Combination Approaches to Estimating Optimal Portfolios

Bradley S. Paye*
Jones Graduate School of Business, Rice University

July 15, 2010

JEL Classification: G1, C13

Keywords: Combined estimators, shrinkage, portfolio choice, mean variance portfolios, statistical decision theory.

Abstract

This paper considers the potential to improve portfolio performance by combining multiple estimators of the unknown optimal asset allocation. When portfolio estimators are evaluated from an economic perspective, I show that the *statistical* problem of optimally combining estimators is equivalent to a standard asset allocation problem. Using this equivalence, I state numerous general results regarding optimal combinations of portfolio estimators. As an application, I consider combining the nonparametric estimator proposed by Brandt (1999) with an ultra-conservative portfolio strategy. Simulation results illustrate that this form of 'shrinkage' can deliver economically significant performance improvements at empirically relevant sample sizes.

---

*Bradley Paye, Jones Graduate School of Business, Rice University, 6100 Main Street - MS531, Houston, TX 77005, Phone: (713) 348-6030, Fax: (713) 348-6296, E-mail: bpaye@rice.edu. I thank Barbara Ostdiek and James Weston for helpful comments, as well as seminar participants at Rice University and HEC Montreal.

Combination Approaches to Estimating Optimal Portfolios

Abstract

This paper considers the potential to improve portfolio performance by linearly combining multiple estimators of the unknown optimal asset allocation. When the relevant loss function is economic, I show that the *statistical* problem of optimally combining estimators is equivalent to a re-parameterized version of the underlying asset allocation problem. Using this equivalence, I state numerous general results regarding optimal combinations of portfolio estimators. As an application, I consider combining the nonparametric estimator proposed by Brandt (1999) with an ultra-conservative portfolio strategy. Simulation results illustrate that this form of 'shrinkage' can deliver economically significant performance improvements at smaller sample sizes.

Introduction

Estimation risk represents a serious concern for data-driven portfolio strategies. When the effects of estimation risk are ignored and portfolio allocation proceeds as though estimated characteristics of asset returns represent the "truth," actual investment performance may be very poor. The literature is rife with approaches designed to mitigate the effects of estimation risk.[1] Unfortunately, recent research questions the effectiveness of these approaches. In the mean-variance setting, DeMiguel, Garlappi, and Uppal (2007) assess the performance of a wide set of portfolio rules, across a number of empirical datasets, and find that none significantly outperforms a simple $1/N$ rule. They conclude that "there are still many 'miles to go' before the gains promised by optimal portfolio choice can actually be realized out-of-sample."

Although it is simple, the $1/N$ portfolio rule adheres to a basic tenet of portfolio theory: an investor facing multiple risks should diversify. Estimation risk represents an additional source of risk, beyond the inherent uncertainty in asset returns that investors face even if the (ex ante) optimal allocation is known. Different estimators possesses different sampling characteristics, and therefore entail different estimation risks. From this perspective, financial intuition suggests that it may be prudent to diversify across multiple estimators, rather than 'put one's eggs in a single basket.' One way to diversify among a set of estimators is to construct a new estimator as a weighted combination. But what is the *optimal* way to combine portfolio estimators? This paper provides a general analysis of this important question.

Determining the optimal combination of a set of estimators requires a metric to value alternative estimators. In this paper, I assume that estimators are evaluated from an *economic* perspective. This approach hypothesizes a particular investor utility function, and equates the

[1]The bulk of existing literature focuses on the mean-variance portfolio problem. Jobson and Korkie (1980) and Michaud (1989) document the poor performance of the standard 'plug-in' estimator in this setting. Proposed approaches for mitigating estimation risk 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 constraints on portfolios (Jagannathan and Ma (2003) and DeMiguel, Garlappi, Nogales and Uppal (2009)), 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)), and shrinkage estimators of the covariance matrix of returns (Ledoit and Wolf (2003, 2004)).

value of any estimator with the expected utility achieved if the estimator is used as a portfolio rule.[2] An economic approach to evaluating portfolio estimators is common, if not prevalent, in recent literature.[3]

The practice of using an investor utility function to value portfolio estimators imposes a *financial* structure upon a *statistical* problem. The resulting financial structure permits economically meaningful quantification of the merits of various portfolio estimators. For example, an empirical paper might report an annualized certainty equivalent return of 1.5% annually for a particular portfolio estimator. The associated interpretation is that an investor would be willing to pay up to 150 basis points annually as a 'management fee' in order to have access to this portfolio rule, as opposed to investing all wealth in a risk-free bond. Similar interpretations are available when comparing alternative estimators.

The paper's main contribution is to show that the *statistical* problem of optimally combining a set of portfolio estimators is mathematically equivalent to a standard portfolio problem. This result is a direct consequence of the financial structure imposed under the economic approach to evaluating portfolio estimators. The equivalent portfolio problem takes the same form as the original portfolio problem, except that the asset payoffs are adjusted to incorporate two separate sources of risk: randomness due to sample variability in the portfolio estimators (estimation risk), and inherent randomness in asset returns.[4]

The equivalence result established in the paper proves very convenient. Using it, I derive a series of general results regarding the optimal combination of portfolio estimators. First, the optimal combining weights may be characterized using standard first-order conditions for optimal portfolio weights. Second, the equivalence result delivers comparative statics results

[2]This expectation involves two separate sources of randomness. The first involves sample variation in the estimator (randomness in estimated portfolio weights), while the other relates to inherent variability in stock returns. Often the expected utility associated with an estimator is converted to the more economically interpretable *certainty equivalent return*. Section 1 of the paper provides a formal development of this concept.

[3]West, Edison and Cho (1993) represents an early empirical paper to value portfolio estimators using economic utility functions. More recent studies include Fleming, Kirby and Ostdiek (2001, 2003), Marquering and Verbeek (2004), Guidolin and Timmermann (2007), Kan and Zhou (2007), DeMiguel, Garlappi and Uppal (2008), Campbell and Thompson (2008), Goyal and Welch (2007), Kirby and Ostdiek (2010) and Tu and Zhou (2010).

[4]The number of assets in the equivalent portfolio problem is equal to the number of estimators that are combined, which differs in general from the number of tradeable risky assets in the economy.

linking optimal combining weights with parameters such as initial investor wealth. As an example, when preferences exhibit linear risk tolerance, the optimal weight on each estimator is linear in wealth, and a form of two-fund separation holds, such that all investors hold the same 'mutual fund' among the set of portfolio estimators.

For the special case of mean-variance preferences, any set of portfolio estimators will generate a standard mean-variance efficient frontier. Since the assets corresponding to the estimators will typically lie inside the the efficient frontier, the same expected return may be achieved at lower variance by forming a nontrivial 'portfolio of estimated portfolios.' In this setting, optimal combining weights may be expressed analytically as a function of the mean and covariance of (suitably transformed) asset returns.

A second special case of interest involves combining an arbitrary (nontrivial) portfolio estimator with the 'ultraconservative' portfolio rule that allocates 100% to the risk-free asset (irrespective of the sample data). This combination problem carries the interpretation of optimal 'shrinkage' toward the ultraconservative portfolio rule. The equivalence result developed in the paper provides several useful insights regarding optimal shrinkage. For example, when a portfolio estimator is unbiased, complete shrinkage is *never* optimal.

Because optimal combining weights depend upon the true probability law governing returns, these weights must themselves be estimated. I suggest a general approach for estimating combining weights that exploits the equivalence between the combination problem and the standard portfolio problem. The approach involves plugging in an estimate of the unknown distribution of risky asset returns, not to directly estimate optimal *portfolio* weights (as is standard practice), but rather to estimate optimal *combining* weights. By simulating a large number of returns for the equivalent portfolio problem, one may recover, with high precision, optimal combining weights under the estimated (plug-in) distribution of returns.

Since the plug-in distribution used to determine combining weights differs from the true, unknown distribution, combining weights are subject to sampling variability. Indeed, this sampling variability may be severe. Intuition again follows from the equivalence result: since the problem of finding optimal combining weights is equivalent to a standard portfolio problem, the same estimation problems that plague data-driven portfolio allocation will plague efforts to recover optimal combining weights. As an alternative to data-driven searches for optimal combining weights, I propose simple averaging of multiple estimators. The forecasting literature provides motivation for this approach, since equal-weighted forecast combinations often perform well relative to more sophisticated combination strategies (see, e.g., the review article by Timmermann (2006).

To illustrate the main concepts developed in the paper, I present a simulation analysis based on a simple portfolio problem involving a single risky stock and a risk-free bond. To highlight the generality of results, the simulation focuses on a portfolio problem outside of the mean-variance setting. Specifically, investor preferences exhibit constant absolute risk aversion. Log excess returns on the risky asset are simulated from a (negatively) skewed and leptokurtotic distribution; consequently the optimal portfolio cannot be expressed using a simple mean-variance tradeoff. As a 'base' estimator, I consider a method of moments estimator in the spirit of Brandt (1999). This nonparametric estimator sets to zero the sample analog of the first-order condition characterizing the optimal portfolio weight. The estimator is consistent for the optimal allocation, but is subject to estimation risk in finite samples.

The simulation explores whether the finite sample performance of the method of moments estimator may be improved by 'shrinking' this estimator toward the ultraconservative portfolio rule. I consider several approaches to shrinkage, including a sophisticated (data-driven) attempt to estimate of the optimal shrinking factor, an *ad hoc* shrinkage method that simply halves the method of moments estimator, and a hybrid approach that averages the portfolio weights obtained under the estimated and *ad hoc* shrinkage approaches.

The results show that the method of moments estimator suffers severely from estimation risk at smaller sample sizes. Shrinkage can deliver economically significant performance improvements, but only the *ad hoc* and hybrid shrinking methods generate considerable improvements. The data-driven shrinkage estimator, by contrast, performs poorly as it generates highly variable estimates of the shrinkage factor unless the sample is very large, at which point there is little benefit to shrinking. Viewing shrinkage as a special case of combining multiple estimators, these results suggest that purely data-driven attempts to combine portfolio estimators may perform poorly. Simple combination strategies, or hybrid approaches that temper the influence of sample data in determining combining weights, are likely to better capture the benefits of diversifying across estimators.

This paper relates to recent studies by Kan and Zhou (2007) and Tu and Zhou (2010). To combat the adverse effects of estimation risk in a mean-variance portfolio problem, Kan and Zhou (2007) propose 'two-fund' and 'three-fund' portfolio rules. The two-fund rule combines the standard plug-in estimator (the sample tangency portfolio) with the risk-free asset, while the three-fund rule combines the sample tangency portfolio, the riskless asset and the estimated global minimum variance portfolio. Tu and Zhou (2010) explore additional estimators that combine the $1/N$ allocation rule with other estimated portfolios, including the three-fund rule. The estimators proposed by Kan and Zhou (2007) and Tu and Zhou (2010) constitute important special cases of the type of combined estimator considered in this paper.

The present paper generalizes the combination approach promoted by Kan and Zhou (2007) and Tu and Zhou (2010) in several ways. First, the set of 'basis' estimators used to form combinations is completely arbitrary in this paper. Second, the portfolio problem is general, and investor preferences need not be summarized by a simple mean-variance tradeoff. This is important, since the deleterious effects of estimation risk may be exacerbated when investors are concerned about additional aspects of the distribution of portfolio returns, including skewness and tail-behavior. Third, the distribution of returns is considered unknown to the investor, even up to a specific parametric model, whereas Kan and Zhou (2007) and Tu and Zhou (2010) assume that returns are multivariate normal.[5]

The paper also provides a new perspective regarding the efficacy of various mean-variance portfolio rules. Although DeMiguel, Garlappi and Uppal (2007) describe the $(1/N)$ rule as "naive", there is a sense in which this rule cleverly exploits the benefits of diversification that combination offers. Rather than combine two or three standard estimators, the $1/N$ rule combines a large number (N) of trivial estimators, each of which allocates 100% of wealth to a particular asset irrespective of sample outcomes. From this perspective, finding a (non-trivial) data-driven portfolio rule that outperforms the $1/N$ strategy amounts to identifying a superior combination strategy. Tu and Zhou (2010) apply what amounts to a second layer of combination, by combining the $1/N$ rule with the three-fund estimator (or alternatives), and

[5]There is a price for this generality. By assuming a normal distribution for returns, Kan and Zhou (2007) are able to show that certain estimators, including the 'plug-in' estimator, are in fact *inadmissible*. Results of this sort are not available in the absence of a strong distributional assumption regarding returns.

find that this estimator does outperform the $1/N$ rule.

The remainder of the paper proceeds as follows. Section 1 describes the portfolio problem and formalizes the economic approach used to value various portfolio estimators. Section 2 considers the problem of forming the optimal linear combination of a set of portfolio estimators and states the fundamental equivalence result. Section 3 applies this equivalence to derive various results regarding the optimal combination of portfolio estimators. Section 4 discusses feasible approaches to estimating combining weights. Section 5 presents the simulation analysis, while Section 6 provides concluding discussion.

1 Evaluating Portfolio Estimators

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 Background

Consider an investor with initial wealth W_0 who must form a portfolio among $N + 1 \in \mathbb{N}$ financial assets. Among these assets, N are risky while the remaining asset offers a risk-free gross return of R_f. The i-th risky asset pays a random gross return of R_i for $i = 1, ..., N$ and the corresponding $N \times 1$ vector of risky returns is denoted R. The position is held for one period and then liquidated.

The $N \times 1$ vector ω represents the investor's portfolio allocation over the risky assets so that ω_i indicates the fraction of wealth allocated to the i-th asset for $i = 1, ...N$. The allocation to the risk-free asset, denoted ω_0, is determined implicitly as $\omega_0 = 1 - \omega' 1_N$ under the standard wealth constraint, where 1_N represents an $N \times 1$ vector of ones. The vector of excess returns on the risky assets is $R_e = (R - R_f 1_N)$.

End-of-period wealth W may be expressed as a function of initial wealth, the portfolio allocation and the excess return using the fundamental relation

$$W = W_0(R_f + \omega' R_e). \tag{1}$$

I maintain the following assumption on the stochastic process governing risky stock returns:

Assumption 1 *The excess return vector R_e on risky assets is independently and identically distributed (iid) with cumulative distribution function F_{R_e}.*

The assumption that R_e is identically distributed is made for convenience and may be relaxed at the cost of additional notation. The assumption that excess returns are *independent* is more substantiative. This assumption rules out any temporal predictability in asset returns, including predictability in mean and variance. Section 6 of the paper discusses extension to a setting with predictable returns. convenience and may be relaxed at the cost of additional notation. The assumption that excess returns are *independent* is more substantiative. This assumption rules out any temporal predictability in asset returns, including predictability in mean and variance. Section 6 of the paper discusses extension to a setting with predictable returns.

Suppose that investor preferences admit an expected utility representation with a strictly increasing, concave and twice continuously differentiable von Neumann-Morgenstern utility function $U(W)$ defined over end-of-period wealth. The investor's portfolio choice problem takes the form:[6]

$$\begin{aligned} & \max_{\omega} \mathbb{E}_{F_{R_e}} [U(W)] \\ = & \max_{\omega} \int U(W_0(R_f + \omega' R_e)) dF_{R_e}. \end{aligned} \tag{2}$$

The optimal portfolio allocation ω^* is characterized by the first-order necessary conditions:

$$\omega^* \equiv \omega \text{ s.t. } \mathbb{E}_{F_{R_e}} \left[U'(W)(R_e) \right] = 0. \tag{3}$$

This formulation of the problem imposes no short sales, borrowing, or other restrictions on the position that may be taken in a particular asset.

The portfolio allocation problem in (2) implicitly assumes that the distribution of returns F_{R_e} is *known* to the investor. Consequently, the problem requires no data and simply involves

[6]Some of the random variables considered later in the paper (e.g., trivial estimators of the optimal portfolio) are discrete random variables. Consequently, I use Stieltjes integrals to represent expectations without the need to specify whether the corresponding random variable is discrete, continuous or mixed. To economize on notation, I do not explicitly distinguish between a random variable and its realized value. This convention leads to a mild abuse of notation in integral representations. For example, I write $\int g(R_e) dF_{R_e}$ in place of the more formal expression $\int g(r_e) dF_{R_e}(r_e)$.

solving for ω^*, possibly numerically. The central focus of this paper, by contrast, is the realistic scenario in which the investor does *not* know F_{R_e}, possibly even up to a specific parametric model.

1.2 Estimators of the Optimal Portfolio

Suppose that a historical sample of excess returns $R_e^T \equiv \left(R_{e,1}^{\prime}, ..., R_{e,T}^{\prime}\right)^{\prime}$ of size T is available for inference regarding ω^*. An estimator $\hat{\omega}$ of the optimal allocation is a (measurable) mapping from realizations of sample returns to $\mathbb{R}^N$.[7] The estimator $\hat{\omega}$ possesses a cumulative distribution function $F_{\hat{\omega}}$. It is important to note that $F_{\hat{\omega}}$ ultimately depends upon both the distribution of excess returns F_{R_e} and the sample size T. That said, this dependence is left implicit to economize on notation.

There are many possible approaches to estimating optimal portfolio weights. A common *plug-in* approach assumes a parametric model $F_{R_e}(r_e;\theta)$ for the distribution of excess returns on risky assets. Given this model and an estimator $\hat{\theta}$ of θ, $\hat{\omega}$ is defined as the solution to the portfolio problem in (2) with θ replaced by $\hat{\theta}$. The plug-in approach encompasses a vast number of different portfolio estimators, which vary not only over the distributional model assumed for returns (e.g., Gaussian versus 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)).[8] Nonparametric estimation methods that make no distributional assumptions are also available. Brandt (1999) develops such an approach based on the moment conditions characterizing the optimal allocation.

1.3 The Risk of a Portfolio Estimator under an Economic Loss Function

From the perspective of statistical decision theory, the performance of an estimator is assessed using *risk*, defined as the expected loss for the estimator, where the expectation is taken with respect to the estimator's sampling distribution. As an example, consider the problem of estimating an arbitrary scalar parameter β. The most common loss function in this setting is

[7] This formulation implies that there are no constraints on positions in the risky assets other than the wealth constraint. The estimated allocation to the riskless asset is implicitly given by $\hat{\omega}_0 = 1 - \hat{\omega}'1$.

[8] The plug-in approach ignores *parameter estimation uncertainty,* i.e., the affect of sampling variation in parameter estimates on the portfolio choice problem.

squared error loss:

$$L(\hat{\beta}, \beta^*) = (\hat{\beta} - \beta^*)^2.$$

Under squared error loss, the corresponding risk function is the well-known *mean squared error* criterion:

$$RISK(\hat{\beta}, \beta^*) = \mathbb{E}_{F_{\hat{\beta}}}\left[(\hat{\beta} - \beta^*)^2\right].$$

For the portfolio allocation problem of interest in this paper, the investor's *utility function* provides the basis for an alternative, economically motivated loss function. The expected utility associated with an arbitrary allocation ω is given by

$$\begin{aligned} EU(\omega) &\equiv \mathbb{E}_{F_{R_e}}\left[U(W)\right] \\ &= \int \left[U(W_0(R_f + \omega' R_e))\right] dF_{R_e}. \end{aligned} \tag{4}$$

An economic loss function may then be defined as the random variable

$$L(\hat{\omega}, \omega^*) = EU(\omega^*) - EU(\hat{\omega}), \tag{5}$$

where ω^* satisfies (3). In light of the optimality of ω^*, the loss function (5) is nonnegative and achieves zero only when $\hat{\omega} = \omega^*$. The term $EU(\omega^*)$ does not depend upon $\hat{\omega}$. Consequently, it may be dropped without affecting risk comparisons among alternative estimators. The resulting loss function carries the natural interpretation that the loss associated with an allocation $\hat{\omega}$ is minus the corresponding expected utility:

$$L(\hat{\omega}) = -EU(\hat{\omega}). \tag{6}$$

Throughout the remainder of the paper, (6) will serve as the relevant loss function.[9]

To compute the risk of $\hat{\omega}$, a second expectation is taken with respect to the distribution of the estimator $\hat{\omega}$:

$$RISK(\hat{\omega}) = -\mathbb{E}_{F_{\hat{\omega}}}\left[EU(\hat{\omega})\right] \tag{7}$$

The risk measure (7) involves a double expectation over the investor's utility function. This suggests that the risk of an estimator carries an alternative interpretation as the expected

[9]Although ω^* has been dropped as an argument from (6), $L(\hat{\omega})$ remains a function of the underlying distribution of excess returns F_{R_e}. For a parametric family $F_{R_e}(r_e; \theta)$, writing $L(\hat{\omega}, \theta)$ indicates this dependence explicitly. Similar comments apply to the risk of an estimator, defined subsequently.

utility of a certain random payoff. To formalize this notion, consider the $2N \times 1$ random vector $Z = (\hat{\omega}', R_e')'$ with CDF F_Z. Define the random variable $\tilde{R}_e \equiv \hat{\omega}' R_e = g(Z)$, with CDF $F_{\tilde{R}_e}$. Under the economic loss function (6), the risk (7) of an arbitrary estimator $\hat{\omega}$ is

$$\begin{aligned} RISK(\hat{\omega}) &= -\int \left(\int U(W_0 \left[R_f + \hat{\omega}' R_e \right]) dF_{R_e} \right) dF_{\hat{\omega}} \\ &= -\int U \left(W_0 \left[R_f + g(Z) \right] \right) dF_Z \\ &= -\mathbb{E}_{F_{\tilde{R}_e}} U \left(W_0 \left[R_f + \tilde{R}_e \right] \right). \end{aligned} \tag{8}$$

The random variable $\tilde{R}_e$ carries an interpretation as the excess return on an asset associated with the estimator $\hat{\omega}$. The following proposition summarizes:

Proposition 1 *Under the economic loss function (6), the risk of an arbitrary estimator $\hat{\omega}$ is equivalent to minus the expected utility achieved when all wealth is invested in an asset held over $T+1$ periods paying an excess return equal to $\tilde{R}_e$.*

I do not claim that the connection established by Proposition 1 is particularly novel. Indeed, many existing empirical studies report 'expected utility values' or 'certainty equivalent returns' associated with various portfolio estimators. Either implicitly or explicitly, such studies rely upon the connection expressed formally in Proposition 1. The primary aim of the present paper is to illustrate the power embedded in this simple connection. Put simply, a large body of financial economic theory may be applied to address *statistical* questions related to the performance of portfolio estimators.

For intuition regarding Proposition 1, imagine that *prior to* obtaining a realization of the random sample R_e^T, an investor allocates 100% of wealth to a hypothetical fund manager, who will return money to the investor after $T+1$ periods. The estimator $\hat{\omega}$ serves as an 'investment rule' that is employed by the fund manager, in the sense that the manager waits T periods, observes the sample realization, and makes an allocation among the N (true) risky financial assets in accordance with $\hat{\omega}$. The investor's ultimate random return depends on two components: 1.) the realization of $\hat{\omega}$ that determines the fund manager's allocation to the actual financial assets; and 2.) the subsequent realization of excess returns on the actual financial assets.

As a concrete example, consider the trivial estimator $\hat{\omega}_0$ that allocates all wealth to the risk-free asset. By Proposition (1), the risk of this estimator is equivalent to minus the expected utility earned by a 100% position in the risk-free asset, regardless of the realization of sample returns. Since $\hat{\omega}_0$ earns a gross return of exactly R^f with no uncertainty, the risk of the estimator $\hat{\omega}_0$ is simply $-U(W_0 R_f)$.[10] The distribution of the excess return $\tilde{R}_e$ associated with this 'ultra-conservative' estimator coincides precisely with the distribution of one of the underlying tradeable asset returns (namely the risk-free asset). In general, however, $\tilde{R}_e$ need not follow the same parametric family as R_e, due to the influence of $\hat{\omega}$.

1.4 Existence Issues

A technical, but important, issue is whether the risk of a portfolio estimator exists, in the sense that $RISK(\hat{\omega})$ is finite. Proposition 1 implies that the existence of $RISK(\hat{\omega})$ is equivalent to the existence of the expected utility of an investment in an asset associated with $\hat{\omega}$ that pays excess return $\tilde{R}_e$. Under mean-variance preferences, the existence of the mean and covariance matrix for $\tilde{R}_e$ is sufficient to ensure the existence of $RISK(\hat{\omega})$. Under alternative preference specifications, the existence of $RISK(\hat{\omega})$ can be a more subtle issue. A particular concern is that $RISK(\hat{\omega}) = -\infty$, which essentially implies that the estimator $\hat{\omega}$ is 'arbitrarily bad.'

Unfortunately, the possibility that $RISK(\hat{\omega}) = -\infty$ is quite real under constant relative risk aversion (CRRA) utility, a common assumption in applied work. The essence of the problem is that CRRA utility tends to $-\infty$ as wealth tends to zero. Consequently, $RISK(\hat{\omega}) = -\infty$ for any estimator that results in negative terminal wealth with positive probability. This can occur even if returns exhibit limited liability when the support of $\hat{\omega}$ is not appropriately constrained. I defer more detailed discussion of this relatively technical issue to a brief Appendix to the paper. In what follows, it is implicitly assumed that all estimators under consideration possess finite risk.

2 Optimal Linear Combination of Portfolio Estimators

Suppose that a 'basis set' of $J > 1$ alternative estimators of the optimal portfolio are available, indexed as $\hat{\omega}_j$ for $j = 1, ..., J$. Consider forming a linear combination of the basis set of estimators. It is convenient for this purpose to collect the estimators $\hat{\omega}_j$ into an $N \times J$ matrix

[10]Note that $\hat{\omega}_0$ possesses a probability mass function rather than a PDF.

$\hat{\Omega}$ as follows:

$$\hat{\Omega} = \left[\begin{array}{ccc} \hat{\omega}_1 & ... & \hat{\omega}_J \end{array} \right].$$

A new estimator formed as a linear combination of the basis set takes the form

$$\hat{\omega}^{\alpha}(R_e^T) = \hat{\Omega}(R_e^T)\alpha(R_e^T), \tag{9}$$

where $\alpha = (\alpha_1, ..., \alpha_J)'$ is a $J \times 1$ vector of combining weights.

The notation in (9) emphasizes the fact that both α_j and $\hat{\Omega}$ are permitted to be functions of the random sample of excess returns. This permits data-driven combinations of the base set of estimators. Of course, this 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 estimators is itself a portfolio estimator, justifying the notation.[11]

Of central interest in this paper is the *optimal* linear combination of the base set of estimators that minimizes risk. The optimal combining weights α^* solve the problem:

$$\min_{\alpha} RISK(\hat{\omega}^{\alpha}) \tag{10}$$

The problem (10) is a statistical problem that involves identifying the linear combination of underlying estimators that minimizes risk as defined by statistical decision theory. Under the economic loss function (6), this statistical problem possesses an important *financial* interpretation. To illustrate, let $F_{\hat{\omega}^{\alpha}}$ represent the CDF of the combined portfolio estimator $\hat{\omega}^{\alpha}$. In parallel with the development of Proposition 1, define the $2N \times 1$ random vector $Z \equiv ((\hat{\omega}^{\alpha})', R_e')'$ with CDF F_Z and let $\tilde{R}_e \equiv \hat{\Omega}' R_e = g(Z)$. We have

$$\begin{aligned} \min_{\alpha} RISK(\hat{\omega}^{\alpha}) &= \min_{\alpha} - \int \left(\int U\left(W_0\left[R_f + (\hat{\omega}^{\alpha})' R_e\right]\right) dF_{R_e} \right) dF_{\hat{\omega}^{\alpha}} \\ &= \max_{\alpha} \int U\left(W_0\left[R_f + \alpha' g(Z)\right]\right) dF_Z \\ &= \max_{\alpha} \mathbb{E}_{F_{\tilde{R}_e}} U\left(W_0\left[R_f + \alpha' \tilde{R}_e\right]\right). \end{aligned} \tag{11}$$

[11]Formally, this follows from the fact that arithmetic operations involving measurable functions result in a measurable function.

The $J \times 1$ random vector $\tilde{R}_e$ carries the interpretation of excess returns on a set of J assets, where the j-th asset pays an excess return equal to $\hat{\omega}_j' R_e$. Using the trivial estimator $\hat{\omega}_0$, the vector of combining weights α may be interpreted as portfolio weights on these assets. Specifically, define $\alpha_0 \equiv 1 - 1_J' \alpha$. The estimator $\alpha_0 \hat{\omega}_0 + \hat{\omega}^\alpha$ is an estimator of the optimal portfolio such that $\sum_0^J \alpha_j = 1$. The estimator $\alpha_0 \hat{\omega}_0 + \hat{\omega}^\alpha$ consequently constitutes a "portfolio of portfolio estimators," where $\alpha_0 \equiv 1 - 1_J' \alpha$ represents the portfolio weight on the 'ultraconservative' asset that allocates all wealth to the risk-free bond. The following proposition summarizes:

Proposition 2 *Let $\hat{\omega}_j$ for $j = 1, ..., J$ represent alternative estimators of the optimal portfolio. Under the economic loss function (6), the problem (10) of finding the risk-minimizing linear combination of estimators $\hat{\omega}_j$ is equivalent to a canonical portfolio optimization problem where the investment options include a risk free asset paying gross return R^f and J risky assets paying excess returns $\tilde{R}_e$.*

For intuition regarding Proposition 2, suppose that an investor has the opportunity to invest funds with J different fund managers. Each manager will invest these funds across the same set of assets and each employs a strategy in accordance with a particular estimator $\hat{\omega}_j$. Finally, an additional ultraconservative manager is available who will allocate all wealth to the risk-free asset. Proposition 2 states that the optimal combination problem boils down to how this hypothetical investor should allocate wealth among the various fund managers.

3 Specific Results Regarding the Combination of Portfolio Estimators

The equivalence between the canonical asset allocation problem and the problem of forming a linear combination of portfolio estimators illustrates why combining estimators may improve performance. The same benefits of diversification that motivate investment across multiple assets suggest that combining multiple estimators of the unknown optimal portfolio may improve statistical performance.

While simple, Proposition 2 provides a powerful tool for characterizing optimal combinations of multiple portfolio estimators in a general setting. In particular, it reveals that the statistical problem of forming an optimal (linear) combination of estimators under an economic loss function is equivalent to a canonical portfolio problem in financial economics. This means that virtually every standard result from classic portfolio analysis may be re-interpreted as a statement regarding the optimal combination of a set of portfolio estimators. The remainder of this section illustrates by providing a number of specific results regarding the optimal combination of portfolio estimators. In each case, the result follows immediately from standard results in portfolio theory upon application of Proposition 2.[12]

3.1 The General Case

The solution to the combination problem is implicitly determined by the first-order conditions for the equivalent portfolio allocation problem:

$$\alpha^* \equiv \alpha \text{ s.t. } \mathbb{E}\left[U'\left(W_0[R_f + \alpha'\tilde{R}_e]\right)\tilde{R}_e\right] = 0. \tag{12}$$

In general, the optimal portfolio of estimated portfolios depends upon the distributional properties of excess returns, the distributional properties of the estimators of the optimal portfolio, investor preferences, and initial wealth.

3.2 Linear Risk Tolerance, Two Fund Monetary Separation, and Wealth Effects

More explicit results are available upon restricting the class of utility functions. In particular, suppose that the investor's utility function exhibits linear risk tolerance (LRT), so that

$$T(y) = \nu + \gamma y \tag{13}$$

where the risk tolerance $T(y)$ is defined as the inverse of the Arrow-Pratt measure of absolute risk aversion $A(y) = -\frac{U''(y)}{U'(y)}$. Under linear risk tolerance, two fund monetary separation holds in the equivalent space of assets, such that the (relative) composition of the investor's optimal portfolio of risky assets does not vary with initial wealth. For different levels of wealth, the optimal portfolio takes the form of linear combinations between the risk-free asset and a single risky asset mutual fund. This gives rise to the following corollary:

[12] References that present and prove the standard portfolio results cited in this section include LeRoy and Werner (2000), Gollier (2004) and Ingersoll (1987).

Corollary 1 *Assume that the investor's utility function exhibits LRT, and consider a non-redundant set of $J > 0$ estimators of the optimal portfolio. The optimal combining weights that solve (10) satisfy*

$$\alpha_j^*(W_0) = (\frac{\nu}{W_0} + \gamma R_f)b_j, \tag{14}$$

where b_j is independent of W_0 and ν.

Since $\alpha_j^*(W_0)$ carries the interpretation of the optimal portfolio weight on the asset corresponding to the j-th estimator, $W_0\alpha_j^*(W_0)$ represents the optimal investment *amount* in this asset. Using (14), we have that

$$W_0\alpha_j^*(W_0) = (\nu + \gamma W_0 R_f)b_j, \tag{15}$$

so that the optimal investment amount is linear in wealth and two-fund separation holds.

Since relative risk aversion is defined as $RRA(y) = \frac{y}{T(y)}$, the special case $\nu = 0$ corresponds to constant relative risk aversion equal to $\frac{1}{\gamma}$. From (14), it is clear that the optimal combining weights do not depend on wealth in this case. For utility functions exhibiting increasing relative risk aversion, such as the exponential utility function, the proportion of wealth invested in the risky asset mutual fund will decrease as initial wealth increases. The opposite holds for utility exhibiting decreasing relative risk aversion. The following corollary summarizes:

Corollary 2 *For utility functions displaying constant relative risk aversion (CRRA), the optimal combining weights that solve (10) are independent of initial wealth. For utility functions displaying increasing (decreasing) relative risk aversion, the proportion of wealth invested in the risky asset mutual fund decreases (increases) as initial wealth increases.*

3.3 Optimally Shrinking a Portfolio Estimator

Consider the case of a single (nontrivial) estimator of the optimal portfolio. In this case, linear "combination" amounts to multiplying an estimator $\hat{\omega}$ by a scaling factor α, so that $\hat{\omega}^\alpha = \alpha\hat{\omega}$. An equivalent interpretation is that the estimator $\hat{\omega}$ is combined with the ultra-conservative estimator $\hat{\omega}_0$. For $0 \leq \alpha \leq 1$, the combined estimator takes the form of a *James-Stein shrinkage estimator* that shrinks toward the portfolio rule $\hat{\omega}_0$. The degree of shrinkage is governed by α, with $\alpha = 1$ representing no shrinkage and $\alpha = 0$ representing complete shrinkage to a 100% position in the risk-free asset.[13]

By Proposition 2, the optimal shrinkage factor α^* is equivalent to the optimal portfolio allocation to a single risky asset that pays an excess return of $\tilde{R}_e = \hat{\omega}' R_e$. Let μ represent the $N \times 1$ vector $\mathbb{E}[R_e]$ and $\mu_{\hat{\omega}} \equiv \mathbb{E}[\hat{\omega}]$. Under Assumption 1, the expected return on $\tilde{R}_e$ is $\tilde{\mu} \equiv \mathbb{E}[\tilde{R}_e] = (\mu_{\hat{\omega}})' \mu$. 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. This yields the following corollary regarding optimal shrinkage:

Corollary 3 *Let $\hat{\omega}$ be an arbitrary non-trivial portfolio estimator. The optimal shrinkage factor α^* associated with $\hat{\omega}$ is greater than, less than, or equal to zero if and only if $\tilde{\mu}$ is greater than, less than, or equal to zero, respectively. If, in addition, $\mu_{\hat{\omega}} = \omega^*$ so that $\hat{\omega}$ is unbiased and ω^* is nonzero, then $\tilde{\mu} > 0$ and $\alpha^* > 0$.*[14]

Corollary 3 implies that a portfolio estimator may be sufficiently biased so that complete shrinkage is optimal. This occurs whenever $\tilde{\mu}$ is exactly zero. If a portfolio estimator is *unbiased*, however, $\alpha^* > 0$ and complete shrinkage is *never* optimal.

If the risk premium $\tilde{\mu}$ is 'small,' then a useful approximation is available for the optimal shrinking factor:

$$\alpha^* \approx \frac{\tilde{\mu}}{\tilde{\sigma}^2} \frac{1}{RRA(W_0 R_f)}, \tag{16}$$

where $\tilde{\sigma}^2 \equiv \mathbb{E}\left(\tilde{R}_e - \tilde{\mu}\right)^2$. The approximation in (16) relies on a first-order Taylor expansion of α^* around zero. A 'small' risk premium consequently refers to a case where higher order terms of the Taylor expansion are negligible. Note that risk aversion enters the formula for α^* in an intuitive way: higher relative risk aversion leads to a lower value of α^*, and consequently stronger shrinkage toward the ultraconservative estimator $\hat{\omega}_0$.

Proposition 2 also delivers several comparative statics results regarding the optimal shrinking factor:

[13] I will refer to such an estimator as a "shrinkage" estimator and to α as a "shrinkage factor" even though α is not formally restricted to the interval $[0, 1]$.

[14] By the optimality of ω^*, $\tilde{\mu} \geq 0$ and will be strictly greater than zero so long as the optimal allocation includes some position in the risky assets.

Corollary 4 *Consider a change in attitude toward risk from $U_1(W)$ to $U_2(W)$. This decreases the optimal shrinking factor α^* for any nontrivial portfolio estimator $\hat{\omega}$ if and only if $U_2(W)$ is more risk-averse than $U_1(W)$ in the Arrow-Pratt sense. Let $\hat{\omega}$ be an arbitrary estimator of the optimal asset allocation, and assume that $U(\cdot)$ is a twice differentiable, strictly concave utility function. Then $\frac{\partial \alpha^*}{\partial W_0} >, =, < 0$ under decreasing relative risk aversion (DRRA), constant relative risk aversion (CRRA) and increasing relative risk aversion (IRRA), respectively.*

The first part of Corollary 4 states that the relation between risk aversion and the optimal shrinking factor observed in the approximation formula (16) extends to the general case. The second portion of Corollary 4 specializes Corollary 2 to the case of a single portfolio estimator. In this case, the partial derivative of the optimal shrinking factor with respect to wealth is determined by relative risk aversion.

3.4 Mean-Variance Preferences

The mean-variance approach to portfolio allocation remains dominant in practice, particularly when a large number of assets is entertained. Let μ and Σ represent the mean and covariance matrix of excess risky returns, respectively. Suppose that investor preferences may be represented in terms of the mean and variance of excess returns on the portfolio, i.e., as

$$U(\omega) = \omega'\mu - \frac{\gamma}{2}\omega'\Sigma\omega \tag{17}$$

where $\gamma > 0$ captures the risk-aversion of the investor. Conditions under which (17) is consistent with expected utility maximization include quadratic utility or elliptical return distributions (see, e.g., Ingersoll (1987)). The investor's problem (2) becomes

$$\max_{\omega} \omega'\mu - \frac{\gamma}{2}\omega'\Sigma\omega.$$

The solution is given by the well-known formula derived by Markowitz (1952, 1959):

$$\omega^* = \frac{1}{\gamma}\mu\Sigma^{-1}. \tag{18}$$

In practice, μ and Σ are unknown and hence ω^* must be estimated. Given a basis set of J estimators, the statistical problem of forming an optimal linear combination of these estimators takes the form

$$\min_{\alpha} RISK(\hat{\omega}^{\alpha}, \omega^{*}) = \min_{\alpha} -\int \left[(\hat{\omega}^{\alpha})' \mu - \frac{\gamma}{2} (\hat{\omega}^{\alpha})' \Sigma (\hat{\omega}^{\alpha}) \right] dF_{\hat{\omega}^{\alpha}}. \tag{19}$$

By Proposition 2, the problem (19) is equivalent to a standard mean-variance portfolio problem for suitably transformed asset returns:

$$\max_{\alpha} \alpha' \tilde{\mu} - \frac{\gamma}{2} \alpha' \tilde{\Sigma} \alpha,$$

where $\tilde{\mu} \equiv \mathbb{E}[\tilde{R}_e]$ and $\tilde{\Sigma} \equiv \mathbb{E}[(\tilde{R}_e - \tilde{\mu})(\tilde{R}_e - \tilde{\mu})']$.

The tools of mean-variance portfolio analysis immediately apply to the optimal combination of the set of estimators. In the new space of asset returns, where the random variation of returns includes both random variation in the estimation data and random variation in the holding period return, the J portfolio estimators generate a standard mean-variance efficient frontier. This provides additional intuition for why combining multiple estimators may be beneficial, since the basis set of estimators will typically lie inside the hyperbola that characterizes the efficient frontier. In such cases, the same expected return may be achieved at lower variance by forming a nontrivial portfolio of estimated portfolios.

A global minimum variance portfolio will exist, representing the linear combination of the J estimators yielding the minimum possible variance in the new asset space. Similarly, a tangency portfolio of portfolio estimators will exist. Assuming that a risk-free asset is available in the actual economy, a risk-free asset will also be available in the transformed portfolio problem (via the estimator $\hat{\omega}_0$). The minimum variance set is therefore characterized in mean-standard deviation space as the usual pair of rays with common intercept at R_f. A mutual-fund result obtains such that optimal combining weights for the J estimators 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 base set of estimators. The formula is of the same form as (18), except that the moments refer to asset returns in the transformed asset space. The following corollary summarizes:

Corollary 5 *Suppose investor preferences are represented by (17) and let $\hat{\omega}_j$ be portfolio estimators for $j = 1, ..., J$. Further assume that $\tilde{\Sigma}$ is nonsingular and that a risk-free asset is available. Then*

1. *(Minimum Variance Portfolio of Estimated Portfolios) The combining weights that deliver the minimum variance of portfolio returns in the new asset space are given by*

$$\alpha^{GMV} = \frac{\tilde{\Sigma}^{-1}1_J}{1_J'\tilde{\Sigma}^{-1}1_J}$$

2. *(Tangency Portfolio of Estimated Portfolios) The combining weights that characterize the tangency point between the minimum variance set and the risky-asset only minimum variance set in the new asset space are given by*

$$\alpha^{TGCY} = \left(\frac{1}{1_J'\tilde{\Sigma}^{-1}\tilde{\mu}}\right)\tilde{\Sigma}^{-1}\tilde{\mu}$$

3. *(Optimal Combining Weights) The optimal linear combination of the J portfolio decision rules is given by*

$$\alpha^* = \frac{1}{\gamma}\tilde{\Sigma}^{-1}\tilde{\mu}. \tag{20}$$

4. *(Mutual Fund Theorem) The optimal linear combination of the J portfolio estimators is proportional to the combining weights characterizing the tangency portfolio in the new asset space. Thus, the relative proportions allocated to each estimator in the optimal combination are independent of risk preferences.*

The condition that $\tilde{\Sigma}$ is nonsingular amounts to requiring that no redundant assets exist among the J new assets. A violation of this condition occurs, for example, 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 $\tilde{\Sigma}$ fails to be invertible.

The formula for the optimal combination in Corollary 5 is expressed in terms of the first and second moments of returns $\tilde{R}_e$. These moments implicitly involve first and second moments of the excess returns on the original assets and first and second moments for the joint distribution of the basis set of portfolio estimators. Consequently, the optimal combination of portfolio estimators in the mean-variance setting depends upon all of the following in addition to investor preferences: 1.) the expected excess return and covariances of actual traded assets; 2.) the expectation of each portfolio estimator; 3.) the covariances of estimated portfolio weights among pairs of assets for each portfolio estimator; and 4.) the covariances of estimated portfolio weights across pairs of portfolio estimators.

4 Feasible Approaches to Combination

Given a set of portfolio estimators, the optimal combining weights α^* depend on unknown properties of the distribution of asset returns. Consequently, combining weights must *themselves* be estimated. By virtue of Proposition 2, estimating the optimal combining weights is equivalent to estimating an optimal portfolio allocation over the set of returns $\tilde{R}_e$. Unfortunately, this implies that the same estimation risk issues that confound data-driven approaches to portfolio allocation will confound estimation of optimal combining weights.[15] An additional difficulty arises from the fact that the asset returns $\tilde{R}_e$ reflect positions initiated prior to the realization of sample data, and held through the entire sample period. In effect, the available data contain only a single realization of returns $\tilde{R}_e$.[16] This section proposes an estimation approach that overcomes this obstacle, and then discusses the relative merits of estimation approaches versus alternative, simpler combination schemes, such as equal-weighting.

4.1 A Plug-in Estimation Approach

Under Assumption 1, all random variation in $\tilde{R}_e$ is ultimately governed by the underlying joint distribution of asset returns F_{R_e}. One feasible approach to estimating α^* is to substitute an estimate $\hat{F}_{R_e}$ in place of the unknown distribution F_{R_e}. Using $\hat{F}_{R_e}$, it is possible to simulate a large sample of asset returns $\tilde{R}_e$. To do so, first simulate a vector of sample data R_e^T from $\hat{F}_{R_e}$, compute $\hat{\omega}_j$ for $j = 1, ..., J$, and form $\hat{\Omega}$. Next, simulate an additional holding-period return R_e from $\hat{F}_{R_e}$ and compute $\tilde{R}_e$ as $\hat{\Omega}' R_e$. Repeating this procedure T^* times yields a sample of returns $\tilde{R}_e^{T^*}$ that may be used to estimate α^*. I now discuss two implementation details: 1.) the choice of estimator $\hat{F}_{R_e}$ used to generate $\tilde{R}_e^{T^*}$; and 2.) the choice of portfolio estimator used to estimate α^* given $\tilde{R}_e^{T^*}$.

[15] There is a 'matryoshka doll' flavor to the problem of estimating optimal combining weights. Indeed, one might consider forming combinations of multiple estimators of the optimal combining weights, and so on, *ad infinitum*. Note that the dimension of the estimation problem associated with combining weights will often be smaller than that in the original asset space. For example, consider an asset allocation problem with 100 risky assets and consider forming a combination of two estimators in this setting. The dimension of the new space of asset returns is two, as opposed to 100.

[16] Technically, if the sample size is N, the holding period for returns $\tilde{R}_e$ is $N + 1$ periods, including the final random return that accrues after the portfolio estimates are realized.

4.1.1 Approaches to Estimating F_{R_e}

Simulating a sample of returns $\tilde{R}_e^{T^*}$ first requires an estimate of F_{R_e}. An obvious candidate for $\hat{F}_{R_e}$ is the empirical distribution function (EDF) of historical asset returns. This amounts to a nonparametric approach to estimating F_{R_e}. An appealing aspect of basing $\hat{F}_{R_e}$ on the EDF is that no distributional model is required. The EDF is a consistent estimator for the true distribution function of returns under mild conditions. Using the EDF, samples $\tilde{R}_e^{T^*}$ may be generated using a bootstrap procedure. This is particularly straightforward when returns are independent, as assumed in this paper.

The primary drawback of basing $\hat{F}_{\tilde{R}_e}$ on the empirical distribution is the curse of dimensionality: as N increases the estimation accuracy of the EDF quickly erodes. Alternatively, $\hat{F}_{\tilde{R}_e}$ may be based on a parametric model, such as the multivariate normal or multivariate student-t distributions. Posing a parametric model for excess 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.[17] Even if a parametric model for F_{R_e} 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, one might specify a factor model for the covariance matrix, or employ a shrinkage estimator as proposed by Ledoit and Wolf (2003).

4.1.2 Estimating the Optimal Combining Weights

Suppose that a simulated set of asset returns $\tilde{R}_e^{T^*}$ is available. The first order condition (12) characterizes the optimal combining weights, and provides an orthogonality condition that may be used as the basis for estimation. This approach amounts to an application of the nonparametric method of moments estimator proposed by Brandt (1999), specialized to the case of i.i.d. returns. The estimator takes the form:

$$\hat{\alpha} = \left\{ \alpha : \frac{1}{T^*} \sum_{i=1}^{T^*} U'\left(W_0[R_f + \alpha' \tilde{R}_{e,i}] \right) \tilde{R}_{e,i} = 0 \right\}. \tag{21}$$

[17]Recall that the conceptual asset returns 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 $\tilde{R}_e$ that are not well-modeled by, e.g., the multivariate normal or multivariate student-t.

For the special case of mean-variance preferences, equation (20) provides an analytic formula for the optimal combining weights as a function of the first and second moments of $\tilde{R}_e$. The optimal combining weights may therefore be estimated using plug-in estimates of these moments. This approach may be used as an approximation even in cases where mean-variance analysis is not strictly valid.

To minimize estimation error when a portfolio estimator is applied to $\tilde{R}_e^{T^*}$, the parameter T^* may be set to a very large value. By doing so, one may ensure that extremely precise estimates are obtained. It is important to bear in mind, however, that what is estimated with high precision is $\alpha^*_{\hat{F}}$, and not α^*, where the discrepancy arises from the fact that asset returns are simulated using $\hat{F}_{R_e}$ rather than the true distribution F_{R_e}. Intuitively, the hope is that if $\hat{F}_{R_e}$ provides a reasonably good estimate of F_{R_e}, then $\alpha^*_{\hat{F}}$ will provide an accurate estimate of α^*.

4.2 Simple Averaging of Estimators

Rather than attempt to estimate optimal combining weights, an *ad hoc* approach to combination is to simply average the various estimators. This is equivalent to setting $\hat{\alpha} = (1/J)1_J$. The primary motivation for such an approach is the limited ability to obtain a precise estimate of $F_{\tilde{R}_e}$ for a typical application with a relatively large set of assets and limited historical data. If the estimator $\hat{F}_{\tilde{R}_e}$ is extremely noisy, this will result in a large degree of statistical uncertainty regarding the optimal combination of the portfolio decision rules. Simply averaging the various portfolio decision rules under consideration may be viewed as a type of shrinkage estimator (for the optimal combining weights) where shrinkage to the $1/J$ weighting is complete. Intuitively, if the truly optimal set of combining weights is not dramatically different from equal weights, then this *ad hoc* estimator may perform well since, even though it suffers from some bias, it exhibits no variance.[18] 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)).

[18] This basic intuition is based on MSE loss which may be decomposed into variance plus squared bias. Of course, in the present setting the loss function is utility-based as opposed to MSE.

4.3 Is Combination Beneficial in Practice?

Ultimately, the extent to which combining estimators improves estimation performance is an empirical question. The existing empirical literature bears witness to the practical benefits of combination. In the context of mean-variance preferences, DeMiguel, Garlappi and Uppal (2007) assess the performance of a wide array of estimated portfolio rules against a 'naive' $1/N$ benchmark, and find that few, if any, rules consistently outperform the benchmark. Tu and Zhou (2010) show that estimators formed as combinations of the $1/N$ rule and other 'sophisticated' portfolio rules can outperform the $1/N$ rule.[19]

It is important to point out that while DeMiguel, Garlappi and Uppal (2007) describe the $1/N$ rule as 'naive', this estimator may be viewed as a type of combined estimator. In particular, the $1/N$ rule combines N different trivial estimators, the i-th estimator allocating 100% of investible wealth to the i-th asset irrespective of the sample outcome. Rather than attempt to estimate optimal combining weights, the $1/N$ exploits the simple averaging idea. From this perspective, recent empirical evidence in the mean-variance setting supports the notion that combined estimators offer an effective means to improve estimation performance. There remains; however, some controversy in the literature regarding what form of combined estimator is superior. Indeed, the answer to this question is likely to vary depending on the specific application.

5 Application: Shrinking a Portfolio Estimator Under CARA Preferences

Most evidence regarding the performance of combined portfolio estimators resides within the mean-variance setting. This paper develops general results regarding combined estimators. As a concrete application of some of the concepts addressed in the paper, I conduct a simulation analysis based on a portfolio choice problem outside of the mean-variance paradigm.

[19]The estimators Tu and Zhou (2010) combine with the $1/N$ rule include estimators proposed by Jorion (1986), MacKinlay and Pastor (2000) and Kan and Zhou (2007). The Kan and Zhou (2007) estimator is *itself* a combined estimator based on the usual maximum likelihood estimator and the estimated global minimum variance portfolio. Kirby and Ostdiek (2010) propose mean-variance timing strategies with low turnover and find that these outperform a $1/N$ rule, even in the presence of transactions costs.

5.1 A Simple Portfolio Problem

I consider a simple portfolio choice problem involving an allocation between a single risky asset (a stock index) and a risk-free bond. The portfolio problem is static and the return on the risky asset is independently and identically distributed, consistent with Assumption 1. Investor preferences adhere to a CARA utility function of the form

$$U(W) = -e^{-\gamma W}. \tag{22}$$

When investor preferences follow (22) and returns are normally distributed, it is well known that optimal portfolios are consistent with mean-variance analysis. The simulation analysis assumes, by contrast, that log returns follow a skew-t distribution (see, e.g., Azzalini and Capitanio (2003)). This assumption permits negative skewness and heavy tails, two features that empirically characterize stock index returns.

The 'base' estimator of the optimal allocation is a nonparametric 'method of moments' estimator in the spirit of Brandt (1999) that chooses the portfolio weight ω to set the empirical first-order condition to zero.[20] Under CARA utility this estimator takes the form:

$$\hat{\omega} \equiv \left\{ \omega : \frac{1}{T} \sum_{i=1}^{T} e^{-\gamma W_0 \left[R_f + \omega R_{e,i}\right]} \left(R_{e,i}\right) = 0 \right\}. \tag{23}$$

Under weak restrictions, the estimator $\hat{\omega}$ is consistent for the optimal allocation ω^* and asymptotically normal (see, e.g., Brandt (1999)). An appealing feature of this nonparametric estimator is that it requires no distributional assumption for (log) returns.

5.2 Shrinking Toward the Ultraconservative Estimator

As discussed earlier in the paper, a special case of forming a combination of portfolio estimators involves 'shrinking' a particular estimator toward the ultraconservative estimator $\hat{\omega}_0$ that allocates all wealth to the risk-free asset, irrespective of the sample data. Such shrinkage estimators take the form:

$$\hat{\omega}^{\hat{\alpha}} = \hat{\alpha} \cdot \hat{\omega}, \tag{24}$$

where $\hat{\alpha}$ represents an estimate of the optimal shrinking factor α^* that satisfies (12).

[20] This estimator may also be viewed as an 'M-estimator' that maximizes sample utility.

I consider several approaches to estimating α^*. The first estimator is based on the simulation procedure described in Section 4.1 of the paper. To generate asset returns corresponding to $\hat{\omega}$, I use the bootstrap procedure described in Section 4.1 with the EDF of excess returns serving as $\hat{F}_{R_e}$. Given simulated returns $\tilde{R}_{e,i}$ for $i = 1, \ldots, T^*$, the optimal shrinking factor is estimated using the method of moments estimator:

$$\hat{\alpha}_1 = \left\{ \alpha : \frac{1}{T^*} \sum_{i=1}^{T^*} e^{-\gamma W_0 \left[R_f + \omega \tilde{R}_{e,i} \right]} \left(\tilde{R}_{e,i} \right) = 0 \right\}.$$

The second shrinkage estimator applies the simple averaging concept discussed in Section 4.2 of the paper. Rather than attempt to estimate the optimal shrinkage factor at all, this estimator simply takes $\hat{\alpha}_2 = 0.5$, which is equivalent to averaging $\hat{\omega}$ and $\hat{\omega}_0$. Of course, it is highly unlikely that the truly optimal shrinkage factor is 0.5; however, this approach avoids any sampling variability associated with estimating the optimal shrinking factor. Finally, I consider a 'hybrid' shrinkage estimator $\hat{\alpha}_3$ defined as $\frac{\hat{\alpha}_1 + \hat{\alpha}_2}{2}$. This estimator applies a second layer of shrinkage, in which the estimated optimal shrinking factor $\hat{\alpha}_1$ is itself shrunk toward the value 0.5. The shrinkage estimators based on $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\alpha}_3$ are denoted $\hat{\omega}^{\hat{\alpha}_1}$, $\hat{\omega}^{\hat{\alpha}_2}$ and $\hat{\omega}^{\hat{\alpha}_3}$, respectively.

5.3 Calibration Details

Simulated risky stock returns are calibrated by fitting a skew-t distribution to quarterly log excess returns on the S&P 500 Index over the period 1952:1 - 2008:4. Figure 1 presents a histogram for log excess returns, along with density plots corresponding to the fitted skew-t distribution and a reference normal distribution. The figure illustrates that the skew-t distribution provides a superior fit relative to the normal distribution, better capturing the negative skewness and leptokurtosis apparent in returns. The risk-free rate is calibrated to 1% (quarterly), a value close to the empirical average of the three-month T-bill rate over the 1952-2008 period. Initial wealth is normalized to unity, while the risk aversion parameter γ is set to three. The analysis considers sample sizes ranging from 60 (5 years of quarterly data) to 1200 (100 years of quarterly data).

The skewness and leptokurtosis of the skew-t distribution imply that the optimal allocation α^* under CARA preferences (22) differs from that prevailing under mean-variance preferences

(18). For the calibrated portfolio problem, the numerically computed optimal portfolio under CARA utility is approximately 78% stock, while the optimal portfolio under mean-variance preferences is approximately 83% stock.[21] This difference between the two optimal portfolios is small, but nontrivial.

5.4 Simulation Results

Figure 2 presents density plots for the method of moments estimator $\hat{\omega}$ at various sample sizes. The variability of the estimated optimal allocation to risky stock is striking. For a sample size of 60, the estimator exhibits extreme variability. Indeed, at this sample size the 5% quantile of the simulated allocations to risky stock is negative (-8.7%), while the 95% quantile is in excess of 200% (213.6%). For a sample size of 240 (20 years of quarterly data), the estimated allocations to stock continue to be relatively volatile, with 5% and 95% quantiles of approximately 34% and 135%, respectively. At a very large sample size of 1200, the estimator $\hat{\omega}$ begins to cluster closely around the optimal allocation of 78% stock.

Table 1 presents additional statistics regarding $\hat{\omega}$, including the bias, standard deviation, and skewness for each sample size. At small sample sizes the estimator is slightly upward biased. As expected, this bias disappears at the largest sample sizes. At smaller sample sizes the estimator is positively skewed, while at larger sample sizes the distribution is roughly symmetric, consistent with an asymptotically normal limiting distribution.

Figure 3 illustrates how shrinkage affects the estimated portfolio weights. The figure provides density plots for $\hat{\omega}$, $\hat{\omega}^{\hat{\alpha}_1}$, $\hat{\omega}^{\hat{\alpha}_2}$ and $\hat{\omega}^{\hat{\alpha}_3}$ at various sample sizes. At lower sample sizes, the densities of the shrinkage estimators peak leftward of the density of $\hat{\omega}$, as the estimated allocation to risky stock is pulled toward zero. As the sample size increases, the data-driven shrinkage estimator $\hat{\omega}^{\hat{\alpha}_1}$ exhibits less shrinkage. This is due to the fact that the variability of $\hat{\omega}$ decreases as the sample size increases, which reduces the optimal degree of shrinkage. By contrast, the density of the naive shrinkage estimator $\hat{\omega}^{\hat{\alpha}_2}$ continues to peak around an allocation of 40% stock even at large sample sizes. This estimator is clearly asymptotically biased and inconsistent. The hybrid estimator $\hat{\omega}^{\hat{\alpha}_3}$ is also inconsistent, but with a lower asympototic bias

[21] The optimal portfolio under CARA preferences is computed by estimating the optimal portfolio for a simulated sample of size 2,000,000.

relative to $\hat{\omega}^{\hat{\alpha}_2}$.

Table 1 presents the certainty equivalent return (CER), expressed in quarterly basis points, corresponding to each estimator. The CER values are based on 400,000 simulated out-of-sample returns. For the lowest sample size of 60, the method of moments estimator $\hat{\omega}$ performs very poorly, earning a CER of around *negative* 7.5 basis points. Consequently, a hypothetical investor would be willing to pay up to 7.5 basis points per quarter to avoid investing using the estimator $\hat{\omega}$ and alternatively allocate all wealth to the risk-free bond. The data-driven shrinkage estimator $\hat{\omega}^{\hat{\alpha}_1}$ performs little better than the raw method of moments estimator, earning a CER of approximately -3.8 basis points. By contrast, the naive and hybrid shrinkage estimators $\hat{\omega}^{\hat{\alpha}_2}$ and $\hat{\omega}^{\hat{\alpha}_3}$ perform substantially better, earning CERs of approximately 35 and 24 basis points, respectively.

To understand this pattern of results, note that a small sample size hampers efforts to estimate the optimal shrinkage factor. Figure 4 shows density plots for the data-driven shrinkage estimator $\hat{\alpha}_1$ at sample sizes of 60, 240, and 1200. For a sample size of 60, the estimated shrinkage factor is both highly variable and left-skewed. This generates extreme variability in the portfolio estimator $\hat{\omega}^{\hat{\alpha}_1}$ at low sample sizes, and poor performance. By contrast, the naive shrinkage estimator $\hat{\alpha}_2$ avoids any sampling variability. When the sample size is 60, the optimal shrinkage factor is approximately 0.64.[22] Despite the fact that the simple shrinkage estimator $\hat{\alpha}_2$ 'overshrinks' relative to the true optimum, its lack of sample variability delivers a large performance improvement.

When the sample size increases to 120, the performance of all estimators improves, as expected. The improvement is dramatic for both the raw estimator $\hat{\omega}$ and the data-driven shrinkage estimator $\hat{\omega}^{\hat{\alpha}_1}$, as these now earn CERs of approximately 30 and 35 basis points, respectively. For this sample size, $\hat{\omega}^{\hat{\alpha}_2}$ and $\hat{\omega}^{\hat{\alpha}_3}$ continue to outperform $\hat{\omega}$, although the CER differential is smaller, on the order of 4-5 basis points quarterly. For large sample sizes, the best performing estimator is $\hat{\omega}$. Intuitively, since the estimator $\hat{\omega}$ is consistent, the optimal shrinking factor approaches one as the sample size increases. With abundant sample data, there is little benefit to shrinkage, and estimators that attempt to shrink do not outperform $\hat{\omega}$. Simulation evidence (not separately reported) shows that for sample sizes of 240 and above the optimal

[22]This approximation is computed via simulation. Detailed results are available from the author upon request.

shrinkage factor is very close to one.[23]

Figure 5 summarizes the performance of the various estimators. The data-driven shrinkage estimator $\hat{\omega}^{\hat{\alpha}_1}$ performs poorly relative to the other shrinkage estimators until the sample size exceeds 240 (20 years of data), and at this point it is already inferior to the raw estimator $\hat{\omega}$. The simple averaging estimator $\hat{\omega}^{\hat{\alpha}_2}$ performs best at very small sample sizes, but significantly underperforms the other estimators for large sample sizes. The hybrid estimator $\hat{\omega}^{\hat{\alpha}_3}$, on the other hand, performs consistently well across a range of sample sizes. At small sample sizes (fewer than 120 observations) it dramatically outperforms both $\hat{\omega}$ and $\hat{\omega}^{\hat{\alpha}_1}$, while at larger sample sizes it either outperforms, or only slightly underperforms, the estimator $\hat{\omega}$. Overall, the results suggest that incorporating some form of shrinkage or prior information regarding combining weights is key to realizing the benefits of diversification available through combination.

The portfolio problem considered here is simple, involving only a single risky asset. Additionally, the application only involves a single 'basis' estimator (the method of moments estimator), while other applications might consider combining many estimators. Nevertheless, the results nicely parallel recent empirical findings in the mean-variance setting. First, an unadjusted estimator suffers severely from estimation risk at smaller sample sizes. Second, a 'naive' shrinkage approach that simply halves the method of moments estimate is difficult to beat at small sample sizes. In the mean-variance setting, DeMiguel, Garlappi and Uppal (2007) find that estimated portfolio strategies do not outperform a $1/N$ allocation rule unless the sample size is quite large. Finally, a 'hybrid' estimator that averages the naive and data-driven shrinkage factors performs consistently well. In a similar vein, Tu and Zhou (2010) find that combining the $1/N$ rule with estimated portfolio rules, some of which already involve combination or shrinkage, leads to performance improvements in the mean-variance setting.

[23] It should also be noted that CER values for the estimators $\hat{\omega}^{\hat{\alpha}_1}$ and $\hat{\omega}^{\hat{\alpha}_3}$ are conservative, as these are based on a bootstrap size of 1500 used to generate returns $\tilde{R}_e$ (see Section 4.1). The bootstrap size could be increased to, e.g., 10,000 to improve the performance of these estimators slightly, at the cost of a significant increase in computational time.

6 Conclusion

The paper considers forming optimal combinations of an arbitrary set of portfolio estimators under an economic loss function. The optimal combination problem is shown to be formally equivalent to a standard asset allocation problem under a transformed set of asset returns. Using this fundamental equivalence, the paper states a variety of general results characterizing optimal combinations.

Unfortunately, optimal combining weights are latent and must be estimated. The paper develops an approach for estimating optimal combining weights that exploits the equivalence between the combination problem and a standard asset allocation problem. Because attempts to estimate optimal combining weights suffer from estimation risk, simple alternatives such as equal-weighting estimators may perform better. This conjecture is supported by the results of a simulation exercise involving a stock versus bond decision under CARA preferences.

I close with two remarks. First, this paper operates under the restrictive assumption that risky asset returns are independently distributed. This assumption is unlikely to be valid in practice. The volatility of asset returns clusters in time, and there is empirical evidence of time-variation in expected returns, although the extent and magnitude of such predictability remains hotly debated. Analyzing the optimal combination of estimators in a dynamic portfolio choice setting with predictable returns represents an interesting avenue for future research. There are some significant differences between that setting and the one addressed in this paper. Among these, the object of estimation in a dynamic portfolio choice problem is a policy function, i.e., a mapping from state variables to portfolio positions. Further, the introduction of conditioning information suggests that optimal combining weights for policy estimators might also vary with the economic state.

Second, this paper follows a large number of recent studies in applying an economic loss function to evaluate portfolio estimators. While this approach has intuitive merit, it may not be suitable for all applications. Investor utility functions generally translate to *asymmetric* loss functions. Asymmetric loss functions favor estimators that are biased in an appropriate direction. In some applications, a 'good' estimator under an economic criterion might be heavily biased. If the primary objective of an empirical study is to report a reasonable estimate of the optimal portfolio allocation, such heavily biased estimates may not be desirable.

A Appendix

A.1 Existence Issues under CRRA Preferences

This appendix provides additional discussion regarding the existence issues raised in Section 1.4 of the paper. Suppose that the investor's utility function takes the CRRA form:

$$U(W) = \frac{W^{(1-\gamma)}}{1-\gamma}, \tag{25}$$

where $W > 0, \gamma > 0$, and where $U(W) = \log(W)$ for the special case $\gamma = 1$. Under CRRA preferences, absolute risk aversion is declining in wealth and relative risk aversion is equal to the constant γ.[24] Note that $U(W)$ in (25) is only defined for $W > 0$ and tends to $-\infty$ as $W \to 0$. This implies that the risk of an estimator $\hat{\omega}$ will not exist whenever the support of $\tilde{R}_e$ extends beyond the interval $(0, \infty)$. In applications of CRRA preferences to problems involving choice under uncertainty, risky payoffs are typically assumed to possess distributions that satisfy this restriction, such as the lognormal distribution. When considering the risk of portfolio estimators, however, it is crucial to bear in mind that the relevant random payoff is $\tilde{R}_e$ rather than R_e. Unfortunately, restricting the support of the return distribution R_e to the interval $(0, \infty)$ does not, in general, confine the support of $\tilde{R}_e$ to the same interval, due to the influence of $\hat{\omega}$.

Consider a simple portfolio allocation between a single risky asset and a risk-free bond under CRRA preferences. Suppose that the support of R_e is bounded on the interval $[-0.5, 0.5]$. Absent estimation risk, this assumption is sufficient to guarantee the existence of expected utility for the optimal asset allocation ω^*. Now consider the risk of an arbitrary portfolio estimator $\hat{\omega}$ of ω^*. If the support of $\hat{\omega}$ extends beyond 2, so that the estimated optimal portfolio takes of position in the risky stock of more than 200% with positive probability, then negative final wealth outcomes are possible and $RISK(\hat{\omega})$ fails to exist. More generally, if the support of $\hat{\omega}$ is unbounded, then $RISK(\hat{\omega})$ fails to exist under CRRA preferences, since negative final wealth outcomes occur with positive probability.[25]

[24]The latter property makes CRRA utility popular in applied work, since it is consistent with stable interest rates and risk premia in the presence of secular growth in consumption and wealth.

[25]This holds so long as excess returns on the risky asset may be negative, which must be the case if there is no arbitrage and the risky asset earns a positive risk premium.

Even if the support of $\tilde{R}_e$ is restricted to $(0, \infty)$, so that final wealth is assured to be positive, $RISK(\hat{\omega})$ may still fail to exist. Geweke (2001) studies the existence of expected utility under CRRA preferences. By virtue of Proposition 1, results established by Geweke (2001) may equivalently be interpreted as results regarding the existence of the risk of an arbitrary portfolio estimator under a loss function based on CRRA preferences:

Corollary 6 *Consider an arbitrary estimator $\hat{\omega}$ of the optimal portfolio in problem (2) where the loss function is (6) with CRRA preferences. Let $Z \equiv \log(W) = \log\left(\tilde{R}_e\right)$ under the normalizations $R_f = 0$ and $W_0 = 1$. If $\gamma = 1$, then $RISK(\hat{\omega})$ exists so long as $\mathbb{E}(Z)$ exists. When $\gamma \neq 1$, then $RISK(\hat{\omega})$ exists if and only if the moment generating function (MGF) of Z, $MGF_Z(t)$, exists at $t = 1 - \gamma$. For general R_f and W_0, a similar result applies with $Z \equiv \log\left(W_0[R_f + \tilde{R}_e]\right)$.*

The question of whether existence of whether a portfolio estimator possesses finite risk is far from obvious, particularly under an economic loss function based on CRRA preferences. Under CRRA preferences, constraining a portfolio estimator may be crucial from a statistical perspective, i.e., irrespective of whether the constraint reflects true economic frictions with respect to investments. Unfortunately, appropriately constraining the estimator requires an auxiliary assumption regarding the support of the distribution of excess returns. Alternative preference specifications that have domain equal to the entire real line, such as constant *absolute* risk aversion (CARA), are more forgiving, since negative wealth outcomes are permitted.

References

Azzalini, A. and A. Capitanio, 2003, Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew-t Distribution, *Journal of the Royal Statistical Society, series B*, 65, 367-389.

Brandt, M.W., 1999, Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach, *Journal of Finance*, 54, 1609-1646.

Campbell, J.Y. and S. B. Thompson, 2008, Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *The Review of Financial Studies*, 21, 1509-1531.

DeMiguel, V., L. Garlappi, J. Nogales and R. Uppal, 2009, A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms, *Management Science*, forthcoming.

DeMiguel, V., L. Garlappi and R. Uppal, 2007, Optimal Versus Naive Diversification: How Inefficient Is the $1/N$ Portfolio Strategy? *The Review of Financial Studies*, forthcoming.

Fleming, J., C. Kirby, and B. Ostdiek, 2001, The Economic Value of Volatility Timing, *Journal of Finance*, 56, 329-352.

Fleming, J., C. Kirby, and B. Ostdiek, 2003, The Economic Value of Voltility Timing Using 'Realized' Volatility, *Journal of Financial Economics*, 67, 473-509.

Frost, P.A. and J. E. Savarino, 1986, An Empirical Bayes Approach to Portfolio Selection, *Journal of Financial and Quantitative Analysis*, 21, 293-305.

Garlappi, L., R. Uppal and T. Wang, 2007, Portfolio Selection with Parameter and Model Uncertainty: A Multi-prior Approach, *The Review of Financial Studies*, 20, 41-81.

Geweke, J., 2001, A Note on Some Limitations of CRRA Utility, *Economics Letters*, 71, 341-345.

Goldfarb, D. and G. Iyengar, 2003, Robust Portfolio Selection Problems, *Mathematics of Operations Research*, 28, 1-38.

Gollier, C., 2001, *The Economics of Risk and Time*, MIT Press.

Goyal, A. and I. Welch, 2008, A Comprehensive Look at the Empirical Performance of the Equity Premium Prediction, *The Review of Financial Studies*, 21, 1455-1508.

Guidolin, M. and A. Timmermann, 2007, Asset Allocation under Multivariate Regime Switching, *Journal of Economic Dynamics and Control*, 3503-3544.

Ingersoll, J.E., 1987, *Theory of Financial Decision Making*, Rowman & Littlefield Publishers, Inc.

Jagannathan, R., and T. Ma, 2003, Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps, *Journal of Finance*, 54, 1651-1684.

Jobson, J.D., and B. Korkie, 1980, Estimation of Markowitz Efficient Portfolios, *Journal of the American Statistical Association*, 75, 544-554.

Jorion, P., 1986, Bayes-Stein Estimation for Portfolio Analysis, *Journal of Financial and Quantitative Analysis*, 21, 279-292.

Jorion, P., 1992, Portfolio Optimization in Practice, *Financial Analysts Journal*, 48, 68-74.

Kan, R. and G. Zhou, 2007, Optimal Estimation for Economic Gains: Portfolio Choice with Parameter Uncertainty, *Journal of Financial and Quantitative Analysis*, 42, 621-656.

Kirby, C. and B. Ostdiek, 2010, It's All in the Timing: Simple Active Portfolio Strategies that Outperform Naive Diversification, Working paper.

Ledoit, O. and M. Wolf, 2003, Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection, *Journal of Empirical Finance*, 10, 603-621.

Ledoit, O. and M. Wolf, 2004, A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices, *Journal of Multivariate Analysis*, 88, 365-411.

LeRoy, S.F. and J. Werner, 2000, *Principles of Financial Economics*, Cambridge University Press.

MacKinlay, A.C. and L. Pastor, 2000, Asset Pricing Models: Implications for Expected Returns and Portfolio Selection, *The Review of Financial Studies*, 13, 883-916.

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

Markowitz, H., 1959, *Portfolio Selection, Efficient Diversification of Investment*, Basil Blackwell Ltd., New York.

Marquering, W. and M. Verbeek, 2004, The Economic Value of Predicting Stock Index Returns and Volatility, *Journal of Financial and Quantitative Analysis*, 39, 407-429.

Michaud, R. O., 1998, *Efficient Asset Management*, Harvard Business School Press, Boston.

Michaud, R.O., 1989, The Markowitz Optimization Enigma: Is Optimized Optimal?, *Financial Analyst Journal*, 45, 31-42.

Sharpe, W., 1963, A Simplified Model for Portfolio Analysis, *Management Science*, 9, 277-293.

Timmermann, A., 2006, Forecast Combinations, in *Handbook of Economic Forecasting, Volume 1 (Handbooks in Economics),*.G. Elliott, C.W.J. Granger and A. Timmermann, eds., North Holland.

Tu, J. and G. Zhou, 2010, Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies, *Journal of Financial Economics*, forthcoming.

West, K.D., H.J. Edison and D. Cho, 1993, A Utility-based Comparison of Some Models of Exchange Rate Volatility,*Journal of International Economics*, 35, 23-45.

Table 1
Simulation Evidence: Shrinking the Method of Moments Estimator

The table presents the bias, standard deviation, skewness and certainty equivalent return (CER, expressed in quarterly basis points) associated with the estimators $\hat{\omega}$, $\hat{\omega}^{\hat{\alpha}_1}$, $\hat{\omega}^{\hat{\alpha}_2}$ and $\hat{\omega}^{\hat{\alpha}_3}$ in the portfolio problem descried in Section 5.1 of the paper. The bias, standard deviation and skewness are based on estimates from 10,000 simulated datasets with corresponding sample size. CER results are based on 400,000 simulated out-of-sample returns, where each estimated portfolio weight from the initial 10,000 simulations is applied to 40 out-of-sample return draws. See Section 5.3 for calibration details regarding the simulation.

Panel A: N= 60

	Bias	St. Dev.	Skew.	CER
$\hat{\omega}$	0.11	0.70	0.80	-7.46
$\hat{\omega}^{\hat{\alpha}_1}$	-0.10	0.69	1.34	-3.83
$\hat{\omega}^{\hat{\alpha}_2}$	-0.34	0.35	0.80	35.23
$\hat{\omega}^{\hat{\alpha}_3}$	-0.22	0.52	1.18	23.73

Panel B: N= 120

	Bias	St. Dev.	Skew.	CER
$\hat{\omega}$	0.05	0.45	0.57	35.59
$\hat{\omega}^{\hat{\alpha}_1}$	-0.11	0.50	0.91	30.65
$\hat{\omega}^{\hat{\alpha}_2}$	-0.37	0.23	0.57	41.48
$\hat{\omega}^{\hat{\alpha}_3}$	-0.24	0.36	0.82	40.51

Panel C: N= 240

	Bias	St. Dev.	Skew.	CER
$\hat{\omega}$	0.03	0.31	0.39	49.25
$\hat{\omega}^{\hat{\alpha}_1}$	-0.09	0.37	0.54	44.06
$\hat{\omega}^{\hat{\alpha}_2}$	-0.38	0.15	0.39	43.48
$\hat{\omega}^{\hat{\alpha}_3}$	-0.23	0.26	0.50	47.48

Panel D: N= 480

	Bias	St. Dev.	Skew.	CER
$\hat{\omega}$	0.01	0.21	0.28	54.48
$\hat{\omega}^{\hat{\alpha}_1}$	-0.07	0.28	0.31	50.68
$\hat{\omega}^{\hat{\alpha}_2}$	-0.39	0.11	0.28	44.23
$\hat{\omega}^{\hat{\alpha}_3}$	-0.23	0.19	0.30	51.06

Panel E: N= 1200

	Bias	St. Dev.	Skew.	CER
$\hat{\omega}$	0.00	0.13	0.17	58.12
$\hat{\omega}^{\hat{\alpha}_1}$	-0.05	0.20	0.21	55.62
$\hat{\omega}^{\hat{\alpha}_2}$	-0.39	0.07	0.17	44.75
$\hat{\omega}^{\hat{\alpha}_3}$	-0.22	0.13	0.20	53.68

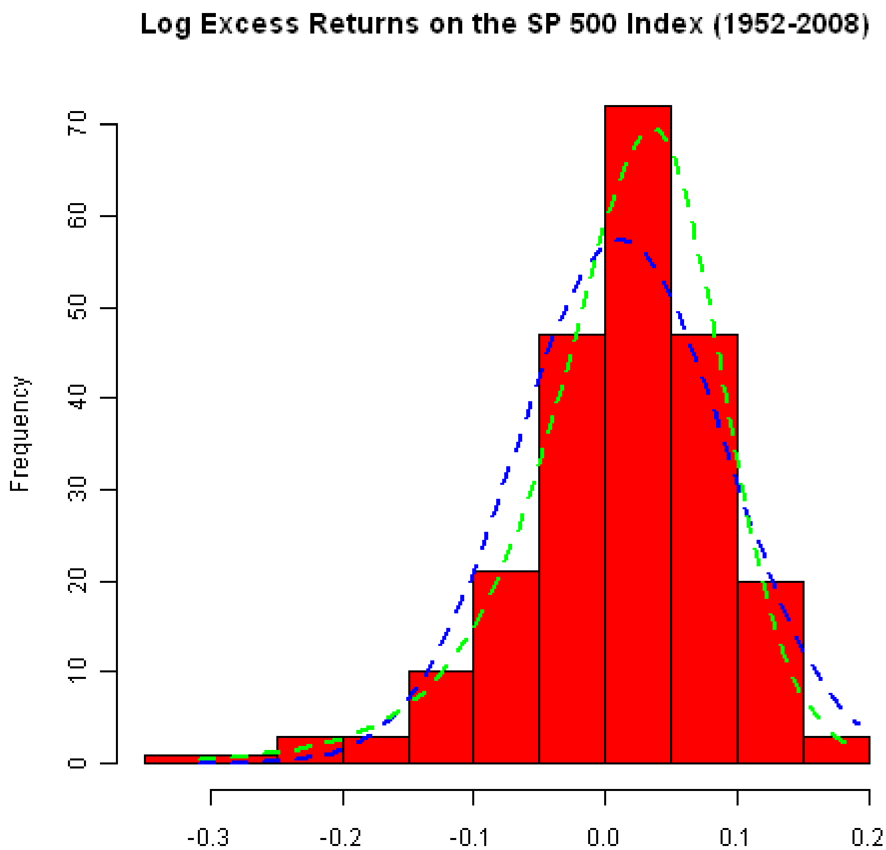

Figure 1: Histogram for quarterly excess returns on the S&P 500 over the sample period 1952 - 2008. The green dashed line provides a density plot for the fitted skew-t distribution, while the blue dashed line provides a reference density plot for the normal distribution.

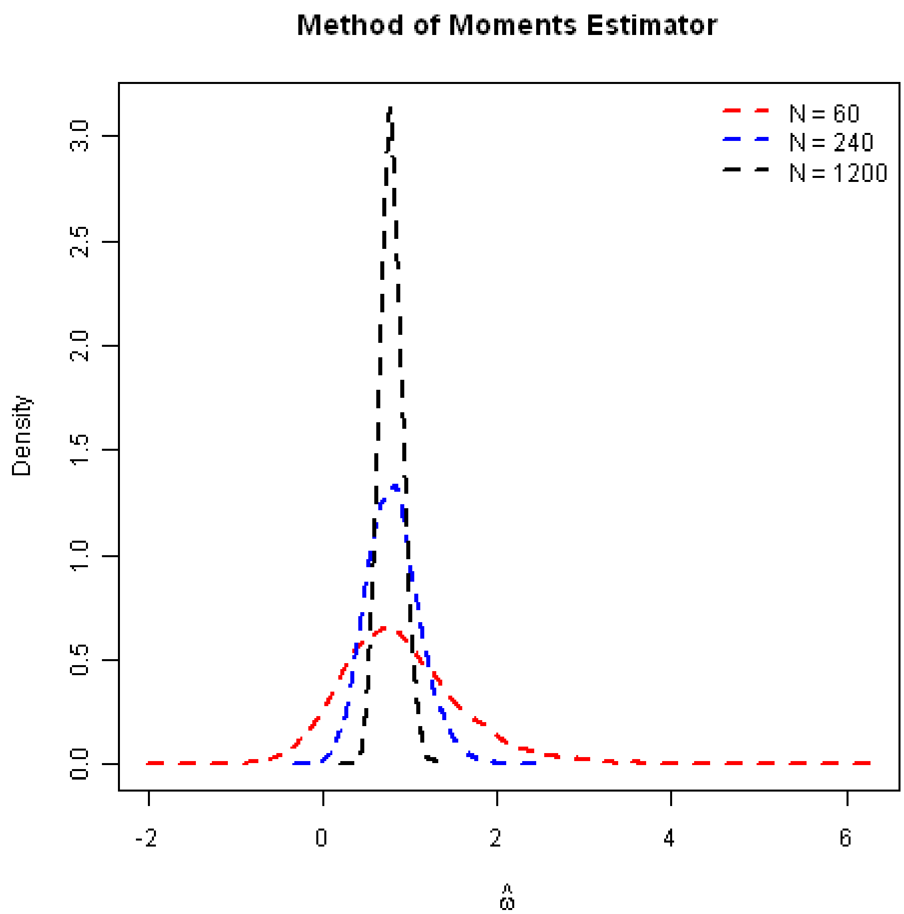

Figure 2: Density plot for the estimator $\hat{\omega}$ at various sample sizes. The density plot is based on 10,000 simulated datasets.

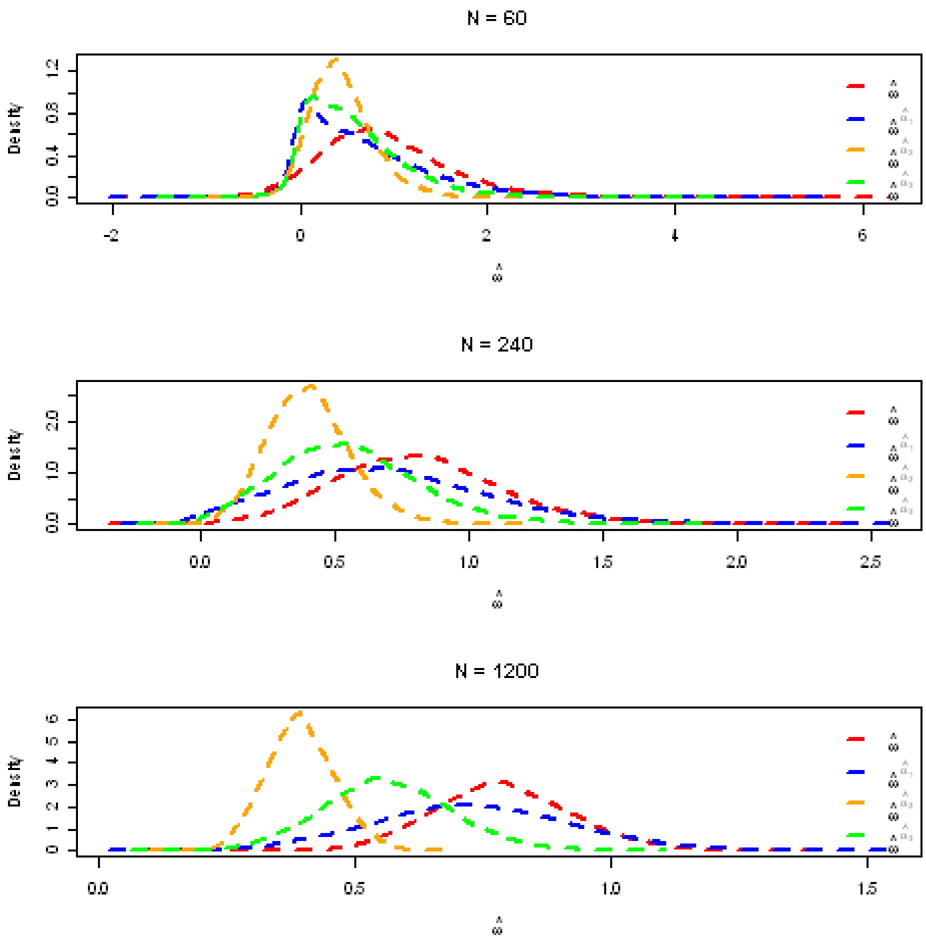

Figure 3: Density plots for the estimators $\hat{\omega}$, $\hat{\omega}^{\hat{\alpha}_1}$ and $\hat{\omega}^{\hat{\alpha}_3}$ at various sample sizes. The density plots are based on 10,000 simulated datasets.

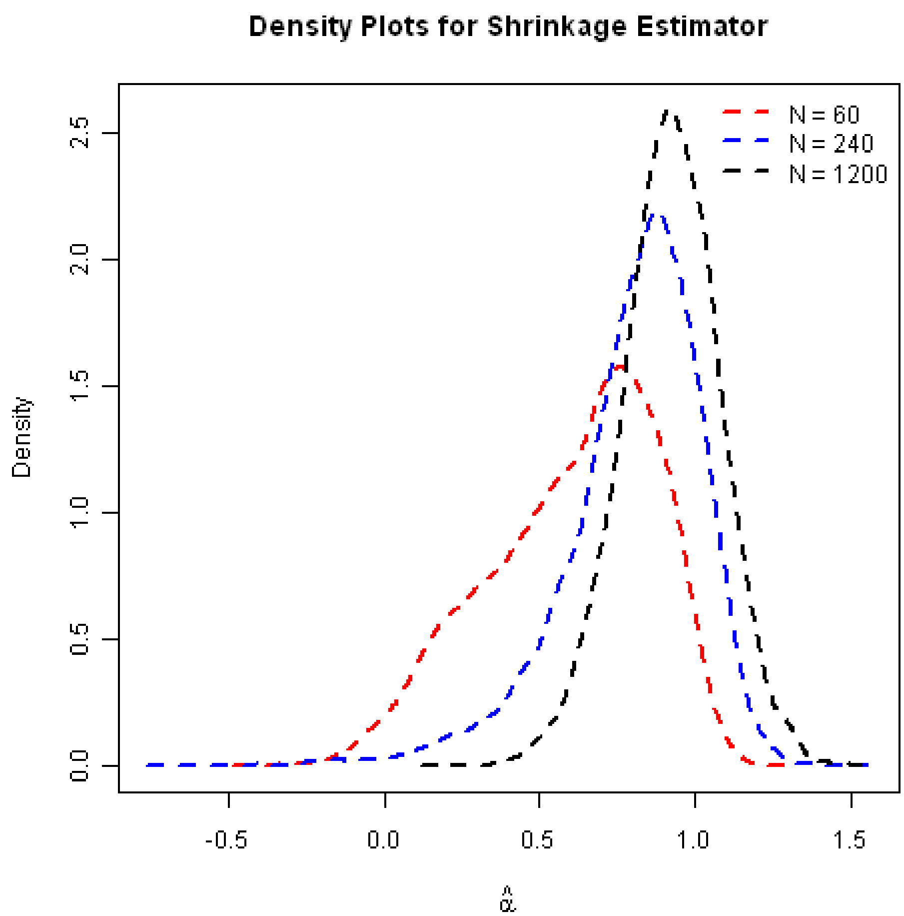

Figure 4: Density plots for the estimated shrinkage factor $\hat{\alpha}_1$ at various sample sizes. This data-drivn shrinkage estimator is discussed in Section 5.2 of the paper. The density plots are based on 10,000 simulations.

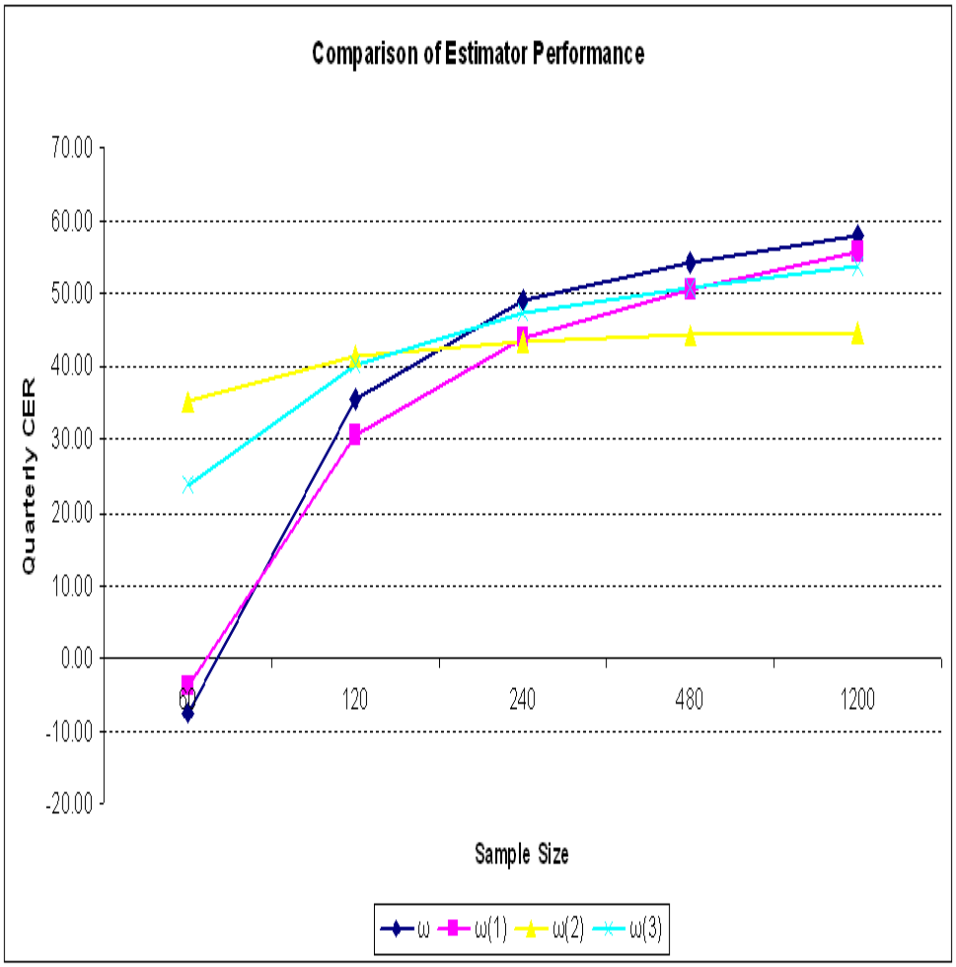

Figure 5: Simulated CER values for the estimators $\hat{\omega}$, $\hat{\omega}^{\hat{\alpha}_1}$ (labeled as $\hat{\omega}(1)$ in the legend), $\hat{\omega}^{\hat{\alpha}_2}$ (labeled as $\hat{\omega}(2)$) and $\hat{\omega}^{\hat{\alpha}_3}$ (labeled as $\hat{\omega}(3)$) at various sample sizes. CER results are based on 400,000 simulated out-of-sample returns, where each estimated portfolio weight from the initial 10,000 simulations is applied to 40 out-of-sample return draws. See Section 5.3 for calibration details regarding the simulation.