1572</td>
</tr>
</tbody>
</table>
Fig. 1: Histogram and density plot for US monthly excess returns. Returns are excess returns on the MSCI US Equity Index over the sample period 1970:1 – 2007:1. The red line plots the density of a normal distribution fit to the data. The green line plots the density of a skew student-t distribution fit to the data, while the blue line plots a nonparametric density estimate.
Fig. 2: Optimal shrinkage factor for the plug-in estimator of the optimal portfolio weight as a function of sample size. The figure plots the optimal shrinkage factor applied to the plug-in estimator of the optimal portfolio allocation to a single risky asset as a function of sample size. The optimal shrinkage factor is computed using results obtained by Kan and Zhou (200x) assuming that the excess return on a single risky asset is normally distributed with a Sharpe ratio calibrated to match excess returns on the MSCI US Equity Index over the monthly sample period 1970x – 2007:1.
Figure 3: Finite sample performance of estimators for the optimal shrinkage factor ($T = 120$). The top panel of the figure displays a nonparametric density estimate for the estimator of the optimal shrinking factor suggested by Kan and Zhou (2007). The bottom panel displays a nonparametric density estimate for the estimator of the optimal shrinking factor as described in Section 4 of the paper with the empirical distribution of returns serving as the basis for generating a “pseudo-sample” of returns. Both plots are based on 2,000 simulation replications with normally distributed excess returns calibrated to MSCI US Equity Index excess returns. The vertical line indicates the theoretically optimal shrinking factor.
Figure 4: Finite sample performance of estimators for the optimal shrinkage factor ($T=240$). The top panel of the figure displays a nonparametric density estimate for the estimator of the optimal shrinking factor suggested by Kan and Zhou (2007). The bottom panel displays a nonparametric density estimate for the estimator of the optimal shrinking factor as described in Section 4 of the paper with the empirical distribution of returns serving as the basis for generating a “pseudo-sample” of returns. Both plots are based on 2,000 simulation replications with normally distributed excess returns calibrated to MSCI US Equity Index excess returns. The vertical line indicates the theoretically optimal shrinking factor.
Fig 5: Certainty equivalent return associated with different estimators of the optimal allocation ($\gamma = 3$). The figure displays CER values (as a monthly percentage) for various estimators of the optimal allocation. There is a single risky asset with normally distributed excess returns calibrated to MSCI US index excess returns. Preferences are mean-variance with $\gamma = 3$. Results in the figure are based on 50,000 replications. The estimator “KZ” refers to a shrinkage factor applied to the Plug-in estimator with the shrinkage factor estimated as proposed by Kan and Zhou (2007). The estimator “KZopt” is similar except that the shrinkage factor is the theoretically optimal shrinkage factor.
Fig 6: Certainty equivalent return associated with different estimators of the optimal allocation ($\gamma = 7$). The figure displays CER values (as a monthly percentage) for various estimators of the optimal allocation. There is a single risky asset with normally distributed excess returns calibrated to MSCI US index excess returns. Preferences are mean-variance with $\gamma = 7$. Results in the figure are based on 50,000 replications. The estimator “KZ” refers to a shrinkage factor applied to the Plug-in estimator with the shrinkage factor estimated as proposed by Kan and Zhou (2007). The estimator “KZopt” is similar except that the shrinkage factor is the theoretically optimal shrinkage factor.
Figure 7: Time series plot of out-of-sample excess returns on the MSCI US stock index for various portfolio decision rules, 1980:1 – 2007:1. Out-of-sample excess returns are generated by applying each estimator using a rolling window of size 120 (10 years) of historical US excess returns data.