

# Subset Optimization for Asset Allocation\*

Benjamin J. Gillen  
Department of Economics  
California Institute of Technology

August 12, 2016

## Abstract

Subset optimization provides a new asset allocation algorithm that's particularly useful when investors face many securities with short return histories. Rather than optimizing weights for all securities jointly, subset optimization constructs Complete Subset Portfolios (CSPs) that naïvely aggregate many "Subset Portfolios," each optimizing weights over a small subset of randomly selected securities. Analytical results characterize the trade-off CSPs balance by implementing a satisficing strategy that sacrifices some potential investor utility so investors can diversify the adverse impact of estimation error on out-of-sample portfolio performance. Simulation and backtest experiments illustrate CSPs' robust performance in several asset universes.

---

\*I'm grateful for helpful commentary and suggestions from Yacine Aït-Sahalia, Khai Chiong, John Cochrane, Jaska Cvitanic, Michael Ewens, Joseph Gerakos, Harry Markowitz, Tatiana Mayskaya, Nicolas Polson, Richard Roll, Alberto Rossi, Allan Timmermann, Ken Whinston, as well as seminar participants at Caltech.

Correspondence: 1200 E California Blvd, MC 228-77; Pasadena, CA 91125 USA  
e-mail: bgillen@caltech.edu; phone: 626-395-4061; fax: 626-795-1547

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Mean-Variance Analysis with Many Securities</b>	<b>4</b>
2.1	Estimation Error and Out of Sample Expected Performance . . . . .	5
2.2	A Brief Review of Quantitative Asset Allocation Strategies . . . . .	6
<b>3</b>	<b>The Algorithm: Complete Subset Portfolios</b>	<b>9</b>
3.1	Population Subset Portfolios . . . . .	9
3.2	Sampling Properties of Subset Portfolio Weights . . . . .	13
3.3	Choosing the Number of Securities and Portfolios . . . . .	15
3.4	Naïve Weighting of Subset Portfolios . . . . .	18
3.5	Constraints and Subset Portfolio Implementation . . . . .	19
<b>4</b>	<b>Simulation Tests of Subset Portfolio Performance</b>	<b>19</b>
4.1	Ground Truth Specifications for Simulated Returns . . . . .	20
4.2	Properties of Simulated Asset Universes . . . . .	22
4.3	Moment Estimators and Portfolio Strategies . . . . .	26
4.4	Simulating Portfolio Strategies' Certainty Equivalent Utility . . . . .	30
<b>5</b>	<b>Backtest Performance of Subset Portfolios</b>	<b>34</b>
<b>6</b>	<b>Robustness and Comparative Statics for Subset Portfolios</b>	<b>37</b>
6.1	Subset Size, Sample Size, and Portfolio Performance . . . . .	38
6.2	Risk Aversion and Subset Portfolio Strategies . . . . .	38
<b>7</b>	<b>Potential Refinements and Extensions</b>	<b>40</b>
7.1	Subset Portfolios with Alternative Objective Functions . . . . .	40
7.2	Weighted and Clustered Selection for Subset Portfolios . . . . .	42
7.3	Enhancing Aggregation for Subset Portfolios . . . . .	43
7.4	Bayesian Refinements for Subset Portfolio Weights . . . . .	43
<b>8</b>	<b>Conclusion</b>	<b>44</b>
	<b>References</b>	<b>44</b>
	<b>Appendix A1: Proofs</b>	<b>49</b>
	<b>Appendix A2: Supplemental Figures and Tables</b>	<b>51</b>
	<b>Separately Presented Tables and Figures as They Appear in the Main Text</b>	<b>59</b>

# 1 Introduction

Despite its prominence as a foundational problem in financial economics, developing reliably implementable solutions to Markowitz (1952)’s classic mean-variance portfolio optimization problem presents a persistently challenging issue. The performance of statistical optimization algorithms suffers from the sensitivity of recommended portfolio weights to estimation error, which becomes particularly severe in the presence of a large number of securities due to the commonly referenced “Curse of Dimensionality.” Indeed, this sensitivity is so severe as to cause some researchers and many professional investment managers to question the relevance of the mean-variance paradigm and the value of incorporating any optimization into the asset allocation decision at all. This paper introduces subset optimization as a simple algorithm for computing portfolio weights that, rather than suffering from the curse of dimensionality, sacrifices some of the investor’s potential utility so as to exploit the large asset universe as a device for diversifying the adverse effects of sampling error on out-of-sample portfolio performance.

The subset optimization strategy presents a new approach to large-scale asset allocation based on satisficing. Rather than jointly optimizing weights for all securities simultaneously, the algorithm randomly selects a small number of securities and forms a “subset portfolio” that optimizes the weights for the selected assets based on sample data. The algorithm proceeds by generating a large number of these subset portfolios, aggregating them into a “Complete Subset Portfolio” by naïvely equally weighting them (an aggregation rule representing the ex-ante optimal combination of exchangeable portfolios). To my knowledge, the subsetting strategy has not yet been considered in application to asset allocation problems or other high-dimensional nonlinear decision models.<sup>1</sup>

---

<sup>1</sup>The closest such algorithm, as suggested by the adopted terminology, is presented in Elliott et al. (2015)’s Complete Subset Regression, a Bayesian Model Averaging strategy that spans regression models with a fixed degree of complexity. The discussion in this paper focuses on the optimization properties of the complete subset portfolio algorithm. Indeed, the complete subset portfolio weights have a natural Bayesian interpretation as the posterior expected weights for an investor whose prior dictates optimal portfolios have exactly  $\hat{N}$  securities but has no information as to which securities should be included in that portfolio. Formulating this representation of the complete subset portfolio algorithm provides a natural mechanism for enhancing the selection of securities and weighting of individual subset portfolios.

The paper’s discussion opens by reviewing the mean-variance problem with many securities, including a cursory survey of some results from the massive literature investigating this problem. Much of this work focuses on improving asset allocation algorithms by developing statistical models that incorporate factors from asset pricing models (Black and Litterman, 1992; Fama and French, 1992; Pastor, 2000) and macroeconomic predictability (Schwert, 1981; Keim and Stambaugh, 1986; Ferson and Harvey, 1999) to improve upon sample estimates of expected returns and to regularize estimated covariance matrices (Ledoit and Wolf, 2004; Carrasco and Noumon, 2012). Especially in large asset universes, the instability of estimated portfolio weights goes beyond “garbage in, garbage out” and is driven largely by the curse of dimensionality (Kritzman, 2006). To address this instability in the mean-variance optimization problem, the literature includes a number of proposed algorithms that are robust to multiple possible data generating processes (Gilboa and Schmeidler, 1989; Goldfarb and Iyengar, 2003; Garlappi et al., 2007; Anderson and Cheng, 2016), penalize or constrain extreme allocations (Jagannathan and Ma, 2003; DeMiguel et al., 2009; Fan et al., 2012), or average across a number of potential data generating processes (Breiman, 1995; Michaud, 1998). Though presented here in its simplest form using sample moments, subset optimization takes a distinctly different approach to the asset allocation problem that can readily incorporate such innovations into the algorithm.

Section 3 formally presents the algorithm, highlighting several key properties that guide the algorithm’s performance. The subset size, or the number of securities selected into each subset, provides an important tuning parameter for the algorithm that balances the efficiency gain from expanded investment opportunities against the cost of sampling error in implemented weights. This section introduces the “complete subset efficient frontier,” a graphical illustration of the expected return realized by an investor given the volatility of complete subset portfolios formed under different subset sizes in the absence of sampling error. In the extreme case where subsets consist of only one asset, the subset portfolio algorithm nests the naïve  $1/N$  rule and the subset efficient frontier consists of a single point. As the subset size increases, the subset efficient frontier expands to match the full efficient

frontier. Between these two extremes, the complete subset efficient frontier characterizes the expected diversification benefits for an investor applying more complex decision rules, much of which are realized by forming subset portfolios of as few as five or ten securities.

When an investor uses data to inform expectations and construct portfolios, the sampling error in observed returns causes estimated portfolio weights to deviate from their expected, optimal, weights. These deviations inject uncertainty into the returns realized by the investor, representing a negative influence on their expected out-of-sample performance as characterized by Kan and Zhou (2007). In a large asset universe where the number of securities may exceed the length of the time-series of returns, the variance of portfolio weights and its negative influence on out-of-sample performance can be unboundedly large. By constraining the number of securities included in the subset optimization problem relative to the observed sample size, subset optimization mitigates this drag on expected performance. As long as the subset size is smaller than the sample size, the variance of individual subset portfolio weights will be bounded, preventing the curse of dimensionality from overcoming the performance of any individual subset portfolio. Complete subset portfolios further reduce the individual subset portfolio weights' variance by averaging these weights across a large number of subsets. Consequently, even in large asset universes, the loss in expected out-of-sample performance due to sampling error for the complete subset portfolio can be bounded (though, as reflected in the complete subset efficient frontier, this stability in portfolio weights comes at the cost of satisficing utility by not fully optimizing over all securities).

Though subset optimization provides new insights into the curse of dimensionality in asset allocation problems, as an algorithmic contribution its performance must be tested by empirical applications. Simulation tests in section 4 provide an ideal sampling environment for characterizing the efficacy of subset optimization against alternative approaches to the asset allocation problem in a variety of asset universe specifications. To characterize how the algorithm would perform in a dynamic implementation, section 5 implements the subset portfolio in a rolling backtest of large asset universes. The algorithm's robustness to tuning parameters is evidenced by expanded simulation tests in section 6. Across all these tests,

the performance of subset portfolios is strikingly robust, and sections 7 and 8 conclude by discussing potential refinements to the algorithm.

## 2 Mean-Variance Analysis with Many Securities

Suppose the investment universe consists of  $N$  securities where asset returns have unknown expected return vector  $\mu$  and unknown covariance matrix  $\Sigma$  that investors endeavor to learn about from observing  $T$  periods of historical returns collected into the information set  $D_T$ . The goal is to find a solution that maximizes the out-of-sample performance of a portfolio under mean-variance utility with risk-aversion parameter  $\gamma$ :

$$w^* \equiv \arg \max_{w \in \Delta^{N-1}} w' \mu - \frac{\gamma}{2} w' \Sigma w \quad (1)$$

A massive literature explores different approaches to solving this problem using investors' subjective expectations for the means and variance-covariance matrix. Under the “plug-in” approach, an investor estimates the population means and covariance matrices in problem (1) from observed data, perhaps adopting a Bayesian prior based on asset pricing theory that adds structure to these estimates and incorporates parameter uncertainty into portfolio risk measures. Approaching the problem this way requires estimating  $N$  means and  $N(N+1)/2$  variances and covariances from a relatively limited sample of observed data. Given posterior beliefs for the expectations ( $\hat{\mu}$ ) and variance-covariance matrix ( $\hat{\Sigma}$ ) of returns, the SEU-optimizing portfolio weights ( $\hat{w}$ ) are:

$$\hat{w} \equiv \arg \max_{w \in \Delta^{N-1}} w' \hat{\mu} - \frac{\gamma}{2} w' \hat{\Sigma} w \quad (2)$$

This problem selects weights to maximize the subjective expected utility based on the investor's beliefs, explicitly conditioning on those beliefs.

## 2.1 Estimation Error and Out of Sample Expected Performance

Consider any statistical portfolio strategy that maps observed data ( $D_T$ ) into individual-asset weights  $\hat{w} : D_T \rightarrow \Delta^{N-1}$ . As statistical estimates, these weights are not deterministic, but depend on the sample drawn from the data generating process that defines  $D_T$ . To characterize the stochastic properties of these weights conditioning on the true data generating process of returns, but not the observed data, denote their expectation  $E[\hat{w}] \equiv \omega$  and variance-covariance matrix  $Var[\hat{w}] \equiv V_{\hat{w}}$ .

This paper's treatment of expected out of sample performance follows the framework of Kan and Zhou (2007), Golosnoy and Okhrin (2009), and Tu and Zhou (2010), who analyze out-of-sample portfolio performance in settings where the ratio  $N/T$  is small. Within this framework, the expected out-of-sample performance of portfolio strategy  $\hat{w}$ , conditional on the true moments of the return generating process can be defined as:

$$\begin{aligned} E\left[\hat{w}'\mu - \frac{\gamma}{2}\hat{w}'\Sigma\hat{w}|\mu, \Sigma\right] &= \omega'\mu - \frac{\gamma}{2}E[\text{tr}(\Sigma\hat{w}\hat{w}')|\mu, \Sigma] \\ &= \omega'\mu - \frac{\gamma}{2}\omega'\Sigma\omega - \frac{\gamma}{2}\text{tr}(\Sigma V_{\hat{w}}) \\ &\equiv U_{\omega} - \frac{\gamma}{2}\text{tr}(\Sigma V_{\hat{w}}) \end{aligned} \tag{3}$$

In the subjective expected utility evaluation guiding equation (2), investors condition on the data and the realized action that they choose based on that data. In the out-of-sample expected performance model for (3), investors evaluate the ex-ante utility of implementing the subjective expected utility weights prior to observing the data. This out-of-sample perspective accounts for the impact of sampling error on an algorithm's performance, reflected in the penalty term  $\frac{\gamma}{2}\text{tr}(\Sigma V_{\hat{w}})$ . In settings where  $N < T - 4$ , Kan and Zhou (2007) analyze plug-in portfolio weights using sample moments, showing the penalty term in equation (3) is  $O(N/T)$  and dwarfs  $U_{\omega}$  as the ratio of  $N$  to  $T$  nears unity. Extending this analysis, Tu and Zhou (2010) construct hybrid portfolio strategies to optimally combine the naïve  $1/N$  rule with the estimated portfolio weights to optimally manage the negative impact of estimation error in the weights on out of sample performance. Okhrin and Schmid (2006) present

analytical derivations for the expectations and covariance matrices of portfolio weights calculated using sample moments, noting that portfolio weight variances are unboundedly large when  $N > T$  and Golosnoy and Okhrin (2009) applies these results to constructing hybrid portfolio strategies.

This paper focuses on settings where  $N$  may be larger than  $T$ , a common practical circumstance to which the results of the previous paragraph do not apply. Here, the ill-conditioned sample covariance matrix causes portfolio weights estimated using sample moments to have an arbitrarily large variance as the optimal weights themselves become unbounded. As this variance causes the penalty in equation (3) to become arbitrarily large, a hybrid strategy that optimally combines any naïve strategy with these estimated weights would place all weight on the naïve rule. In effect, given access to a statistically optimized portfolio, investors would prefer to invest exclusively in naïve strategies and completely ignore the data on returns. The subset algorithm avoids this degeneracy by simplifying the problem and using data to construct optimal portfolios consisting of only a subset  $\hat{N} \subset N$  of the available securities. By naïvely weighting a large number of subset portfolios, the complete subset portfolio weights diversify the impact of estimation error across subsets, effectively driving sampling error in portfolio weights toward zero.

## 2.2 A Brief Review of Quantitative Asset Allocation Strategies

As a textbook problem for financial decision making, researchers have explored a host of strategies and techniques for asset allocation, with many of these surveyed by Brandt (2010). Many of these algorithms adopt a Bayesian formulation of the asset allocation decision problem, exploring how incorporating information from economic theory in the form of prior beliefs for investors can enhance portfolio weights. Other approaches reframe the decision problem to directly regularize portfolio weights.

Applying Bayesian inference techniques to the asset allocation problem first appears with Markowitz (1959)'s discussion of Savage (1954)'s expected utility axioms. Since then, the literature has explored structuring inference through both Stein (1955) shrinkage and more

formal Bayesian analysis. Jorion (1986) and Frost and Savarino (1986) proposed early shrinkage strategies for expected returns, with Ledoit and Wolf (2004, 2013) presenting shrinkage estimators for covariance matrices and Jagannathan and Ma (2003) relating the shadow costs of constraints to a regularized covariance matrix. Dating back to Black and Litterman (1992)'s analysis of the reverse optimization prior based on market portfolio weights, a number of estimators using Bayesian posterior expectations have since entered the literature. Prominent contributions along this line include Pastor (2000), Pastor and Stambaugh (2000), and Pastor and Stambaugh (2002)'s presentation of posterior beliefs when investors' prior beliefs are stated in terms of factor models.

Much of the challenge in implementing asset allocation strategies in markets arises due to dynamic features in the return generating process and addressing the influence of changing economic conditions on asset returns has inspired a large literature exploring predictability in expected returns, dynamic volatility, or time-varying factor exposures. Some of the earlier work in this literature, including Fama (1981), Schwert (1981), and Keim and Stambaugh (1986), focuses on the predictability of asset-class returns using predetermined macroeconomic variables. Rapach and Zhou (2013) surveys the literature discussing forecasting stock returns, with notable contributions from Ferson and Harvey (1993), Ferson and Harvey (1999), and Avramov and Chordia (2006). A closely related literature analyzes predictability as a mechanism for momentum in mutual fund manager performance, with Avramov and Wermers (2006) analyzing US mutual funds and Banegas et al. (2013) investigating European mutual funds. Anderson and Cheng (2016) present a robust approach to modeling dynamics in asset returns that allows for multiple priors and rich model dynamics. This approach relates to the literature on robust optimization from Goldfarb and Iyengar (2003) and explores both model averaging and min-max approaches to learning in multi-prior environments from Gilboa and Schmeidler (1989) and Garlappi et al. (2007). Subset portfolios can readily incorporate dynamic estimates for return as well as robust optimization principles when computing subset weights, as discussed very briefly in section 7.

Another branch of the literature directly estimates portfolio weights themselves. Brandt

et al. (2009) derive optimal portfolio policies in terms of the cross-sectional characteristics of individual securities, constructing a mapping from security characteristics to portfolio holdings, a practice that goes back to Sharpe (1964)'s demonstration of the optimality of the market-weighted portfolio within the Capital Asset Pricing Model (CAPM). The Black and Litterman (1992) approach utilizes the CAPM to relate these market-capitalization weights to a firm's expected return within a Bayesian framework. Avramov and Zhou (2010) survey Bayesian strategies for incorporating investor beliefs about weights into the asset allocation decision, a natural framework in which one could interpret the complete subset portfolio weights. Many of these techniques directly regularize portfolio weights by way of penalized optimization, as in DeMiguel et al. (2009), Carrasco and Noumon (2012), and Fan et al. (2012).

Other work considers hybrid optimization strategies for portfolio formation. Within the out-of-sample performance framework, Tu and Zhou (2010) derive the optimal combination for an investor constructing three funds representing the sample estimated portfolio weights that maximize the Sharpe Ratio, the weights that minimize the variance, and the naïve  $1/N$  rule. Golosnoy and Okhrin (2009) also presents a shrinkage strategy for estimated portfolio weights toward the  $1/N$  rule. In principle, each of these hybrid strategies could be embedded within the subset portfolio algorithm.

The subset portfolio algorithm shares some similarities with bootstrap aggregation algorithms proposed by Michaud (1998) for asset allocation and Breiman (1995) for statistical decision rules generally. These resampling algorithms bootstrap observed returns to generate a large number of samples from the data generating process. For each bootstrap draw, these algorithms calculate optimized weights, aggregating across the bootstrap by averaging the optimized weights. It is well known that, absent non-negativity constraints, resampling portfolio weights simply introduces noise to the sample estimated weights. However, these algorithms are remarkably effective when implemented with non-negativity constraints. The link between these algorithms, however, subtly derives from the role non-negativity constraints play in resampling algorithms. When weights are constrained to be non-negative,

only a small number of securities actually receive positive portfolio allocations. This effectively randomized selection forms subset portfolios out of those securities that are assigned positive allocations, with bootstrap aggregation then averaging across the weights in these subset portfolios.

### 3 The Algorithm: Complete Subset Portfolios

Algorithm (1) presents the complete subset portfolio strategy, which is remarkably straightforward and easy to implement for any number of securities  $N$ . Rather than computing optimized weights assigned to all securities simultaneously, subset portfolios optimize weights over a relatively small, randomly-selected, subset of securities. This section opens by characterizing how this restriction on the optimization problem reduces the potential benefits from diversification by evaluating the algorithm’s performance when investors know the exact distribution of returns. Considering the algorithm’s performance in settings where investors do not know this distribution, subsection 3.2, illustrates how sampling error in subset portfolio weights affects the out-of-sample performance for complete subset portfolios. The section closes with a discussion of practical issues for implementing the subset algorithm: tuning parameters, the aggregation rule for subset portfolios, and accommodating constraints in the optimization problem.

#### 3.1 Population Subset Portfolios

Intuitively, subset portfolios can be interpreted as a satisficing algorithm in which investors sacrifice some benefits of diversification to mitigate the risk of estimation error. Subset weights clearly fail to achieve the global optimal utility for the investor, so how much utility is the investor sacrificing and how does this loss vary with the number of securities in each subset? Analyzing the properties of subset portfolios in the absence of estimation provides some insight into this question.

Define a “population subset portfolio” as the subset portfolio for investors who know

---

**Algorithm 1** Calculating Complete Subset Portfolio Weights

---

**0. Fix Subset Size.**

For any number of securities  $N$  and length of the return time-series for each security  $T$ , fix the subset size  $\hat{N} \ll \min(N, T)$ .

**1. Construct Individual Subset Portfolios.**

For  $b = 1, 2, \dots, B$

- I. *Select subset of securities.* Uniformly randomly select  $\hat{N}$  securities without replacement. Identify the set of securities by the index  $\mathcal{N}_b$ .
- II. *Compute optimal subset portfolio weights.* Solve:

$$\hat{w}_b = \arg \max_{w \in \Delta^{\hat{N}-1}} w' \hat{\mu}_b - \frac{\gamma}{2} w' \hat{\Sigma}_b w$$

where  $\hat{\mu}_b$  and  $\hat{\Sigma}_b$  represent the sample means and variance-covariance matrix for the securities in  $\mathcal{N}_b$ .

Next  $b$

**3. Equally-Weight Individual Subset Portfolios for Complete Subset Portfolio.**

For security  $i$  assign the weight  $\hat{w}_i^* \equiv \frac{1}{B} \sum_{b=1}^B \hat{w}_{i,b}$ , where  $\hat{w}_{i,b}$  is the weight assigned to asset  $i$  in subset portfolio  $\hat{w}_b$  and equal to zero for subsets that do not include asset  $i$ .

---

the population distribution of returns. To construct a population subset portfolio, select a subset of  $\hat{N} \ll \min(N, T)$  securities and let  $\mu_b$  and  $\Sigma_b$ , respectively, denote the (true) population means and covariance matrix for the selected securities' returns. Let  $U_b^*$  denote the maximized utility in the subset optimization problem:

$$w_b^* \equiv \arg \max_{w \in \Delta^{\hat{N}-1}} w' \mu_b - \frac{\gamma}{2} w' \Sigma_b w \tag{4}$$

Borrowing terminology from Elliott et al. (2015), define the “complete” population subset portfolio of size  $\hat{N}$  by equally weighting all  $N$  choose  $\hat{N}$  possible subset portfolios.<sup>2</sup> Denote

---

<sup>2</sup>From a computational perspective, spanning the population of all  $\frac{N!}{\hat{N}!(N-\hat{N})!}$  such combinations of securities is clearly an infeasible prospect. Subsection 3.3 will address this issue, invoking the law of large numbers to argue the sufficiency of generating a reasonably large number of subsets for computing complete subset portfolio weights.

the complete population subset portfolio weights  $\bar{w}_{\hat{N}}^*$  and utility  $U_{\hat{N}}^* \equiv \bar{w}_{\hat{N}}^{*\prime} \mu - \bar{w}_{\hat{N}}^{*\prime} \Sigma \bar{w}_{\hat{N}}^*$ .

Note that subset portfolios nest the  $1/N$  rule when their subset size is set at the  $\hat{N} = 1$  extreme, at which point investors' subset portfolios each consist of a single security. At the other extreme where  $\hat{N} = N$ , subset portfolio weights are equivalent to sample estimated portfolio weights. In this way, subsetting portfolio weights reflect the trade-off between the simplicity of the naïve  $1/N$  rule and the complexity of fully optimized weights across all securities.

To characterize the effect of increasing subset portfolio size in the absence of estimation error, consider the properties of complete population subset portfolios when  $\gamma$  is arbitrarily high and the investor seeks the Global Minimum Variance (GMV) portfolio. As can be formally established by induction on subset size, the complete population subset GMV portfolio's variance declines as the subset size increases. Assuming no two securities have the exact same expected return, this monotonicity extends to variance minimization subject to a minimal expected return, a problem equivalent to the objective presented in Algorithm 1. The next theorem states formally that, for a targeted portfolio expected return  $\underline{\mu}$ , the variance of the complete subset portfolio that minimizes portfolio variance with expectation at least  $\underline{\mu}$  declines monotonically in subset portfolio size.

**Theorem 1.** *Suppose  $N$  assets' returns have mean  $\mu$  and variance  $\Sigma$ . Let  $\bar{w}_{\hat{N}_i}^*$  represent the complete population subset portfolio weights where the subset optimization step (II) minimizes the variance for a subset portfolio of size  $\hat{N}_i$  subject to a minimal expected return constraint,  $\underline{\mu}$ :*

$$\bar{w}_{\hat{N}_i}^* (\underline{\mu}) = \lim_{B \rightarrow \infty} \frac{1}{B} \sum_{b=1}^B w_b^* (\hat{N}_i, \underline{\mu}) \quad w_b^* (\hat{N}_i, \underline{\mu}) = \arg \min_{w \in \Delta^{\hat{N}_i-1}} w' \Sigma_b w \text{ subject to } w' \mu_b \geq \underline{\mu}$$

*Then, if  $\hat{N}_1 > \hat{N}_2$  the variance of the complete population subset portfolio of size  $\hat{N}_1$  is lower than the variance of the complete subset portfolio of size  $\hat{N}_2$ :*

$$\bar{w}_{\hat{N}_1}^{*\prime} \Sigma \bar{w}_{\hat{N}_1}^* < \bar{w}_{\hat{N}_2}^{*\prime} \Sigma \bar{w}_{\hat{N}_2}^*$$

This theorem motivates a definition for the “complete population subset efficient frontier,” a graphical presentation plotting the complete subset portfolio volatility for a given expected return target.<sup>3</sup> In plots consolidating the complete subset efficient frontier for different subset sizes, the  $1/N$  rule appears as a point representing the complete population subset portfolio of size one. As the subset size grows, the subset efficient frontier expands and converges to the efficient frontier generated by including all securities in the portfolio. To illustrate this property, Figure 1 plots the population subset efficient frontier for the simulated asset universe generated by US Stocks with additional frontiers for other universes presented in section 4.1.

**Definition 1** (The Complete Population Subset Efficient Frontier of size  $\hat{N}$ ). *The Complete Population **Subset Efficient Frontier of Size  $\hat{N}$**  plots the tradeoff between expected return and volatility an investor obtains by implementing the complete subset portfolio while using the population means, variances, and covariances for securities.*

Clearly, in the absence of estimation error, investors want the largest choice set possible to maximize the benefits of diversification. However, the marginal benefit from adding each new security to a portfolio are typically declining with the number of assets already included in the portfolio. Many researchers have considered the number of assets necessary to obtain a “diversified portfolio,” but this question lacks a definitive answer because it critically depends on the structure of returns in the market.<sup>4</sup> Note, however, that while individual subset portfolios may consist of a relatively small number of securities, aggregated subset portfolios’ weights span the entire market.

---

<sup>3</sup>Without short-sales constraints, the properties of mean-variance analysis extend to each subset portfolio to map this frontier. For each subset, the algorithm need only compute the minimum variance and maximum Sharpe ratio portfolios. To achieve a given expected return target, the efficient subset portfolio is a simple weighted average of these two portfolios, the variance of which is easily calculated.

<sup>4</sup>Evans and Archer (1968) present early simulation results that as few as ten securities suffices for diversification, though analytical results from Elton and Gruber (1977) suggest these simulations may understate the benefits of diversification. Statman (1987) cites Evans and Archer (1968) and four textbooks suggesting as few as ten securities are enough to form a diversified portfolio before presenting analytical results indicating substantial benefits are available from forming portfolios including 30 to 40 stocks.

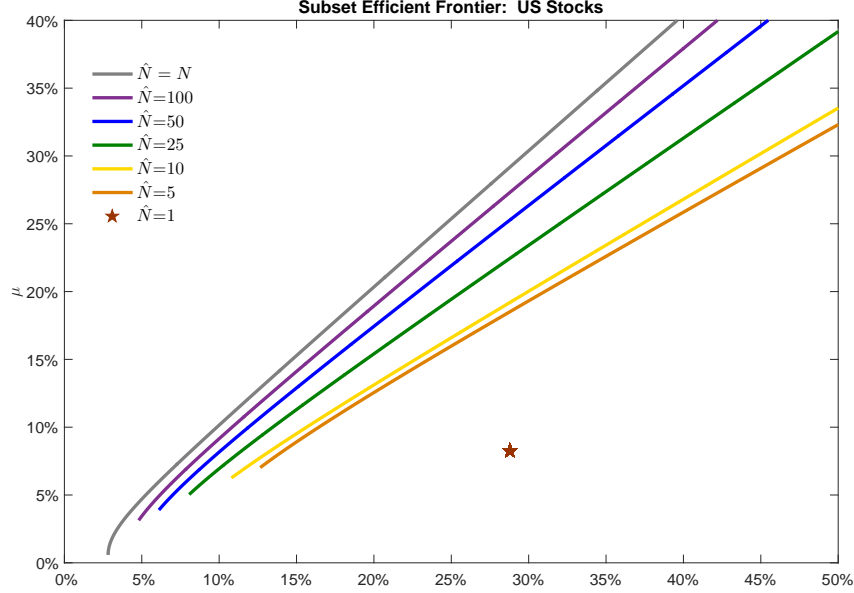


Figure 1: Population Complete Subset Efficient Frontier for Simulated US Stock Data  
This figure presents the mean-variance tradeoffs for investors implementing complete subset portfolios with access to population means and variances calibrated to a simulation universe representing  $N = 1,063$  US Stocks. For each subset, investors minimize portfolio variance allowing for short sales subject to the expected return target and complete subset portfolios aggregate individual subsets with equal weights across subsets. The  $1/N$  rule appears as a point representing the complete population subset portfolio of size one. The line for each subset size plots the mean (y-axis) and volatility (x-axis) of the complete subset portfolio that minimizes the variance for that level of expected return. As the subset size increases, the complete subset efficient frontier expands and converges to the efficient frontier generated by including all securities in the portfolio.

### 3.2 Sampling Properties of Subset Portfolio Weights

To consider the setting where population means and variances are unavailable to the investor, let  $\hat{\mu}_b$  and  $\hat{\Sigma}_b$  denote unbiased sample estimates for the means and covariances of assets selected into subset  $\mathcal{N}_b$ . Letting  $\mathbf{1}$  denote a vector of ones, the unbiased estimator for the population subset portfolio weight  $w_b^*$  is:

$$\hat{w}_b = \frac{\hat{\Sigma}_b^{-1} \mathbf{1}}{\mathbf{1}' \hat{\Sigma}_b^{-1} \mathbf{1}} + \frac{T - \hat{N} - 1}{\gamma (\hat{N} - 1)} \hat{R}_b \hat{\mu}_b, \text{ with } \hat{R}_b = \hat{\Sigma}_b^{-1} - \frac{\hat{\Sigma}_b^{-1} \mathbf{1} \mathbf{1}' \hat{\Sigma}_b^{-1}}{\mathbf{1}' \hat{\Sigma}_b^{-1} \mathbf{1}} \quad (5)$$

Since  $\hat{w}_b$  is unbiased for  $w_b^*$ , the analog to equation (3) for the expected out of sample

performance is the true optimal utility ( $U_b^*$ ) minus the sampling error penalty for the weights:

$$\begin{aligned} E \left[ \hat{w}'_b \mu - \frac{\gamma}{2} \hat{w}'_b \Sigma \hat{w}_b | \mu, \Sigma \right] &= w_b^{*\prime} \mu - \frac{\gamma}{2} w_b^{*\prime} \Sigma w_b^* - \frac{\gamma}{2} \text{tr}(\Sigma V_{\hat{w}_b}) \\ &= U_b^* - \frac{\gamma}{2} \text{tr}(\Sigma V_{\hat{w}_b}) \end{aligned} \quad (6)$$

Here  $V_{\hat{w}_b}$  again denotes the variance-covariance matrix for the estimated subset portfolio weights. With  $\hat{N}$  much smaller than  $T$ , the entries in  $V_{\hat{w}_b}$  are bounded by a constant times  $T^{-1}$ , consistent with Kan and Zhou (2007)'s result that the out-of-sample certainty equivalent penalty for sampling error in each subset portfolio is  $O(\hat{N}/T)$ .

Let  $\hat{w}_{\hat{N}}$  denote the complete subset portfolio weights, constructed by averaging across all  $N$  choose  $\hat{N}$  subset portfolios:

$$\hat{w}_{\hat{N}} = \lim_{B \rightarrow \infty} \frac{1}{B} \sum_{b=1}^B \hat{w}_b \quad (7)$$

Since  $\hat{w}_b$  is unbiased for  $w_b^*$  for each  $b$ , the complete subset portfolio weights  $\hat{w}_{\hat{N}}$  are unbiased for the complete population subset portfolio weights  $w_{\hat{N}}^*$ . Consequently, the expected out-of-sample performance of complete subset portfolios is also equal to the complete population subset utility minus a penalty based on the variability of the aggregated weights.

$$E \left[ \hat{w}'_{\hat{N}} \mu - \frac{\gamma}{2} \hat{w}'_{\hat{N}} \Sigma \hat{w}_{\hat{N}} | \mu, \Sigma \right] = U_{\hat{N}}^* - \frac{\gamma}{2} \text{tr}(\Sigma V_{\hat{w}_{\hat{N}}}) \quad (8)$$

To extend Kan and Zhou (2007)'s result to complete subset portfolios, consider how averaging the weights from many subset portfolios reduces the total sampling error in the complete subset portfolio. Since each security is selected into only approximately  $\hat{N}/N\%$  of the subset portfolios, averaging across subset portfolios effectively scales each subset portfolio's weight by  $\hat{N}/N$ . Consequently, the entries in  $V_{\hat{w}_{\hat{N}}}$  are bounded by a constant times  $\frac{\hat{N}^2}{N^2 T}$ , going to zero as either the universe size ( $N$ ) or the sample size ( $T$ ) becomes large. Since there are  $N^2$  terms in the  $\text{tr}(\Sigma V_{\hat{w}_{\hat{N}}})$  penalty, the out-of-sample certainty equivalent penalty for sampling error in subset portfolios across all securities remains  $O(\hat{N}^2/T)$ . The following theorem, which is proved in the appendix, formalizes this result:

**Theorem 2.** *Suppose an investor who observes  $T$  observations of returns on  $N$  securities with mean  $\mu$  and covariance matrix  $\Sigma$  uses equation (5) to calculate the weights of each subset portfolio of fixed size  $\hat{N}$  with  $T - 4 \geq \hat{N} \geq 4$  and denote the complete subset portfolio weights  $\hat{w}_{\hat{N}}$ .*

1. *The estimated complete subset portfolio weights are unbiased for the population complete subset portfolio weights:*

$$E [\hat{w}_{\hat{N}}] = \bar{w}_{\hat{N}}^*$$

2. *The sample variance of a single subset portfolio weight is bounded by a constant times  $T^{-1}$ , so that  $\text{Var}(\hat{w}_b) = O(T^{-1})$ .*
3. *The expected out of sample performance for the estimated complete subset portfolio equals that of the population subset portfolio minus a penalty term that is bounded by a constant times  $\hat{N}^2/T$ :*

$$E \left[ \hat{w}'_{\hat{N}} \mu - \frac{\gamma}{2} \hat{w}'_{\hat{N}} \Sigma \hat{w}_{\hat{N}} \right] = U_{\hat{N}}^* - O(\hat{N}^2/T)$$

### 3.3 Choosing the Number of Securities and Portfolios

The subset algorithm includes two tuning parameters that must be selected prior to implementation: The subset size ( $\hat{N}$ ) and the number of subset portfolios ( $B$ ). This subsection illustrates how these two parameters might influence portfolio performance.

Increasing the size of subset portfolios expands the subset efficient frontier while also increasing the penalty in expected out-of-sample performance due to sampling error, so the subset portfolio size should balance these offsetting forces. General results for how best to strike this balance aren't available as the optimal subset size depends on properties of the return generating process itself. In specially constructed settings, such as the constant correlation model with expected returns proportional to security variances, it is straightforward to show that the variance of returns on complete population subset portfolios of size  $\hat{N}$  is

$O(\hat{N}^{-1})$ . From Theorem 2, the magnitude of the penalty term due to sampling error in portfolio weights is  $O(\hat{N}^2/T)$ . The scaling of these terms has the same order when the subset size is  $O(T^{1/3})$ . However, the exact subset size will depend on constants determined by the true distribution of returns, so choosing the implemented subset size presents primarily a practical issue. Though this paper only considers implementations for subset portfolios of a fixed size, cross-validation or simulation techniques could provide guidance for selecting the subset size in different settings.

To develop some intuition for how subsetting improves portfolio performance, consider first the setting when the subset size is small. In this setting, there's very little noise in estimated weights and so performance is quite reliable across data samples. As the subset size increases, the increased variability in weights translates into increased variability in portfolio mean and variance. To illustrate this effect, consider simulating 100 sample histories with 100 months of returns on US Stock Data and computing subset optimal portfolio weights for a mean-variance investor with  $\gamma = 4$  using subsets of size  $\hat{N} \in \{5, 50, 100\}$ . For these simulated portfolio weights, Figure 2 plots the true mean and variance for the portfolio's out-of-sample performance. If the investor were to hold the  $1/N$  portfolio, their expected return would be 8.05% with volatility of 16.63%, with the associated indifference curve plotting mean-variance combinations that make the investor equally well-off to holding the  $1/N$  portfolio. When  $\hat{N} = 5$ , the simulated portfolios' mean-variance combinations are reasonably tightly clustered above and to the left of this indifference curve. As  $\hat{N}$  grows, this cloud spreads out and the estimated portfolio weights often deliver means and variances that make an investor much worse-off than the  $1/N$  rule.

The other parameter to determine for implementation is the actual number of subset portfolios to generate for the algorithm,  $B$ . The complete population subset portfolios take the limit as  $B \rightarrow \infty$ , but in practice, computing the full set of  $\frac{N!}{\hat{N}!(N-\hat{N})!}$  possible subsets in the complete subset portfolio presents an impossible calculation. However, completely spanning this set isn't necessary once a sufficiently large sample of subsets is drawn as the law of large numbers drives convergence in complete subset portfolio weights. For this

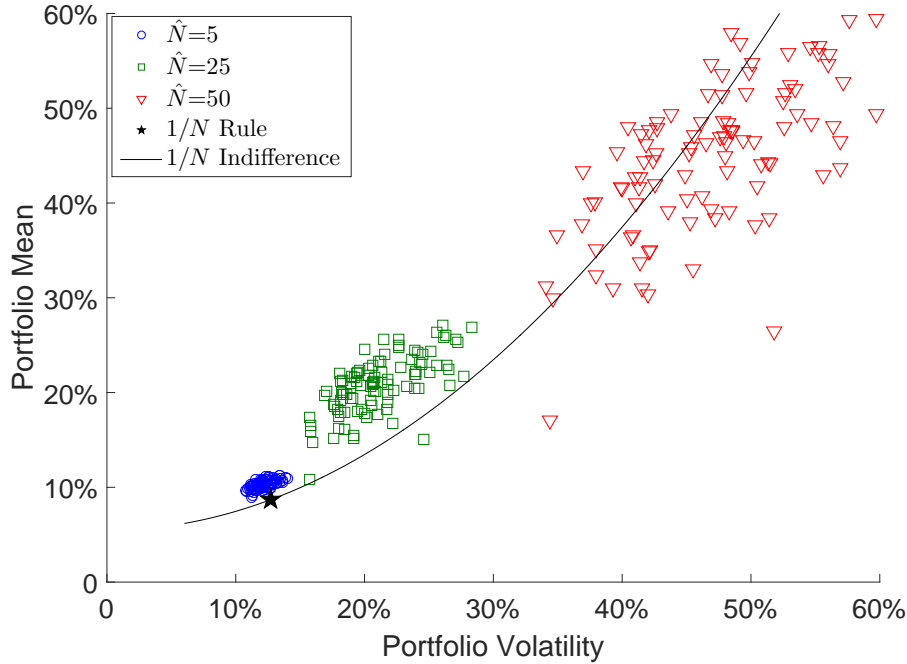


Figure 2: Subset Size, Sampling Error in Weights, and Portfolio Performance

This figure presents the mean-variance tradeoffs for investors choosing the subset size when implementing complete subset portfolios securities with population means and variances calibrated to a universe of  $N = 1,063$  US Stocks. The star represents the mean and volatility for the  $1/N$  rule with the black indifference curve reflecting the combinations of means and variances that lead to indifference for an investor with mean-variance preferences and risk aversion parameter  $\gamma = 4$ . From 100 simulation samples with 100 months' returns each, the blue circles represent the true means and variances from implementing the estimated complete subset portfolio weights with  $\hat{N} = 5$  and the green squares and red triangles represent the same with  $\hat{N} = 25$  and  $\hat{N} = 50$ , respectively. Though the performance of the strategy with  $\hat{N} = 5$  always lies above and to the left of the  $1/N$  rule's indifference curve, as the subset size becomes large, estimated portfolio weights, and consequent performance, become noisier and leads to potentially bad outcomes for investors.

reason, the implemented value for  $B$  can be effectively determined during implementation using rules of thumb based on the variance in the weights computed across subset draws. One approach could apply Gelman and Rubin (1992) or Geweke (1991)-style tests for convergence of Markov Chain Monte-Carlo samplers often used in Bayesian estimation. Alternatively, the number of subsets could be determined so that each security appears in a fixed number of subset portfolios. For instance, if each security is selected into 10,000 subset portfolios, then the standard error of its weight due to subset selection will be reduced by a factor of 1%. Such a specification would suggest generating  $10,000 * N / \hat{N}$  subsets, a cumbersome but easily

parallelized computational exercise.

### 3.4 Naïve Weighting of Subset Portfolios

Given a universe of subset portfolios, an investor interested in maximizing their subjective expected utility would then endeavor to learn these portfolios' means and covariances to optimally combine the subset portfolios themselves. Rather than applying the naïve equal weighting proposed in Algorithm (1), this investor is simply trading the problem of assigning weights to a universe of  $N$  individual securities for the problem of assigning weights to a universe of  $B$  subset portfolios.

From an ex-ante perspective, selecting weights for the  $B$  subset portfolios presents an easier problem than the  $N$  individual securities. Because subset securities are selected uniformly randomly, the return series for each of the  $B$  subset portfolios are ex ante exchangeable. This exchangeability allows the returns on two subset portfolios to be weighted as if they are independent and identically distributed conditional on the true values of  $\mu$  and  $\Sigma$ . Consequently, exchangeability implies the naïve equally weight applied to subset portfolios is ex ante optimal.<sup>5</sup>

In this way, constructing subset portfolios rotates the  $N$  individual securities into  $B$  locally-optimized portfolios with ex-ante exchangeable returns. As will be discussed in section 7, one could imagine numerous techniques for refining the aggregation rule for subset portfolios. The current work focuses on the equal weighting rule as the reference model, not only because this rule is well-motivated by its ex-ante optimality, but also because it provides the simplest presentation of the complete subset strategy.

---

<sup>5</sup>This result follows from de Finetti's Representation theorem applied to a diffuse prior with exchangeability, as presented in Jacquier and Polson (2012). Though this equal-weighting is optimal ex-ante, Jacquier and Polson (2012) also characterize how investors update their beliefs based on observed returns and use that information to optimize their exposure to securities in the market.

### 3.5 Constraints and Subset Portfolio Implementation

The discussion to this point assumes weights in individual securities are not restricted by non-negativity or gross exposure constraints. This simplifying assumption facilitates establishing analytical properties of subset portfolio optimization, but fails to account for realistic constraints on investment policies. Such constraints complicate analyzing subset portfolio properties as there is no unbiased estimator for constrained subset optimal portfolio weights. However, by constraining portfolio weights themselves, these constraints also limit the sampling error in those weights, tempering the penalty associated with implementing potentially larger-sized subsets.

From an implementation perspective, incorporating constraints into the optimization problem is mainly a computational issue that can be addressed in a variety of ways. Marginal constraints on individual securities could be implemented by simply censoring subset portfolio weights and rescaling portfolio weights to sum to unity. While nonlinear constraints on aggregated weights can be challenging to implement, linear constraints that can be accommodated with each subset portfolio will be trivially satisfied by the aggregated subset portfolio. Beyond externally-dictated regulatory or policy constraints, statistical constraints controlling portfolio exposure to estimated risk factors are also readily implemented within the subset portfolio algorithm. Indeed, to the extent that statistically-hedged portfolios are left with residual exposure due to estimation error in factor loadings, a complete subset portfolio can provide a more reliable hedge by diversifying this exposure across subset portfolios.

## 4 Simulation Tests of Subset Portfolio Performance

As the most relevant test of the subset portfolio algorithm lies in its implementation, this section compares the algorithm's performance with that of a number of alternative strategies in a series of simulation experiments. These experiments closely match the information environment in which the out-of-sample performance measure was derived and follow similar

implementations by Chopra and Ziemba (1993), Markowitz and Usmen (2003), Liechty et al. (2008), and Fan et al. (2012), among others. Calibrating the population return generating process for simulated data to historical excess returns from various asset universes allows these tests to exactly evaluate the expected out-of-sample performance of different portfolio strategies.

## 4.1 Ground Truth Specifications for Simulated Returns

The simulation testing methodology begins by calibrating a data generating process to three asset universes consisting of US Stocks, European Mutual Funds, and European Stocks. The simulated returns for asset  $i$  in period  $t$  are based on a factor model:

$$r_{i,t} = \beta_i' r_{F,t} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma_i^2) \quad (9)$$

where  $r_{F,t} \sim N(\mu_F, \Gamma_F)$  is a vector of normally-distributed returns on systematic factors with mean  $\mu_F$  and covariance matrix  $\Gamma_F$ ,  $\beta_i$  is a vector of factor loadings for asset  $i$ , and  $\epsilon_{i,t}$  represents an i.i.d. idiosyncratic return for asset  $i$  that's independent of  $r_{F,t}$  and the idiosyncratic returns for all other assets  $\epsilon_{j,\tau}, \forall j \neq i$ . The simulated data generating process calibrates these factor to historical data from CRSP (US Stock Data), DataStream (European Stock Data, reported in Banegas et al. (2013)), and Lipper (European Mutual Fund Data, also reported in Banegas et al. (2013)).

The first simulation model is “Restricted,” corresponding to a properly-specified four-factor model with no mispricing ( $\alpha = 0$ ) where the priced factors include the usual Market (Mkt) factor, a Small-Minus-Big (SMB) size factor, a High-Minus-Low Book-to-Market Ratio (HML) value factor, and a Winning-Minus-Losing (WML) momentum factor, all of which come from Ken French’s Data Library. These factors’ expectations,  $\mu_F$ , and covariance matrix,  $\Gamma_F$ , are calibrated to historical sample moments denoted  $\hat{\mu}_F$  and  $\hat{\Gamma}_F$ , respectively. For asset  $i$ , let the index set  $D_i$  represent the sample of  $T_i$  historical periods in which the return on asset  $i$  is observed, with  $r_i$  denoting the vector of these returns. Let  $X_{i,s} \equiv$

$[r_{Mkt,s}, r_{SMB,s}, r_{HML,s}, r_{WML,s}]$  denote the factor returns for period  $s \in D_i$ , and construct  $X_i \equiv [X_{i,1}, \dots, X_{i,T_i}]$  so that  $X_i$  and  $r_i$  are conformable. The parameters for the simulation DGP,  $\hat{\beta}_i \equiv [\hat{\beta}_{i,Mkt}, \hat{\beta}_{i,SMB}, \hat{\beta}_{i,HML}, \hat{\beta}_{i,WML}]'$ , and  $\hat{\sigma}_i^2$  are then given by the usual OLS formula:

$$\hat{\beta}_i = (X_i' X_i)^{-1} X_i' r_i, \text{ and, } \hat{\sigma}_i^2 = \frac{1}{T_i - 4} \sum_{t \in \tau_i} (r_{i,t} - \hat{\beta}_i' r_{F,t})^2. \quad (10)$$

Then the return for asset  $i$  in simulation period  $\tau$  for the ‘‘Restricted’’ model is:

$$r_{i,\tau} = \hat{\beta}_i' r_{F,\tau} + \epsilon_{i,\tau}, \text{ where } r_{F,\tau} \sim N(\hat{\mu}_F, \hat{\Gamma}_F) \text{ and } \epsilon_{i,t} \sim N(0, \hat{\sigma}_i^2) \quad (11)$$

and the investor observes  $r_{i,\tau}$  and  $r_{F,\tau}$  for all  $i = 1, \dots, N$  and for all simulation periods  $\tau$ .

The second simulation model is ‘‘Augmented’’ with ten principal components extracted from individual security returns to reflect misspecification in the factor model itself. Let  $\tilde{r}_{F,t}$  denote the time  $t$  vector of four systematic factors and ten principal component factors and denote the historical sample means for the factors by  $\hat{\mu}_F$  and the sample covariance matrix by  $\hat{\Gamma}_F$ . Using analogous definitions for  $\tilde{X}_{i,s}$  and  $\tilde{X}_i$ , let  $\hat{\beta}_i$  represent the augmented factor loadings for asset  $i$  estimated by OLS and  $\hat{\sigma}_i^2$  denote the corresponding estimated residual variance. Then the return for asset  $i$  in simulation period  $\tau$  for the ‘‘Augmented’’ model is:

$$r_{i,\tau} = \hat{\beta}_i' \tilde{r}_{F,\tau} + \tilde{\epsilon}_{i,\tau}, \text{ where } \tilde{r}_{F,\tau} \sim N(\hat{\mu}_F, \hat{\Gamma}_F) \text{ and } \epsilon_{i,t} \sim N(0, \hat{\sigma}_i^2) \quad (12)$$

and the investor observes  $r_{i,\tau}$  and  $r_{F,\tau}$  for all  $i = 1, \dots, N$  and for all simulation periods  $\tau$ , but not the full vector of factors  $\tilde{r}_{F,\tau}$ . By allowing for unobserved factors to drive returns for each security, the augmented model not only allows individual securities to have non-zero  $\alpha$  but also allows for extra-benchmark correlation that would not be captured in a benchmark model.

## 4.2 Properties of Simulated Asset Universes

This subsection presents some results characterizing the properties of the three asset universes used in the simulation tests. The first universe draws on 31,219 US Stocks with returns from 1963-2015 reported in the CRSP database. After screening for only those stocks that have returns in 75% of the sample, this asset universe includes a total of 1,063 stocks. The second universe selects 878 out of 4,955 European Equity Mutual Funds with returns from 1988-2008 reported in the Lipper database as analyzed in Banegas et al. (2013). The last universe consists of 1,357 out of 14,617 European Stocks with returns from 1988-2008 reported in Datastream, analyzed as part of Banegas et al. (2013)'s study. For systematic factors, the market excess return (Mkt), small-minus-big capitalization (SMB), high-minus-low book-to-market ratio (HML), and winning-minus-losing (WML) momentum factors for the US and Europe come from Ken French's website.

Table 1 reports summary statistics describing the cross-section of expectations and risk factors for each of the three universes in the restricted and unrestricted simulation models. The US stock universe has less heterogeneity in security returns than the European stock universe, but more than European mutual funds. Key return statistics for each factor appear in Table 2, indicating slightly higher risk premia for the Market, HML, and MOM factors in Europe than the US.

Because the simulation universe is constructed from securities that have survived for over 75% of the data sample, these factors are only used for calibrating the model itself rather than as performance benchmarks for the portfolios. The influence of this survivor bias is evident in the average returns reported in the summary statistics of table 1, which typically exceed those of the market benchmark. As a notable exception, however, the average European mutual fund underperforms its market benchmark while maintaining neutral factor loadings. Though the universe constituents itself might be biased relative to the neutral benchmark, the simulation analysis evaluates portfolio allocation strategies based on their *relative* performance with respect to the  $1/N$  rule and other strategies rather than their absolute performance with respect to market benchmarks.

Table 1: Simulation Universe Cross-sectional Return Properties

This table reports the cross-sectional return properties and risk exposures for simulated asset universes. Panels A, B, and C report on the universe of US Stocks (1963-2015) from the CRSP database, European Stocks (1988-2008) from Datastream, and European Equity Mutual Funds (1988-2008) from the Lipper database, respectively. The Restricted Moments column reports the mean and volatility of returns calibrated to a restricted (zero-alpha) four-factor model where factor loadings are calculated by regressing assets' excess returns on benchmark factors, ignoring missing data, with idiosyncratic volatility equal to the residual standard deviation. The usual four benchmark factors (Mkt, SMB, HML, Mom) come from Ken French's website, with benchmark expected returns and covariances calibrated to historical sample moments. The Augmented Moment Model extracts ten latent priced factors from the security universe using principal components and calibrates security returns to a fourteen factor (four benchmarks + 10 PCAs). The column Alpha reports the mispricing in the expected return when the augmented model is fit to only the four observed benchmark factors. Appendix Table A1 reports on the expectations and factor loadings for the augmented factors.

Panel A: CRSP Stock Universe ( $N = 1,063$ Assets)									
	Restricted Moments		Benchmark Factor Loadings				Augmented Moment Model		
	Mean	Vol	Mkt	SMB	HML	MOM	Alpha	Mean	Vol
Average	8.05%	38.92%	0.96	0.47	0.41	(0.08)	0.22%	8.28%	38.86%
Std Deviation	2.98%	15.48%	0.32	0.53	0.40	0.21	0.98%	3.09%	15.40%
1%-Quantile	1.35%	13.37%	0.27	(0.49)	(0.79)	(0.77)	-2.38%	1.46%	13.36%
10%-Quantile	4.42%	22.17%	0.56	(0.15)	(0.07)	(0.34)	-0.82%	4.54%	22.17%
50%-Quantile	7.96%	35.76%	0.96	0.43	0.45	(0.05)	0.20%	8.05%	35.75%
90%-Quantile	11.86%	57.93%	1.36	1.13	0.86	0.13	1.24%	12.25%	57.70%
99%-Quantile	15.55%	92.81%	1.75	1.97	1.25	0.31	3.17%	16.45%	92.43%
Panel B: European Stock Universe ( $N = 1,357$ Assets)									
	Restricted Moments		Benchmark Factor Loadings				Augmented Moment Model		
	Mean	Vol	Mkt	SMB	HML	MOM	Alpha	Mean	Vol
Average	8.70%	33.58%	0.88	0.63	0.37	(0.03)	-0.02%	8.69%	33.59%
Std Deviation	4.49%	10.71%	0.35	0.52	0.50	0.23	3.02%	4.40%	10.71%
1%-Quantile	-4.62%	13.81%	0.20	(0.61)	(1.40)	(0.64)	-7.90%	-3.68%	13.80%
10%-Quantile	3.27%	22.29%	0.45	0.02	(0.20)	(0.32)	-3.22%	3.31%	22.27%
50%-Quantile	9.00%	31.71%	0.86	0.61	0.44	(0.00)	-0.21%	8.89%	31.73%
90%-Quantile	13.93%	48.46%	1.36	1.25	0.88	0.23	3.57%	13.95%	48.39%
99%-Quantile	18.40%	66.98%	1.82	2.12	1.26	0.49	9.66%	18.82%	67.01%
Panel C: European Mutual Fund Universe ( $N = 878$ Assets)									
	Restricted Moments		Benchmark Factor Loadings				Augmented Moment Model		
	Mean	Vol	Mkt	SMB	HML	MOM	Alpha	Mean	Vol
Average	7.05%	17.09%	1.00	0.21	(0.05)	(0.01)	-0.14%	6.90%	17.00%
Std Deviation	1.87%	2.85%	0.16	0.34	0.24	0.09	1.83%	2.05%	2.73%
1%-Quantile	2.86%	12.41%	0.60	(0.24)	(0.72)	(0.22)	-3.77%	2.23%	12.56%
10%-Quantile	4.99%	13.99%	0.81	(0.13)	(0.30)	(0.12)	-2.34%	4.50%	14.09%
50%-Quantile	6.91%	16.63%	1.00	0.11	(0.07)	(0.01)	-0.31%	6.87%	16.52%
90%-Quantile	9.47%	20.98%	1.20	0.74	0.27	0.10	2.56%	9.36%	20.68%
99%-Quantile	12.29%	25.67%	1.32	1.19	0.52	0.25	5.00%	13.00%	25.27%

Knowing the means and variances of the return generating process allows the investor to calculate the true optimal portfolio weights, which can then characterize the maximum potential benefit from diversification relative to benchmark strategies. Table 3 presents the performance statistics and portfolio characteristics for  $1/N$  rule, and the optimized portfolios that minimize variance and maximize the Sharpe Ratio. It's important to note that, absent estimation error, these universes represent environments in which an investor can

Table 2: Simulation Universe Benchmark Return Properties

This table reports the historical average, volatility, Sharpe Ratio, and Certainty Equivalent Utility for the benchmarks defining systematic factors in the US and European simulation universes. These four benchmark factors include market excess return (Mkt), small-minus-big capitalization (SMB), high-minus-low book-to-market ratio (HML), and winning-minus-losing (WML) momentum factors for the US and Europe taken from Ken French’s website. The annualized Expected Return and Volatility is calculated from the sample mean and standard deviation of historical monthly returns, while the Certainty Equivalent (CE) utility is calculated based on the mean-variance utility function with no estimation error.

	Panel A: US Benchmark Factors				Panel B: Europe Benchmark Factors			
	Mkt	SMB	HML	MOM	Mkt	SMB	HML	MOM
Expected Return	6.17%	2.36%	4.24%	8.50%	7.79%	-1.00%	7.46%	12.70%
Volatility	15.93%	10.83%	10.35%	15.43%	14.21%	8.23%	7.90%	13.19%
Sharpe Ratio	0.39	0.22	0.41	0.55	0.55	(0.12)	0.94	0.96
CE Utility								
$\gamma = 1$	4.91%	1.77%	3.70%	7.30%	6.78%	-1.34%	7.14%	11.83%
$\gamma = 2$	3.64%	1.19%	3.16%	6.11%	5.77%	-1.68%	6.83%	10.96%
$\gamma = 4$	1.10%	0.01%	2.09%	3.73%	3.75%	-2.36%	6.21%	9.22%
$\gamma = 8$	-3.97%	-2.33%	-0.05%	-1.03%	-0.29%	-3.71%	4.96%	5.74%

substantially improve their utility by deviating from naïve weights. Without non-negativity constraints, the maximized Sharpe Ratios range from two to four times as large as the Sharpe Ratio for the  $1/N$  rule and the volatility of the minimum variance portfolio is an order of magnitude smaller. Even constraining portfolio weights to be non-negative still allows for Sharpe Ratios almost twice as large as and volatilities equal to half of the benchmark  $1/N$  rule. The optimized weights themselves also deviate substantially from the  $1/N$  rule, with unconstrained optimal weights ranging from -2.95% to 3.37% in the augmented universe and constrained weights equal to zero for over 75% of the securities in each universe.

Lastly, consider the utility loss from subsetting rather than optimizing over the full universe of securities. Figure 3 plots the population complete subset efficient frontier for each of the six simulation universes with subsets of size 5, 10, 25, 50, 100, and  $N$ . These graphs highlight the opportunities (or limited presence thereof) for improved portfolio performance by deviating from the equal weighted  $1/N$  rule. For the augmented universes, with a richer covariance structure and additional return heterogeneity, the subset efficient frontier expands further and faster than for the restricted universes. Further, European markets, which feature greater heterogeneity among stocks and mutual funds with different country exposures,

Table 3: Simulation Universe Asset Optimized Portfolio Return Properties

This table reports the expected performance properties, including the annualized Expected Return, Volatility, and Sharpe Ratio, along with quantiles from the cross-sectional distribution of optimized weights for each of the six simulation universes. The optimized portfolios minimize variance (MinVar) and maximize Sharpe Ratio (MaxSR) allowing for arbitrary short positions (Unconstrained) and subject to no-shorting restrictions (Non-Negative), along with the 1/N portfolio for reference. Data for the universe of US Stocks (1963-2015) comes from the CRSP database, European Stocks (1988-2008) from Datastream, and European Equity Mutual Funds (1988-2008) from the Lipper database. The Restricted model of returns fits a zero-alpha four-factor model by regressing assets' excess returns on benchmark factors, with idiosyncratic volatility equal to the residual standard deviation. The four benchmark factors (Mkt, SMB, HML, Mom) come from Ken French's website, with factor expected returns and covariances equaling historical sample moments. The Augmented model extracts an additional ten latent priced factors from the security universe using principal components and calibrates security returns to a fourteen factor (four benchmarks + 10 PCAs).

Panel A.1: CRSP Stocks (Restricted)						Panel A.2: CRSP Stocks (Augmented)				
	1/N	Unconstrained		Non-Negative		1/N	Unconstrained		Non-Negative	
		MinVar	MaxSR	MinVar	MaxSR		MinVar	MaxSR	MinVar	MaxSR
Expectation	8.05%	0.58%	10.81%	2.74%	7.48%	8.28%	0.58%	14.37%	2.98%	8.15%
Volatility	16.63%	2.54%	10.92%	5.53%	9.40%	16.70%	2.82%	14.10%	6.06%	10.17%
Sharpe Ratio	0.484	0.230	0.990	0.495	0.796	0.496	0.204	1.019	0.491	0.801
Weight Distribution										
5%-Quantile	0.09%	-0.29%	-0.47%	0.00%	0.00%	0.09%	-0.33%	-0.75%	0.00%	0.00%
25%-Quantile	0.09%	-0.09%	-0.11%	0.00%	0.00%	0.09%	-0.10%	-0.24%	0.00%	0.00%
50%-Quantile	0.09%	0.01%	0.07%	0.00%	0.00%	0.09%	0.02%	0.09%	0.00%	0.00%
75%-Quantile	0.09%	0.16%	0.25%	0.00%	0.00%	0.09%	0.18%	0.37%	0.00%	0.00%
95%-Quantile	0.09%	0.65%	0.77%	0.00%	0.42%	0.09%	0.67%	0.99%	0.02%	0.56%
Panel B.1: European Stocks (Restricted)						Panel B.2: European Stocks (Augmented)				
	1/N	Unconstrained		Non-Negative		1/N	Unconstrained		Non-Negative	
		MinVar	MaxSR	MinVar	MaxSR		MinVar	MaxSR	MinVar	MaxSR
Expectation	8.70%	0.70%	16.01%	3.30%	10.88%	8.69%	0.54%	43.42%	3.17%	14.32%
Volatility	12.52%	2.04%	9.71%	4.01%	7.83%	12.70%	2.79%	24.94%	6.24%	11.19%
Sharpe Ratio	0.695	0.345	1.648	0.822	1.388	0.684	0.195	1.741	0.508	1.280
Weight Distribution										
5%-Quantile	0.07%	-0.21%	-0.27%	0.00%	0.00%	0.07%	-0.40%	-1.24%	0.00%	0.00%
25%-Quantile	0.07%	-0.07%	-0.07%	0.00%	0.00%	0.07%	-0.14%	-0.41%	0.00%	0.00%
50%-Quantile	0.07%	0.01%	0.03%	0.00%	0.00%	0.07%	0.03%	0.08%	0.00%	0.00%
75%-Quantile	0.07%	0.13%	0.20%	0.00%	0.00%	0.07%	0.18%	0.57%	0.00%	0.00%
95%-Quantile	0.07%	0.45%	0.55%	0.00%	0.52%	0.07%	0.52%	1.54%	0.02%	0.20%
Panel C.1: European Mutual Funds (Restricted)						Panel C.2: European Mutual Funds (Augmented)				
	1/N	Unconstrained		Non-Negative		1/N	Unconstrained		Non-Negative	
		MinVar	MaxSR	MinVar	MaxSR		MinVar	MaxSR	MinVar	MaxSR
Expectation	7.05%	0.49%	24.39%	4.48%	10.33%	6.90%	0.17%	179.40%	2.99%	13.27%
Volatility	13.94%	2.09%	14.85%	7.18%	11.01%	14.07%	2.83%	92.75%	8.11%	14.79%
Sharpe Ratio	0.506	0.232	1.643	0.624	0.938	0.490	0.059	1.934	0.369	0.897
Weight Distribution										
5%-Quantile	0.11%	-1.26%	-2.16%	0.00%	0.00%	0.11%	-2.95%	-28.17%	0.00%	0.00%
25%-Quantile	0.11%	-0.49%	-0.90%	0.00%	0.00%	0.11%	-1.02%	-10.11%	0.00%	0.00%
50%-Quantile	0.11%	0.02%	0.01%	0.00%	0.00%	0.11%	0.12%	0.11%	0.00%	0.00%
75%-Quantile	0.11%	0.72%	1.03%	0.00%	0.00%	0.11%	1.15%	9.82%	0.00%	0.00%
95%-Quantile	0.11%	1.63%	2.77%	0.00%	0.00%	0.11%	3.37%	28.26%	0.00%	0.00%

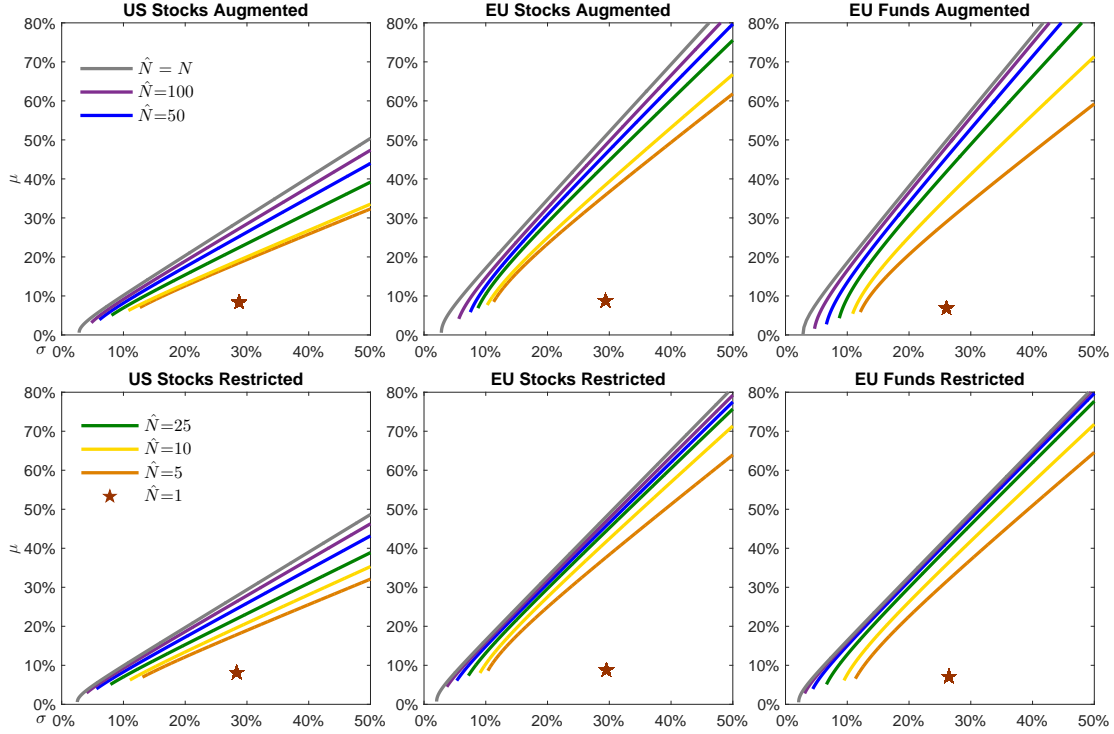


Figure 3: Population Subset Efficient Frontier for Simulated Asset Universes

This population subset efficient frontier characterizes the tradeoff between the expected return on a portfolio  $\mu$  and the portfolio's volatility  $\sigma$  when return generating process moments are known by the investor. The star represents the mean and variance of the  $1/N$  portfolio, corresponding to subset portfolios of size one. Different colors correspond to different subset portfolio sizes. Data for the universe of US Stocks (1963-2015) comes from the CRSP database, European Stocks (1988-2008) from Datastream, and European Equity Mutual Funds (1988-2008) from the Lipper database. The Restricted model of returns fits a zero-alpha four-factor model by regressing assets' excess returns on benchmark factors, with idiosyncratic volatility equal to the residual standard deviation. The four benchmark factors (Mkt, SMB, HML, Mom) come from Ken French's website, with factor expected returns and covariances equaling historical sample moments. The Augmented model extracts an additional ten latent priced factors from the security universe using principal components and calibrates security returns to a fourteen factor (four benchmarks + 10 PCAs).

demonstrate greater expansion in the efficient frontier as subset size grows than US stocks. In all universes, though the benefits from moving beyond subsets of size 100 is negligible, there is clear potential for improvement beyond the  $1/N$  rule even with subsets containing only five assets.

### 4.3 Moment Estimators and Portfolio Strategies

For each simulation iteration, investors observe sixty months of randomly generated returns on all securities ( $\{\{r_{i,\tau}\}_{\tau=1}^{60}\}_{i=1}^N$ ) and the benchmark factors ( $\{r_{F,\tau}\}_{\tau=1}^{60}$ ) and use that data

to choose their portfolio weights. Though implementing all proposed approaches to asset allocation discussed in section 2.2 is a hopeless undertaking, the tests here span a rich collection of such techniques intended to compare the performance of subset portfolios in a competitive environment. This subsection reviews the set of implemented portfolio strategies, including many of the models presented in DeMiguel et al. (2007).

Many statistical portfolio strategies optimize subjective mean-variance utility based on an investor’s risk-aversion coefficient  $\gamma$  as presented in equation (1), which are implemented with risk aversion parameter  $\gamma = 2$  (section 6.2 evaluates the influence of this specification on subset strategy performance). A full investment budget constraint restricts the weights to sum to unity. In the “unconstrained” problem, the absolute value of the weight assigned to any single security is restricted to be less than 1,000% to address the frequent simulation samples with unbounded optimal portfolio weights (most commonly encountered when implementing non-subset strategies). The “constrained” problem implements strategies subject to the common no-shorting condition that restricts weights to be non-negative.

#### **4.3.1 Benchmark Fixed Portfolio Strategy**

Within the simulation setting, the the naïve Equally Weighted (**EW**) portfolio assigning weights of  $1/N$  to all securities in the investment universe provides the most natural benchmark. Conceptually, this naïve strategy is optimal when the expected returns are proportional to total portfolio risk of a security (DeMiguel et al., 2007), when investors have extremely ambiguous beliefs (Pflug et al., 2012), or as anchoring the subset portfolio algorithm with portfolios of size one. DeMiguel et al. (2007) illustrate the strategy’s robust performance in a variety of settings, finding it to present a challenging benchmark for asset allocation, particularly in US markets.

#### **4.3.2 Subjective Expected Utility Portfolio Strategies**

The simulations implement a number of Bayesian and “plug-in” portfolio strategies for solving the mean-variance optimization problem (1) using estimated moments for population

moments, both with informative and diffuse prior specifications. Given the well-known motivation that forming optimal portfolio weights using sample means and variances results in poor performance, the subjective expected utility maximizing strategies focus on Bayesian parameter estimates. Variations in the prior structure imposed by these estimators on means, variances, and covariances differentiates these models.

The first three fitted models implement a Bayes-Stein shrinkage for assets' expected returns following Jorion (1986) and Frost and Savarino (1986). These hierarchical Bayesian models shrink the sample average of the individual security's return toward the grand mean average return taken across all securities:

$$\hat{\mu}_{i,BS} = \phi_i \hat{\mu}_{i,S} + (1 - \phi_i) \bar{\mu}, \quad \phi_i \equiv \frac{\sigma_{\mu,i}^{-2}}{\sigma_{\mu,i}^{-2} + \bar{\sigma}_{\mu}^{-2}}, \quad \hat{\mu}_{i,S} \equiv \frac{1}{T} \sum_{t=1}^T r_{i,t}, \quad \bar{\mu} \equiv \frac{1}{N} \sum_{i=1}^N \hat{\mu}_{i,S} \quad (13)$$

$$\sigma_{\mu,i}^2 \equiv \left( \frac{1}{T-1} \sum_{t=1}^T (r_{i,t} - \hat{\mu}_{i,S})^2 \right) / T, \quad \bar{\sigma}_{\mu}^2 \equiv \frac{1}{N} \sum_{n=1}^N \sigma_{\mu,i}^2 + \frac{1}{N} \sum_{n=1}^N (\hat{\mu}_{i,S} - \bar{\mu})^2$$

The three Bayes-Stein Mean models differ in their estimator for the covariance matrix of returns. The Bayes-Stein Sample (**BS-S**) model estimates the covariances between securities using the sample covariance matrix estimator. The Bayes-Stein Bayes-Stein (**BS-SO**) model estimates the covariances between securities using the Stein-optimal covariance matrix estimator proposed in Gillen (2014) with bandwidth parameter of 1%. Finally, the Bayes-Stein Ledoit-Wolf (**BS-LW**) model uses the Ledoit and Wolf (2004) shrinkage covariance matrix estimator that shrinks the sample covariance matrix towards a constant correlation covariance matrix.

The next class of Bayesian models incorporate the structure implied by a factor model for expected returns following Pastor (2000). The "Data and Model" posterior expectations and covariances are updated using a natural conjugate regression specification where investors' prior beliefs characterize the degree of mispricing in a four-factor model. Investor beliefs are encapsulated in a prior for the parameter reflecting mispricing of securities, denoted  $\alpha$ , which is assumed to take the conjugate normal distribution with mean 0 and standard

deviation  $\sigma_\alpha \in \{0.01\%, 1\%, 10\%\}$  while prior beliefs about factor loadings are diffuse. At the lowest extreme,  $\sigma_\alpha = 0.01\%$  imposes a dogmatic belief in the factor model **(DM-D)** that is exactly right in the baseline simulation model but misspecified in the augmented model. The intermediate value of  $\sigma_\alpha = 1\%$  represents an informed belief in the factor model **(DM-I)**, and the extreme an uninformed belief in the factor model **(DM-U)**. Each of these models impose the covariance matrix restriction that idiosyncratic returns are uncorrelated.

### 4.3.3 Minimum Variance Portfolio Strategies

Chopra and Ziemba (1993) identify sampling error in expected returns as a primary source of poor performance in plug-in portfolio strategies. This result suggests the global minimum variance portfolio often delivers better performance than portfolios that seek to optimize investor’s expected utility. Completely ignoring the expected return on individual securities, three minimum variance portfolios use the different estimators for the covariance matrix of securities applied to the Bayes-Stein plug-in portfolios. The first of these uses the sample covariance matrix **(GMV-S)**, the second uses the Stein optimal covariance matrix **(GMV-SO)**, and the last uses Ledoit and Wolf (2004)’s shrinkage estimator with a constant correlation prior **(GMV-LW)**.

### 4.3.4 Subset Portfolio Strategies

The simulation tests implement only the simplest formulation of complete subset portfolios to focus exclusively on the algorithm’s role in refining portfolio performance. The reference specification for subset portfolios uses sample estimates for expectations and the variance-covariance matrix. All subsets are formed with a fixed number of ten securities in each subset and averaging across 100,000 subsets (so that on average, each security is selected into approximately 1,000 subset portfolios). Robustness tests considered in section 6 explore different specifications of the subset strategy, including varying the number of securities in each subset and the number of subsets generated to illustrate how these tuning parameters affect subset portfolio performance. These robustness checks indicate that the reference

specification presented here provides a conservative characterization of subset portfolios’ potential performance.

#### 4.4 Simulating Portfolio Strategies’ Certainty Equivalent Utility

The most relevant metric with which to evaluate a portfolio’s performance is the true certainty equivalent utility (CEU) realized by the investor choosing the strategy. Given true expectations  $\mu$  and covariance matrix  $\Sigma$  of returns, the CEU for an investor implementing strategy  $k$  in the  $m^{\text{th}}$  simulation is the expected out-of-sample performance for the estimated portfolio weights  $\hat{w}_{k,m}$ :

$$CEU_{k,m} = \hat{w}'_{k,m}\mu - \frac{\gamma}{2}\hat{w}'_{k,m}\Sigma\hat{w}_{k,m} \quad (14)$$

Table 4 Panels A and B report the simulated average and standard deviation of each strategy’s CEU, with Panel C characterizing the frequency with which the portfolio strategy delivered a CEU greater than that achieved by implementing the  $1/N$  rule. Additional information about the average simulated expected return, volatility and Sharpe Ratios for these strategies are reported in the appendix Table A2.

The robust performance of subsetting relative to alternative estimators is striking. In terms of average CEU from Long-Short portfolios, the complete subset portfolio ranks as the highest or second-highest performing strategy across all universes and simulation specifications. Complete subset portfolios’ long-short strategies deliver up to 9% improvement in the CEU over that of the  $1/N$  rule. Though the magnitude of improvement for complete subset portfolios’ long-only strategies over the  $1/N$  rule is somewhat muted 50 to 150bp<sup>6</sup>. Still, that outperformance is reliable across simulations, with long-only complete subset portfolios’ out-of-sample performance exceeding that of the  $1/N$  rule in well over 90% of the simulation samples for US Stock environments and effectively 100% of simulations for European stocks and mutual funds.

---

<sup>6</sup>Since constraining the weights bounds their sampling variance, reducing the sensitivity of portfolio performance to estimation error, it seems likely the optimal subset size for each subset portfolio would much larger than ten (as implemented here). Section 6.1 will present additional discussion of this effect, along with simulation evidence regarding the performance of subset portfolios with different size specifications.

Table 4: Simulated Certainty Equivalent Utility from Portfolio Strategies

This table reports statistics on the Certainty Equivalent Utility (CEU) for an investor with mean-variance preferences 1 and risk aversion parameter  $\gamma = 2$  when implementing the portfolio algorithms presented in section 4.3 based on simulated 60-month histories of returns. All portfolios implement the budget constraint that weights sum to unity, with the Long-Short portfolios constraining weights to be between -1000% and 1000% and the Long-Only portfolios constraining weights to be non-negative. Panel A reports the average CEU across 500 simulations for the Restricted Four-Factor model and Augmented Factor model that includes an additional 10 principal components of returns as described in section 4.1. Panel B reports the simulated standard deviation of the CEU and Panel C reports the simulated frequency with which the strategy's CEU exceeds that of the 1/N portfolio.

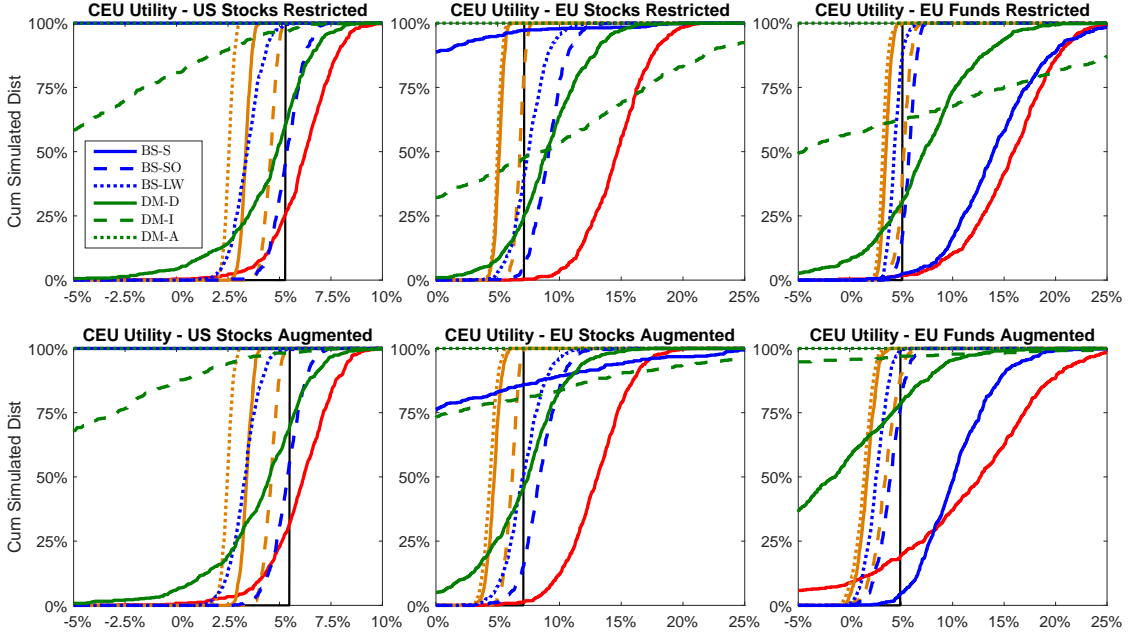
		Panel A: Simulated Average Certainty Equivalent Utility From Portfolio Strategy												
		Restricted Factor Models					Augmented Factor Models							
		Long-Short Portfolios		Long-Only Portfolios			Long-Short Portfolios		Long-Only Portfolios					
$T = 60, \gamma = 2$		EUFund	EUStock	EUFund	EUStock	USStock	EUFund	EUStock	EUFund	EUStock	USStock	EUFund	EUStock	USStock
1/N		5.11%	7.13%	5.11%	7.13%	5.29%*	4.92%	7.08%	4.92%	7.08%	5.49%*	4.92%	7.08%*	5.49%*
Subset		14.80%**	14.77%**	6.23%**	6.22%	8.64%**	9.52%**	12.70%**	9.52%**	12.70%**	6.18%**	6.09%	8.39%**	5.86%**
GMV		3.55%	5.12%	3.36%	4.61%	3.89%	1.81%	4.66%	3.41%	3.10%	3.41%	3.10%	5.12%	3.93%
GMV-SO		5.27%	6.82%	4.54%	5.35%	4.71%	3.37%	6.09%	4.54%	4.35%	4.54%	4.35%	6.24%	4.79%
GMV-LW		3.25%	4.96%	2.54%	4.09%	2.54%	1.48%	4.35%	2.45%	2.63%	2.45%	2.63%	2.93%	2.68%
Bayes-Stein														
BS-S		14.47%*	-69.65%	-131.36%	4.66%	4.11%	10.70%**	-76.62%	-132.00%	3.15%	5.32%	3.15%	5.32%	4.14%
BS-SO		5.80%	9.24%*	5.40%*	5.43%	5.01%	3.92%	8.44%*	5.37%	4.44%	6.75%	4.44%	6.75%	5.07%
BS-LW		4.41%	7.55%	3.43%	4.17%	2.79%	2.78%	7.09%	3.31%	2.70%	3.49%	2.70%	3.49%	2.91%
Data-Model														
DM-D		6.98%	9.11%	4.53%	5.74%	4.95%	-4.10%	7.14%	4.04%	5.30%	5.93%	5.30%	4.80%	4.80%
DM-I		-26.72%	1.91%	-10.77%	6.64%*	4.83%	-327.04%	-41.18%	-14.31%	6.32%*	6.72%	6.32%*	4.82%	4.82%
DM-U		-1,065%	-3,459%	-1,745%	7.28%**	3.59%	-5,894%	-8,040%	-2,221%	7.75%**	7.02%	7.75%**	7.02%	3.73%
		Panel B: Simulated Standard Deviation of Certainty Equivalent Utility												
		Restricted Factor Models					Augmented Factor Models							
		Long-Short Portfolios		Long-Only Portfolios			Long-Short Portfolios		Long-Only Portfolios					
$T = 60, \gamma = 2$		EUFund	EUStock	EUFund	EUStock	USStock	EUFund	EUStock	EUFund	EUStock	USStock	EUFund	EUStock	USStock
1/N		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Subset		4.03%	2.59%	1.48%	0.40%	0.29%	8.99%	2.61%	1.44%	0.43%	0.42%	0.43%	0.25%	0.25%
GMV		0.45%	0.39%	0.25%	0.43%	0.44%	0.80%	0.57%	0.27%	0.57%	0.46%	0.57%	0.46%	0.46%
GMV-SO		0.70%	0.41%	0.30%	0.51%	0.37%	1.10%	0.59%	0.36%	0.53%	0.66%	0.53%	0.37%	0.37%
GMV-LW		0.47%	0.35%	0.23%	0.53%	0.20%	0.82%	0.58%	0.24%	0.57%	0.45%	0.57%	0.19%	0.19%
Bayes-Stein														
BS-S		4.20%	88.45%	137.50%	0.43%	0.47%	3.61%	91.17%	130.47%	0.56%	0.83%	0.56%	0.49%	0.49%
BS-SO		0.74%	1.27%	0.70%	0.51%	0.38%	1.12%	1.66%	0.70%	0.54%	0.61%	0.54%	0.43%	0.43%
BS-LW		0.65%	1.44%	0.73%	0.51%	0.28%	1.07%	1.68%	0.64%	0.57%	0.64%	0.57%	0.27%	0.27%
Data-Model														
DM-D		4.78%	3.58%	2.17%	1.15%	0.86%	12.26%	3.85%	2.67%	1.44%	1.91%	1.44%	0.91%	0.91%
DM-I		74.68%	21.95%	16.44%	1.02%	0.84%	391.63%	62.80%	16.46%	1.22%	2.07%	1.22%	0.91%	0.91%
DM-U		935.60%	1661.28%	834.40%	1.30%	1.37%	5486.10%	4236.92%	1260.57%	1.93%	2.45%	1.93%	1.10%	1.10%
		Panel C: Frequency of Certainty Equivalent Utility > 1/N Certainty Equivalent Utility												
		Restricted Factor Models					Augmented Factor Models							
		Long-Short Portfolios		Long-Only Portfolios			Long-Short Portfolios		Long-Only Portfolios					
$T = 60, \gamma = 2$		EUFund	EUStock	EUFund	EUStock	USStock	EUFund	EUStock	EUFund	EUStock	USStock	EUFund	EUStock	USStock
Subset		96%*	99%**	75%**	99%**	90%**	84%*	99%**	74%**	100%**	100%**	100%**	93%**	93%**
GMV		0%	0%	0%	13%	0%	0%	0%	0%	0%	1%	0%	1%	1%
GMV-SO		60%	20%	0%	68%	5%	8%	3%	0%	16%	8%	16%	1%	1%
GMV-LW		0%	0%	0%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Bayes-Stein														
BS-S		99%**	4%	0%	16%	0%	95%**	14%	0%	0%	1%	0%	1%	1%
BS-SO		84%	95%*	53%*	71%	21%	21%	84%*	44%*	18%	34%	18%	16%	16%
BS-LW		16%	58%	0%	4%	0%	3%	45%	0%	0%	0%	0%	0%	0%
Data-Model														
DM-D		65%	76%	41%	69%	39%*	24%	59%	31%	63%	34%	63%	23%	23%
DM-I		33%	51%	4%	94%	31%	1%	15%	1%	88%	48%	88%	28%*	28%*
DM-U		0%	0%	0%	95%**	8%	0%	0%	0%	94%*	53%*	94%*	3%	3%

\*\*-Highest ranked strategy

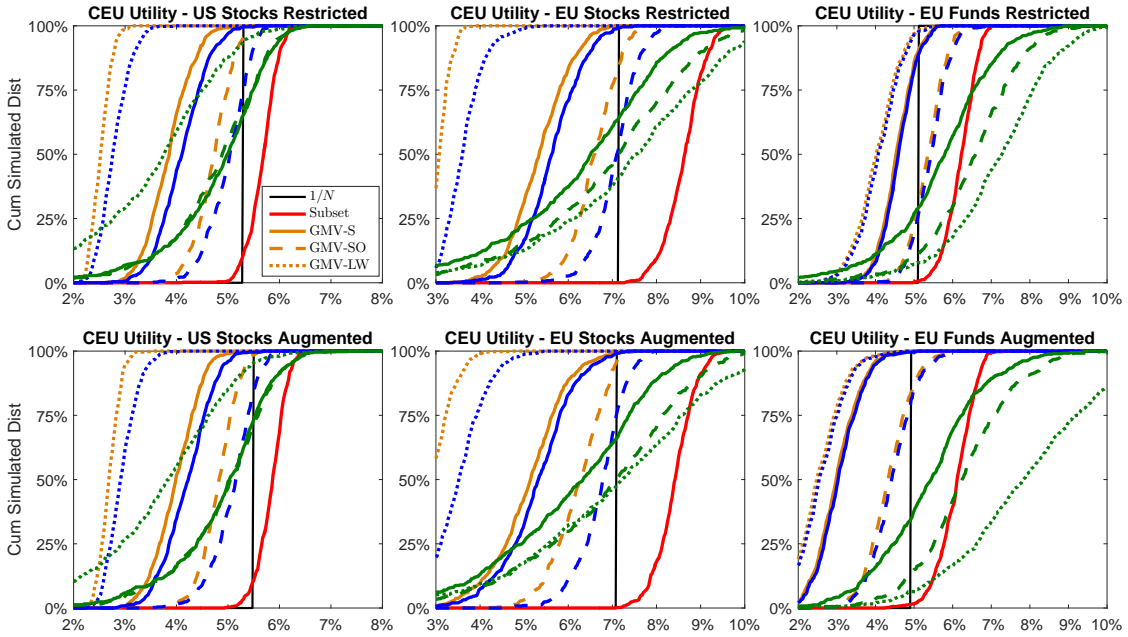
The alternative strategies' performance varies widely depending on the investment universe. The uninformed Data & Model investor (DM-U) performs well outside of the US Stock universe when constrained to non-negativity portfolios, delivering higher average CEU than any other strategy in the simulation for European Mutual Funds. This same strategy delivers disastrous results in long-short portfolios, where the variability of portfolio performance cause the average CEU to be an order of magnitude below -100%, implying investors would prefer to simply give away all their wealth to following these strategies. The Bayes-Stein models perform well in some settings but not others, with the BS-SO model performing reasonably well in the European universes. However, the only strategies that deliver reasonable performance across all settings are the GMV portfolios, though these consistently underperform the  $1/N$  rule.

Looking at the simulated variability of the investor's CEU, complete subset portfolios display a good deal variability across simulation draws, especially for long-short portfolios. Still, it's performance is quite stable relative to that of Data & Model strategies and the BS-S Bayes-Stein strategy using the sample covariance matrix. Further, subset portfolios' relative performance holds up across simulation draws, and in several specifications subset portfolios outperform the  $1/N$  rule in every simulation sample. In the most adverse environment, complete subset portfolios outperform the  $1/N$  rule in over 74% of simulation samples. To that end, these results replicate the near-optimality of the  $1/N$  rule, especially compared to existing portfolio formation strategies, but also illustrate how subset portfolios can systematically outperform naïve strategies.

Figure 4.4 presents a graphical characterization of the sensitivity of a strategy's performance to sampling error by plotting the cumulative distributions of simulated certainty equivalents across the different strategies. Here, the subset portfolio strategy's consistent outperformance is demonstrated by the stochastic dominance relationship between the distribution of subset portfolios' CEU and that of other strategies. Across all specifications, the CEU for subset portfolios exceeds that of any other strategy in over 75% of the simulations.



Panel A: Long-Short Portfolios



Panel B: Long-Only Portfolios

Figure 4: Simulated Distribution of Portfolio Certainty Equivalent Utilities,  $T = 60, \gamma = 2$

This figure plots the distribution of Certainty Equivalent Utility (CEU) for an investor with mean-variance preferences (1) and risk aversion parameter  $\gamma = 2$  when implementing the portfolio algorithms presented in section 4.3 based on 500 simulated 60-month histories of returns. All portfolios implement the budget constraint that weights sum to unity, with the Long-Short portfolios constraining weights to be between -1,000% and 1,000% and the Long-Only portfolios constraining weights to be non-negative. The black line represents the CEU of the 1/N rule while the solid red line represents the cumulated distribution for the simulated CEU of Subset portfolios with size  $\hat{N} = 10$ . The blue lines represent the performance of Empirical Bayesian estimators, with the different dashed presentations corresponding to different covariance matrices. The green lines characterize that of the Data and Model estimators, with the different dashed presentations corresponding to different prior beliefs.

## 5 Backtest Performance of Subset Portfolios

While simulation tests evaluate the subset algorithm in an ideal static environment, live financial markets feature dynamic variation in the return generating process that may complicate the algorithm’s implementation. To this end, a backtest experiment evaluates a strategy’s performance in light of the sort of dynamic uncertainty investors have faced in the past and might expect to face in the future. This approach provides additional validation for the strategy’s performance while also characterizing properties like turnover or drawdown for the implemented strategy.

This section implements a rolling backtest with the full sample of data used to define simulation universes from the previous section without filtering for data coverage so there are no concerns with survivor bias. Portfolio weights are reoptimized monthly using a rolling window with sixty months of returns. At each reoptimization period, stock investors filter the universe for those assets with no missing returns in the trailing sixty months and then selects those among the highest 30% of market capitalizations. Mutual fund investors select the 30% of funds with the lowest estimated volatilities (unfortunately the coverage of Assets Under Management and Expense Ratios in the Lipper database is too thin to provide a viable filter for investability). The investor calculates weights for these assets using each of the portfolio strategies considered in the simulation study, constraining weights for each security to be non-negative and less than 10%. When assets in the portfolio are delisted, investors realize the delisting return and reallocate the holdings to the risk-free security.

Table 5 reports summary performance statistics for each of the implemented strategies across the three universes. The geometric means for the complete subset portfolios of size  $\hat{N} = 100$  consistently rank among the top two strategies, delivering the most consistent outperformance over the  $1/N$  rule of any strategy across all three universes. The Data and Model strategies also perform reasonably well, particularly in the stock universes, but the relative performance of these models is sensitive to both the universe and the prior parameterization. Though the Minimum Variance and Empirical Bayesian strategies deliver relatively high Sharpe ratios, their total return performance is somewhat underwhelming.



The complete subset portfolio with a smaller size  $\hat{N} = 10$  delivers solid performance across all three universes and regularly outperforms the  $1/N$  rule, though not to the extent of the more active complete subset portfolio of size  $\hat{N} = 100$ .

Looking into the results more closely, the US stock universe presents a clearly challenging environment to improve on the  $1/N$  rule, which outperforms the Market portfolio in 55% of months. The backtested efficiency of the  $1/N$  rule is consistent with the simulation tests in which only the Subset strategy could consistently outperform the  $1/N$  rule. The minimum variance portfolios and Empirical Bayes strategies were effective in generating high Sharpe Ratios, but their average returns are relatively low. Though the uninformed Data and Model strategy performs well, the informed and dogmatic models' performance is somewhat underwhelming. These results indicate that Bayesian portfolio estimators, with a properly tuned prior, can match the performance of subset portfolios. In unreported results, subset optimization using moments from the Uninformed Data and Model estimator improves the performance of both the Complete Subset Portfolio and the Uninformed Data and Model strategies.

The European Stock universe presented more opportunity for active (relative to the  $1/N$  rule) portfolios to deliver improved performance. Here, the uninformed data and model strategy again performed well, delivering the highest geometric average return across strategies, albeit with slightly lower Sharpe Ratios. Complete Subset portfolios provided the second-highest geometric average return and most consistent outperformance of the  $1/N$  rule. Minimum variance and Empirical Bayesian strategies delivered the highest Sharpe Ratios as well as some extreme outliers in terms of benchmark-adjusted performance as measured by the portfolio alpha.

Due to (on average) poor returns in individual funds, the  $1/N$  rule applied to European Mutual Funds frequently underperforms its market portfolio benchmark. Overcoming the average performance of individual securities in the universe, not only do subset portfolios consistently outperform the  $1/N$  rule, but they also outperform the market benchmark in more than half of the months, presenting the only strategy that doesn't deliver negative

alpha for the portfolio. In this universe, complete subset portfolios of size  $\hat{N} = 100$  dominate the other strategies in terms of geometric returns, Sharpe Ratios, and benchmark outperformance.

The cost of implementing active portfolios lies in their turnover. Here, subset portfolios are comparable to active strategies but, at 200% per year, is likely to generate significant trading costs. However, these costs can be managed by reducing the frequency of reoptimization and rebalancing, as results from unreported tests found quarterly rebalancing delivers comparable performance with roughly half the turnover.

The results of this section could be readily extended to evaluate their robustness with respect to a number of design specifications and universes. Unreported robustness checks have considered the threshold for market capitalizations, the maximum weight in securities, frequency of rebalancing, and estimation window size. All these tests deliver qualitatively similar results to those reported here and none demonstrate poor performance for subset optimization. Other unreported backtests implement subset optimization for sorted portfolios from Ken French's website, with the complete subset portfolios delivering higher Sharpe Ratios than the  $1/N$  rule in every universe. These myriad tests demonstrate well the robust performance of complete subset portfolios.

## 6 Robustness and Comparative Statics for Subset Portfolios

The previous sections' subset portfolio implementation adopted a fixed subset size and number of subsets portfolios for simplicity of presentation. This section explores the subset portfolio algorithm's robustness to these tuning parameters using simulation tests from section 4. Given the focus on the relative performance of subset portfolios under different parameterizations and relative to the  $1/N$  rule, results for other strategies are unreported.

## 6.1 Subset Size, Sample Size, and Portfolio Performance

Increasing subset size expands the population subset efficient frontier but also increases the sensitivity of portfolio weights to sampling error. To evaluate this trade-off, this subsection implements simulation tests for subset portfolios of various sizes ( $\hat{N} \in \{1, 10, 25, 50\}$ ) with different sample sizes available ( $T \in \{30, 60, 120, 240\}$ ), generating 400 simulations for each specification.

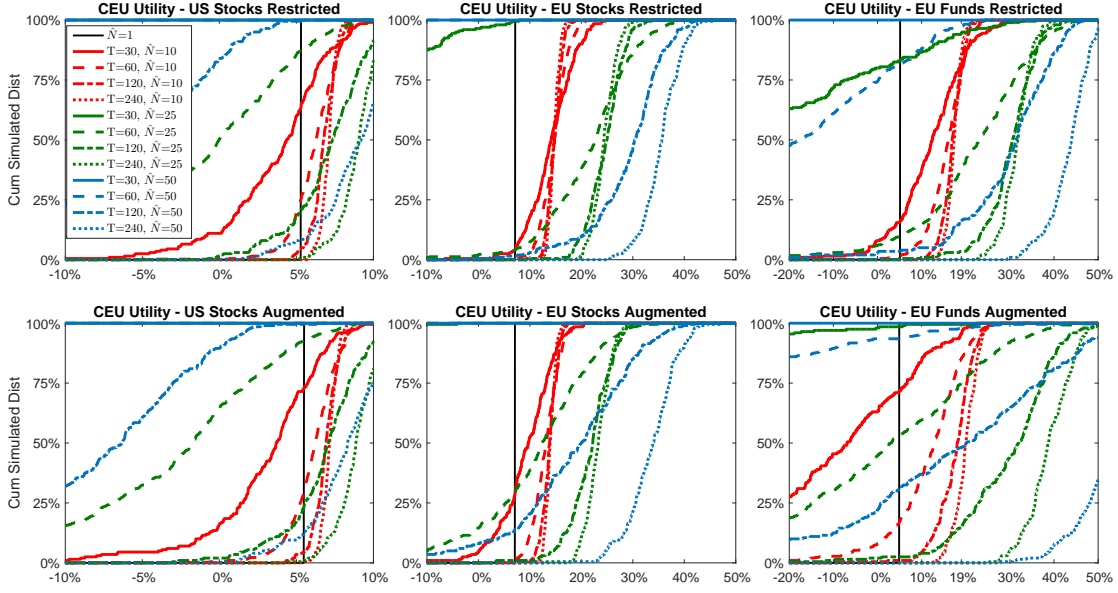
The figure 6.1 summarizes the results of these simulations by presenting the cumulative densities for the out-of-sample Certainty Equivalent Utility by subset size. For each fixed subset size, the simulated distribution of CEU's shift right and become more concentrated on the population complete subset utility. Looking across subset sizes, portfolio performance clearly deteriorates substantially when the subset size is large relative to the sample size. When  $T = 30$ , smaller subsets outperform larger subsets and even  $\hat{N} = 10$  often underperforms the  $1/N$  portfolio in simulations based on US Equities. However, as the sample size increases, larger subset sizes more consistently realize their potential to outperform smaller subset sizes. Additional tables in the appendix present further performance statistics from these experiments, including average certainty equivalent utilities, expected returns, volatilities, and Sharpe Ratios. Appendix tables A3 and A4 present results for simulation experiments where the sample size is fixed at  $T = 60$  but the subset sizes vary substantially, with  $\hat{N} \in \{1, 2, 5, 10, 25, 50, 100\}$ . Tables A5 and A6, also in the appendix, present the additional performance statistics for the specifications presented in figure 6.1.

## 6.2 Risk Aversion and Subset Portfolio Strategies

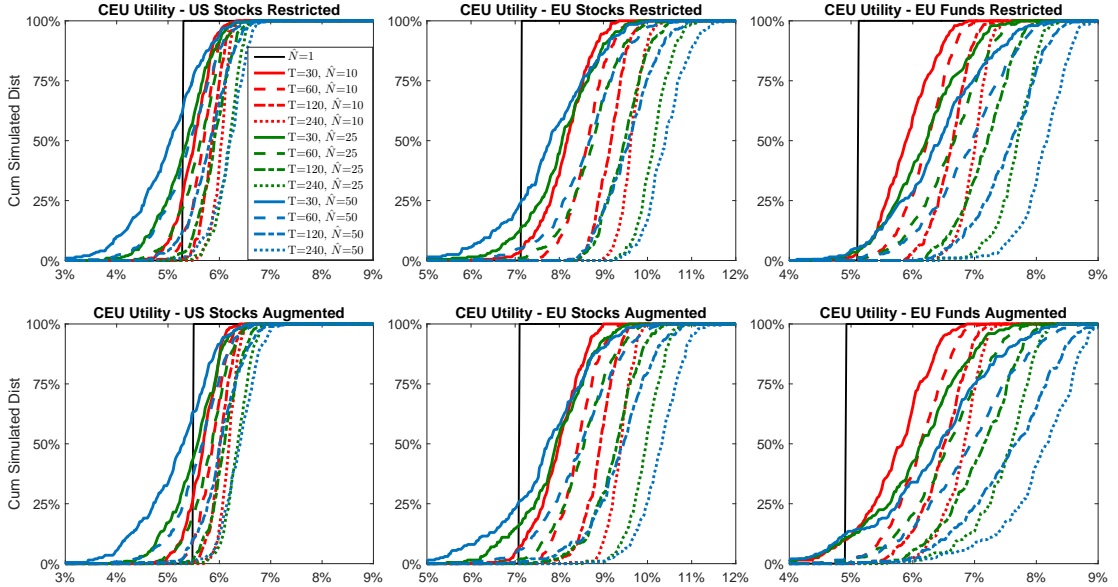
Complete subset portfolios allow investors to incorporate their risk preferences into the portfolio optimization process through the parameter  $\gamma$ . For extremely low levels of  $\gamma$ , investors become risk neutral and seek higher expected returns.<sup>7</sup> For extremely low levels

---

<sup>7</sup>Without constraints on individual security weights, the risk neutral investor's portfolio weights are undefined when  $\gamma = 0$ . With non-negativity constraints, this investor seeks to invest all of their portfolio in the security with the highest expected return. This property has interesting implications for the risk neutral complete subset portfolio weights subject to non-negativity constraints, as the weight a security receives in



Panel A: Long-Short Portfolios



Panel B: Long-Only Portfolios

Figure 5: Simulated Distribution of Portfolio Certainty Equivalent Utilities,  $T = 60, \gamma = 2$   
This figure plots the distribution across 400 simulations of Certainty Equivalent Utilities (CEU) realized by an investor with mean-variance objective from Equation (1) parameterized by risk aversion coefficient  $\gamma = 2$  access to  $T \in \{30, 60, 120, 240\}$  months of normally-distributed return data. Portfolio weights are calculated by the subset portfolio algorithm using sample estimates for asset expectations and covariances with subset sizes  $\hat{N} \in \{1, 10, 25, 50\}$  both allowing for long and short positions (Panel A) and restricting weights to be non-negative (Panel B). The six subplots within each panel relate to the US Stock, European Stock, and European Mutual Fund Universes calibrated according to the Restricted four-factor model and the PCA Augmented fourteen factor models presented in section 4.1.

of  $\gamma$ , investors seek out the minimum variance portfolio, comparable to the global minimum variance portfolio strategies.

Table 6 reports the average expected returns, volatilities, and Sharpe Ratios from these simulation experiments. The comparative performance of these portfolios match expectations, with lower values for  $\gamma$  delivering higher expected returns and higher levels delivering lower volatilities. Interestingly, the volatility of portfolios with  $\gamma = 50$  very closely matches that of the minimum variance portfolio from the infeasible subset efficient frontier. Appendix table A7 presents summary statistics relating to the certainty equivalence of the portfolio's utility.

## 7 Potential Refinements and Extensions

The subsetting approach is very flexible and provides a portable mechanism for large-scale allocation decisions that can be applied to other problems such as forecast combination and optimal risk sharing. As the current project's scope focuses on a reference implementation of the subset strategy and illustrating its robustness to various design parameters, this exposition hasn't fully explored potential refinements to the subset strategy even as it applies exclusively to the asset allocation problem. This section discusses potentially interesting approaches for refining and extending the subset strategy.

### 7.1 Subset Portfolios with Alternative Objective Functions

Though the subset application developed here focuses on optimizing the mean-variance objective function, a number of other utility functions could be readily implemented. For instance, a fund manager may be interested in optimizing their alpha relative to their tracking error. Other risk measures, such as Value-at-Risk are readily incorporated into the subset algo-

---

each subset is an indicator variable of whether that security has the highest expected return. For subsets of size  $\hat{N}$ , the complete subset portfolio weight for a security is proportional to the cross-sectional quantile of that security's average return raised to the  $\hat{N} - 1$  power. In this way, long-only complete subset portfolios for risk-neutral investors represent a transformation of the rank ordering of average historical returns.

Table 6: Subset Portfolio Simulated Expectations, Volatilities, and Sharpe Ratios: Variable Subset and Sample Size

This table reports statistics on the expected return (Panel A), volatility (Panel B), and Sharpe Ratio (Panel C) realized by an investor with mean-variance preferences 1 after observing  $T = 60$  months of returns and implementing subset portfolios of size  $\hat{N} = 10$  under variable values for the risk aversion parameter  $\gamma = 2$ . The reported performance averages across 400 simulations for the Restricted Four-Factor model and Augmented Factor model that includes an additional 10 principal components of returns as described in section 4.1. All portfolios implement the budget constraint that weights sum to unity, with the Long-Short portfolios constraining weights to be between -1000% and 1000% and the Long-Only portfolios constraining weights to be non-negative.

		Panel A: Simulated Average Expected Return by Risk Aversion ( $\gamma$ )							
		Restricted Factor Models			Augmented Factor Models				
		Long-Short Portfolios		Long-Only Portfolios		Long-Only Portfolios			
$T = 60, \hat{N} = 10$	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
1/ $N$	7.05%	8.70%	8.05%	7.05%	8.70%	8.05%	6.90%	8.69%	8.28%
Mean-Variance									
$\gamma = 0.1$	2661.83%	1920.65%	737.57%	8.25%	10.18%	9.25%	3570.26%	1951.76%	773.87%
$\gamma = 1$	36.43%	27.34%	13.59%	8.20%	10.34%	8.87%	49.24%	27.43%	14.07%
$\gamma = 2$	21.32%	17.76%	9.93%	8.14%	10.27%	8.51%	27.42%	17.66%	10.22%
$\gamma = 5$	12.26%	12.00%	7.73%	7.91%	9.89%	7.81%	14.31%	11.80%	7.91%
$\gamma = 10$	9.23%	10.09%	7.00%	7.61%	9.41%	7.24%	9.95%	9.85%	7.14%
$\gamma = 25$	7.42%	8.93%	6.56%	7.15%	8.79%	6.76%	7.33%	8.67%	6.68%
$\gamma = 50$	6.82%	8.55%	6.41%	6.91%	8.51%	6.60%	6.45%	8.28%	6.52%
		Panel B: Simulated Average Volatility of Return by Risk Aversion ( $\gamma$ )							
		Restricted Factor Models			Augmented Factor Models				
		Long-Short Portfolios		Long-Only Portfolios		Long-Only Portfolios			
$T = 60, \hat{N} = 10$	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
1/ $N$	13.94%	12.52%	16.63%	13.94%	12.52%	16.63%	14.07%	12.70%	16.70%
Mean-Variance									
$\gamma = 0.1$	3364.43%	2349.64%	2390.62%	14.38%	14.06%	19.85%	6304.21%	3250.47%	2629.84%
$\gamma = 1$	40.59%	27.82%	29.24%	14.12%	13.34%	18.04%	73.51%	36.19%	31.30%
$\gamma = 2$	22.78%	17.03%	18.58%	13.86%	12.80%	16.79%	38.71%	21.04%	19.47%
$\gamma = 5$	13.38%	11.58%	13.32%	13.24%	11.77%	14.70%	19.26%	13.29%	13.63%
$\gamma = 10$	11.07%	10.22%	11.97%	12.61%	10.89%	13.23%	14.03%	11.41%	12.16%
$\gamma = 25$	10.11%	9.56%	11.30%	11.94%	10.02%	12.10%	11.88%	10.61%	11.44%
$\gamma = 50$	9.87%	9.38%	11.10%	11.72%	9.72%	11.78%	11.44%	10.41%	11.24%
		Panel C: Simulated Average Sharpe Ratio for Return by Risk Aversion ( $\gamma$ )							
		Restricted Factor Models			Augmented Factor Models				
		Long-Short Portfolios		Long-Only Portfolios		Long-Only Portfolios			
$T = 60, \hat{N} = 10$	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
1/ $N$	0.506	0.695	0.484	0.506	0.695	0.484	0.490	0.684	0.496
Mean-Variance									
$\gamma = 0.1$	0.832	0.827	0.303	0.575	0.727	0.468	0.592	0.616	0.288
$\gamma = 1$	0.934	1.003	0.468	0.583	0.778	0.493	0.697	0.775	0.452
$\gamma = 2$	0.972	1.069	0.588	0.588	0.805	0.509	0.733	0.853	0.535
$\gamma = 5$	0.944	1.054	0.590	0.599	0.842	0.533	0.756	0.893	0.589
$\gamma = 10$	0.847	0.994	0.588	0.604	0.866	0.548	0.713	0.864	0.590
$\gamma = 25$	0.738	0.936	0.581	0.599	0.879	0.559	0.617	0.818	0.584
$\gamma = 50$	0.693	0.913	0.578	0.589	0.876	0.561	0.564	0.796	0.581

rithm. Multi-objective strategies balancing expectations, risk, and factor exposures could also be adopted.

Subsetting has particularly interesting potential as a device for constructing portfolios that hedge against specific risk factors, such as a factor-neutral or portable alpha investment strategy. These strategies often realize residual factor exposure due to estimation error in the factor exposures used when constructing the hedge. By diversifying across a large number of subset portfolios that are each factor-neutral, subsetting could enhance the robustness of the hedge with respect to estimation error in these estimated exposures.

## **7.2 Weighted and Clustered Selection for Subset Portfolios**

While the uniform selection in subsetting admits the exchangeability property under which the equal weighting of subsets is ex-ante optimal, weighted selection nests other natural portfolio strategies. For instance, if securities are selected proportionally to their market capitalizations, subsets of size one would correspond to the market portfolio weights. In this light, subsetting can provide a device for mean-variance enhancement of fundamental-driven long-only portfolio weights. One conceptual challenge arises as the exchangeability property may no longer hold, suggesting a weighting of subset portfolios according to the likelihood of their constituents' selection. Second, the selection weights themselves need to be determined. When constructing a fund focusing on outperforming a specific benchmark, the benchmark weights themselves provide a natural candidate.

One limitation of the subset strategy lies in its treatment of all securities as if they were identical. This problem is particularly relevant to investors interested in spanning a set of asset classes, as many subsets would consist entirely of securities in a single asset class. To address this, investors may wish to form clusters of securities, for instance by industry, asset class, investment objective, or security fundamentals. Forming subsets by selecting a single security from each cluster then ensures the aggregated portfolio will achieve diversification across asset classes, providing a mechanism for applying subset portfolios to asset-class level allocation problems.

### 7.3 Enhancing Aggregation for Subset Portfolios

Another set of potential enhancements and extensions could investigate improvements upon the equal-weighting of each subset portfolio. One interesting question in this regard could investigate the properties of iterated applications of the subset portfolio algorithm. For instance, consider taking  $N$  assets and constructing  $B_1$  subset portfolios of size  $\hat{N}$ . Using the sample return data on these  $B_1$  subset portfolios, then construct  $B_2$  subset portfolios of size  $\hat{N}$ . One particularly interesting question is whether an iterated application of the subsetting algorithm would eventually lead the weights for individual securities to converge.

Another interesting enhancement might filter subset portfolios based on their interim expected CEU, concentrating the aggregated portfolio in high-CEU subset portfolios. Like optimizing the weights for subset portfolios, filtering subset portfolios will enhance the aggregated portfolio's estimated utility. The cost of this enhancement lies in an indeterminate increase in the variability of portfolio weights. Considering issues raised in Novy-Marx (2016), some care would be necessary in implementing this filtration to avoid spurious results driven by data mining on randomly generated signals.

### 7.4 Bayesian Refinements for Subset Portfolio Weights

The complete subset portfolio weights have a natural interpretation as the posterior expectation of a Bayesian trying to learn about optimal portfolio weights while uncertain about the optimal portfolio's size and constituents. To the extent that Bayesian posterior expectations or hybrid shrinkage portfolios can enhance the CEU for each individual subset portfolio, adopting these strategies when optimizing subset weights could further enhance aggregated portfolio performance. Such an interpretation also provides a natural mechanism for developing strategies for weighting subset portfolios.

## 8 Conclusion

This paper proposes subset optimization as a new algorithm for asset allocation that’s particularly useful in settings with many securities and short return histories. As a satisficing algorithm, subset optimization sheds new light on the relationship between the number of securities in the universe and the performance of data-driven asset allocation. Given DeMiguel et al. (2007), it’s well-known that naïve  $1/N$  weighting performs well in a variety of asset universes. Subset optimization provides a principled approach to defining the “ $N$ ” constituents in the universe to which investors ought apply this naïve strategy. In so doing, subset optimization overcomes the curse of dimensionality by exploiting the large asset universe to diversify the adverse impact of estimation error on portfolio performance.

When analyzing the properties of portfolio strategies based on aggregating multiple signals, Novy-Marx (2016) succinctly describes the main motivation for implementing subset portfolios: “The basic tenants of Markowitz’s (1952) modern portfolio theory hold, and efficient combinations of high Sharpe ratio assets have even higher Sharpe ratios.” Subset portfolios provide a simple and effective approach to enhancing the out-of-sample expected performance of investment portfolios by forming approximately efficient combinations of high Sharpe ratio portfolios. Though presented here in its most simple and straightforward implementation, substantial potential remains for refining the algorithm in application to asset allocation as well as for adapting the subset approach to other applications.

## References

- Evan W. Anderson and Ai-Ru (Meg) Cheng. Robust Bayesian portfolio choices. *Review of Financial Studies*, 29(5):1330–1375, 2016.
- Doron Avramov and Guofu Zhou. Bayesian portfolio analysis. *Annual Review of Financial Economics*, 2:25–47, 2010.
- Doron Avramov and Tarun Chordia. Predicting stock returns. *Journal of Financial Economics*, 82(2):387–415, 2006.
- Doron Avramov and Russ Wermers. Investing in mutual funds when returns are predictable. *Journal of Financial Economics*, 81:339–377, 2006.

- Ayelen Banegas, Benjamin Gillen, Allan Timmermann, and Russ Wermers. The cross-section of conditional mutual fund performance in European stock markets. *Journal of Financial Economics*, 108(3):699–726, 2013.
- Fischer Black and Robert Litterman. Global portfolio optimization. *Financial Analysts Journal*, 48:28, 1992.
- Michael W. Brandt. Portfolio choice problems. In Yacine Aït-Sahalia and Lars Peter Hansen, editors, *Handbook of Financial Econometrics: Tools and Techniques*, number 1, chapter 5, pages 269–336. Elsevier B.V., 2010.
- Michael W. Brandt, Pedro Santa-Clara, and Rossen Valkanov. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies*, 22(9):3411–3447, 2009.
- Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1995. ISSN 1573-0565.
- Marine Carrasco and Nérée Noumon. Optimal portfolio selection using regularization. *Mimeo*, 2012.
- Vijay Kumar Chopra and William T. Ziemba. The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19(2):6–11, 1993.
- Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the  $1/N$  portfolio strategy? *Review Financial Studies*, 2007.
- Victor DeMiguel, Lorenzo Garlappi, Francisco Nogales, and Raman Uppal. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55:798–812, 2009.
- Graham Elliott, Antonio Gargano, and Allan Timmermann. Complete subset regressions with large-dimensional sets of predictors. *Journal of Economic Dynamics and Control*, 54: 86 – 110, 2015. ISSN 0165-1889.
- Edwin J Elton and Martin J Gruber. Risk reduction and portfolio size: An analytical solution. *The Journal of Business*, 50(4):415–437, 1977.
- John L. Evans and Stephen H. Archer. Diversification and the reduction of dispersion: An empirical analysis. *The Journal of Finance*, 23(5):761–767, 1968.
- Eugene F Fama. Stock returns, real activity, inflation, and money. *The American Economic Review*, 71(4):545–565, 1981.
- Eugene F Fama and Kenneth R French. The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465, 1992.

- Jianqing Fan, Jingjin Zhang, and Ke Yu. Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107:592–606, 2012.
- Wayne E Ferson and Campbell R Harvey. The risk and predictability of international equity returns. *Review of Financial Studies*, 6(3):527–566, 1993.
- Wayne E Ferson and Campbell R Harvey. Conditioning variables and the cross section of stock returns. *The Journal of Finance*, 54(4):1325–1360, 1999.
- Peter A Frost and James E Savarino. An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, 21(3):293–305, SEP 1986.
- Lorenzo Garlappi, Raman Uppal, and Tan Wang. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies*, 20(1):41–81, JAN 2007.
- Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992.
- John Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Federal Reserve Bank of Minneapolis Staff Report No 148*, 1991.
- Itzhak Gilboa and David Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141 – 153, 1989.
- Benjamin J Gillen. An empirical bayesian approach to stein-optimal covariance matrix estimation. *Journal of Empirical Finance*, 29:402–420, 2014.
- Donald Goldfarb and Garud Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, FEB 2003.
- Vasyl Golosnoy and Yarema Okhrin. Flexible shrinkage in portfolio selection. *Journal of Economic Dynamics and Control*, 33:317–328, 2009.
- Eric Jacquier and Nicholas G Polson. Asset allocation in finance: A Bayesian perspective. *Hierarchical models and MCMC: a Tribute to Adrian Smith*, 2012.
- Ravi Jagannathan and Tongshu Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58(4):1651–1683, AUG 2003.
- Philippe Jorion. Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3):279–292, SEP 1986.
- Raymond Kan and Guofu Zhou. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656, SEP 2007.
- Donald B Keim and Robert F Stambaugh. Predicting returns in the stock and bond markets. *Journal of Financial Economics*, 17(2):357–390, 1986.

- Mark Kritzman. Are optimizers error maximizers? *The Journal of Portfolio Management*, 32(4):66–69, 2006.
- Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, FEB 2004.
- Olivier Ledoit and Michael Wolf. Optimal estimation of a large-dimensional covariance matrix under Stein’s loss. *Mimeo*, 2013.
- John Liechty, Campbell R. Harvey, and Merrill W. Liechty. Bayes vs. resampling: A rematch. *Journal of Investment Management*, 2008.
- Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, March 1952.
- Harry Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. Blackwell Publishing, 1959.
- Harry Markowitz and Nilufer Usmen. Resampled frontiers vs diffuse Bayes: An experiment. *Journal of Investment Management*, 1:9–25, 2003.
- Richard Michaud. *Efficient Assset Management: A Practial Guide to Stock Portfolio Optimization*. Oxford University Press, 1998.
- Robert Novy-Marx. Testing strategies based on multiple signals. *Mimeo*, 2016.
- Yarema Okhrin and Wolfgang Schmid. Distributional properties of portfolio weights. *Journal of Econometrics*, 134:235–256, 2006.
- Lubos Pastor. Portfolio selection and asset pricing models. *Journal of Finance*, 55(1):179–223, FEB 2000.
- Lubos Pastor and Robert Stambaugh. Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, 56(3):335–381, JUN 2000.
- Lubos Pastor and Robert Stambaugh. Investing in equity mutual funds. *Journal of Financial Economics*, 63(3):351–380, MAR 2002.
- Georg Ch Pflug, Alois Pichler, and David Wozabal. The  $1/N$  investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance*, 36(2):410–417, 2012.
- David Rapach and Guofu Zhou. Forecasting stock returns. In Graham Elliott and Allan Timmermann, editors, *Handbook of Economic Forecasting*, volume 2, Part A of *Handbook of Economic Forecasting*, chapter 6, pages 328 – 383. Elsevier, 2013.
- Leonard Savage. *The Foundations of Statistics*. Dover Publications, 1954.
- G William Schwert. The adjustment of stock prices to information about inflation. *The Journal of Finance*, 36(1):15–29, 1981.

- William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964.
- Meir Statman. How many stocks make a diversified portfolio? *Journal of Financial and Quantitative Analysis*, 22(03):353–363, 1987.
- Charles Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1955.
- Jun Tu and Guofu Zhou. Incorporating economic objectives into Bayesian priors: Portfolio choice under parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 45: 959–986, 2010.

## Appendix A1: Proofs

**Theorem 1.** *Suppose  $N$  assets' returns have mean  $\mu$  and variance  $\Sigma$ . Let  $\bar{w}_{\hat{N}_1}^*$  and  $\bar{w}_{\hat{N}_2}^*$  represent the complete population subset portfolio weights where the subset optimization step (II) minimizes the variance for a subset portfolio of size of size  $\hat{N}_1 > \hat{N}_2$  subject to a minimal expected return constraint,  $\underline{\mu}$ :*

$$\bar{w}_{\hat{N}}^*(\underline{\mu}) = \lim_{B \rightarrow \infty} \frac{1}{B} \sum_{b=1}^B w_b^*(\hat{N}, \underline{\mu}) \quad w_b^*(\hat{N}, \underline{\mu}) = \arg \min_{w \in \Delta^{\hat{N}-1}} w' \Sigma_b w \text{ subject to } w' \mu_b \geq \underline{\mu}$$

*Then the variance of the complete population subset portfolio of size  $\hat{N}_1$  is lower than the variance of the aggregated subset portfolio of size  $\hat{N}_2$ :*

$$\bar{w}_{\hat{N}_1}^{*'} \Sigma \bar{w}_{\hat{N}_1}^* < \bar{w}_{\hat{N}_2}^{*'} \Sigma \bar{w}_{\hat{N}_2}^*$$

*Proof.* The proof proceeds inductively, constructing larger subset portfolios by replicating smaller subset portfolios.

Let  $\mathcal{W}_{\hat{N}}$  denote the complete set of subset portfolios used to calculate  $\bar{w}_{\hat{N}}^*$ . To construct the complete set of subset portfolios of size  $\hat{N} + 1$ , first select a population subset portfolio  $w_b^* \in \mathcal{W}_{\hat{N}}$  and recall the notation of  $\mathcal{N}_b$  as representing its constituent assets. Now iteratively replicate  $\mathcal{N}_b$  and augment it by including one of the excluded assets, constructing  $N - \hat{N}$  replications:  $\tilde{\mathcal{N}}_{b,1}, \dots, \tilde{\mathcal{N}}_{b,N-\hat{N}}$ . Let  $\tilde{w}_{b,i}^*$  denote the population optimal portfolio weights for subset  $\tilde{\mathcal{N}}_{b,i}$ .

The complete population subset portfolio weights of size  $\hat{N} + 1$  can be calculated by averaging these  $N - \hat{N}$  subset weights. Further, whereas  $\tilde{w}_{b,i}^*$  solves:

$$\arg \min_{w \in \Delta^{\hat{N}-1}} w' \Sigma_{b,i} w \text{ subject to } w' \mu_{b,i} \geq \underline{\mu}$$

$w_b^*$  solves the more constrained problem:

$$\arg \min_{w \in \Delta^{\hat{N}-1}} w' \Sigma_{b,i} w \text{ subject to } w' \mu_{b,i} \geq \underline{\mu} \text{ and } w_i = 0.$$

Clearly, the minimizing variance from  $\tilde{\mathcal{N}}_{b,i}$  is lower than the constrained minimizer in  $\mathcal{N}_b$ , so that  $\tilde{w}_{b,i}^{*'} \Sigma_{b,i} \tilde{w}_{b,i}^* \leq w_b^{*'} \Sigma_b w_b^*, \forall b, i$ . Since the variance for an equal-weighted combination of these  $N - \hat{N}$  subset portfolios less than the maximal variance of any such portfolio, the variance of the portfolio  $\bar{\bar{w}}_b \equiv \frac{1}{N-\hat{N}} \sum_{i \in N \setminus \mathcal{N}_b} \tilde{w}_{b,i}^*$  is less than the variance of the portfolio  $w_b^*$ .

This argument applies for any subset  $b$ , extending directly to the complete subset portfolio. The key property necessary for this extension is the simple result that the variance of an equal weighted portfolio is less than the average variance of the portfolio's constituents. Since the subset portfolio weights of size  $\hat{N}$  are simply  $\bar{w}_{\hat{N}}^* = \frac{1}{B} \sum_{b=1}^B w_b^*$  and the subset port-

folio weights of size  $\hat{N} + 1$  are  $\bar{w}_{\hat{N}+1}^* = \frac{1}{B} \sum_{b=1}^B \bar{w}_b$ , the result is established for  $\hat{N}_1 = \hat{N}_2 + 1$  with induction completing the proof.  $\square$

**Theorem 2.** *Suppose an investor observes  $T$  observations of returns on  $N$  securities with mean  $\mu$  and covariance matrix  $\Sigma$  uses equation (5) to calculate the weights of each sample subset portfolio and holds the complete subset portfolio weights  $\hat{w}_{\hat{N}}$ . Then:*

1. *The estimated complete subset portfolio weights are unbiased for the population complete subset portfolio weights:*

$$E[\hat{w}_{\hat{N}}] = \bar{w}_{\hat{N}}^*$$

2. *The sample variance of an estimated subset portfolio weight is  $O(T^{-1})$ .*

3. *The expected out of sample performance for the estimated complete subset portfolio equals that of the population subset portfolio minus a  $O(\hat{N}^2/T)$  penalty:*

$$E\left[\hat{w}'_{\hat{N}}\mu - \frac{\gamma}{2}\hat{w}'_{\hat{N}}\Sigma\hat{w}_{\hat{N}}\right] = U_b^* - \frac{\gamma}{2}O\left(\hat{N}^2/T\right)$$

*Proof.* The first result holds trivially by the unbiasedness of subset portfolio weight estimates and the linearity of the expectation operator. The second result follows from Okhrin and Schmid (2006)'s Theorem 1, Part (b) that characterizes the variance of the weights when  $T > \hat{N} + 4$  and  $\hat{N} > 4$ , rewritten by defining  $\sigma_{b,GMV}^2$  as the variance of the minimum variance portfolio for the securities in subset  $\mathcal{N}_b$ :

$$\begin{aligned} \text{Var}(\hat{w}_b) &= c_0 R_b \sigma_{b,GMV}^2 + c_1 R_b \mu_b \mu_b' R_b + c_2 \mu_b' R_b \mu_b R_b + c_3 R_b \\ c_0 &= \frac{1}{T - \hat{N} - 1} \quad c_1 = \frac{(T - 1)^2 (T - \hat{N} + 1)}{\gamma^2 (T - \hat{N}) (T - \hat{N} - 1)^2 (T - \hat{N} - 3)}, \quad c_2 = c_1 \frac{T - \hat{N} - 1}{T - \hat{N} + 1}, \\ c_3 &= \frac{1}{T} c_1 + \frac{1}{T} c_2 (\hat{N} - 1) + \frac{(T - 1)^2}{\gamma^2 T (T - \hat{N} - 1)^2}, \quad \hat{R}_b = \hat{\Sigma}_b^{-1} - \frac{\hat{\Sigma}_b^{-1} \mathbf{1} \mathbf{1}' \hat{\Sigma}_b^{-1}}{\mathbf{1}' \hat{\Sigma}_b^{-1} \mathbf{1}} \end{aligned}$$

Holding  $\hat{N}$  fixed, note that the constants  $c_0$ ,  $c_1$ ,  $c_2$ , and  $c_3$  are each  $O(T^{-1})$ , completing the proof for part 2.

Proving Part 3 begins with the result from equation (8):

$$\begin{aligned} E\left[\hat{w}'_{\hat{N}}\mu - \frac{\gamma}{2}\hat{w}'_{\hat{N}}\Sigma\hat{w}_{\hat{N}}\right] &= \bar{w}_{\hat{N}}^*\mu - \frac{\gamma}{2}\bar{w}_{\hat{N}}^*\Sigma\bar{w}_{\hat{N}}^* - \frac{\gamma}{2}\text{tr}(\Sigma V_{\hat{w}_{\hat{N}}}) \\ &= U_{\hat{N}}^* - \frac{\gamma}{2}\text{tr}(\Sigma V_{\hat{w}_{\hat{N}}}) \\ &= U_{\hat{N}}^* - \frac{\gamma}{2}\mathbf{1}'\Sigma \circ V_{\hat{w}_{\hat{N}}}\mathbf{1} \end{aligned}$$

where  $\circ$  represents the element-wise Hadamard product and  $\mathbf{1}$  is a vector of  $N$  ones.

To analyze this penalty term, first consider the variance of a single asset's weight in the complete subset portfolio,  $\hat{w}_{\hat{N},i}$ :

$$\begin{aligned} \text{Var} \left( \hat{w}_{\hat{N},i} \right) &= \text{Var} \left( \frac{1}{B} \sum_{b=1}^B \hat{w}_{b,i} \right) \\ &= \frac{1}{B^2} \sum_{a=1}^B \sum_{b=1}^B \text{Cov} (\hat{w}_{a,i}, \hat{w}_{b,i}) \mathbb{1}_{\{i \in \mathcal{N}_a \cap \mathcal{N}_b\}} \\ &\leq \frac{1}{B^2} \sum_{a=1}^B \sum_{b=1}^B \sqrt{\text{Var} (\hat{w}_{a,i}) \text{Var} (\hat{w}_{b,i})} \mathbb{1}_{\{i \in \mathcal{N}_a \cap \mathcal{N}_b\}} \end{aligned}$$

where  $\mathbb{1}_{\{i \in \mathcal{N}_a \cap \mathcal{N}_b\}}$  is an indicator variable indicating that asset  $i$  is selected into both subsets  $a$  and  $b$ . From the result (2),  $\text{Var} (\hat{w}_{b,i})$  is  $O(T^{-1})$  and there is some constant  $\phi_i$  such that  $\text{Var} (\hat{w}_{b,i}) < \frac{\phi_i}{T}$ , further simplifying the calculation allowing the application of a law of large numbers for the complete subset portfolio weight:

$$\begin{aligned} \lim_{B \rightarrow \infty} \text{Var} \left( \hat{w}_{\hat{N},i} \right) &\leq \lim_{B \rightarrow \infty} \frac{\phi_i}{TB^2} \sum_{a=1}^B \sum_{b=1}^B \mathbb{1}_{\{i \in \mathcal{N}_a \cap \mathcal{N}_b\}} \\ &= \lim_{B \rightarrow \infty} \frac{\phi_i}{TB^2} \sum_{a=1}^B \sum_{b=1}^B \text{Pr} \{i \in \mathcal{N}_a \cap \mathcal{N}_b\} \\ &= \phi_i \frac{\hat{N}^2}{TN^2} \end{aligned}$$

As a covariance matrix, the order of the diagonal entries in  $V_{\hat{w}_{\hat{N}}}$  are  $O\left(\frac{\hat{N}^2}{TN^2}\right)$  also bounds the order of the off-diagonal terms. Let  $\bar{\phi} = \max \phi_i$  and  $\bar{\sigma}$  denote the maximal entry in  $\Sigma$ , then the entries of the Hadamard product each be bounded by  $\bar{\sigma} \bar{\phi} \frac{\hat{N}^2}{TN^2}$ . Noting that there are  $N^2$  entries in the penalty's sum completes the proof, bounding the penalty by:

$$\mathbf{1}' \Sigma \circ V_{\hat{w}_{\hat{N}}} \mathbf{1} \leq \bar{\sigma} \bar{\phi} \frac{\hat{N}^2}{T}$$

□

## Appendix A2: Supplemental Figures and Tables

Table A1: Simulation Universe Cross-sectional Factor Exposures for Augmented Models

This table reports the cross-sectional return properties and risk exposures for simulated asset universes. Panel A reports on 1,063 out of 31,219 US Stocks with returns from 1963-2015 reported in the CRSP database, Panel B reports on 1,357 out of 14,617 European Stocks with returns from 1988-2008 reported in Datastream, analyzed as part of Banegas et al. (2013)'s study, and Panel C reports on 878 out of 4,955 European Equity Mutual Funds with returns from 1988-2008 reported in the Lipper database as analyzed in Banegas et al. (2013), where each universe was screened for only those stocks that have returns in 75% of the sample. The rows report cross-sectional averages, standard deviations, and selected quantiles for the parameters identified in the column labels. The Restricted Moments column reports the mean and volatility of returns calibrated to a restricted (zero-alpha) four-factor model where factor loadings are calculated by regressing assets' excess returns on benchmark factors, ignoring missing data, with idiosyncratic volatility equal to the residual standard deviation. These four benchmark factors include market excess return (Mkt), small-minus-big capitalization (SMB), high-minus-low book-to-market ratio (HML), and winning-minus-losing (WML) momentum factors for the US and Europe taken from Ken French's website, with benchmark expected returns and covariances calibrated to historical sample moments. The Augmented Moment Model extracts ten latent priced factors using principal components and calibrates security returns to a fourteen factor (four benchmarks + 10 PCAs). The Alpha reports the mispricing in the expected return when the augmented model is calibrated only to the four observed benchmark factors. Appendix Table A1 reports on the expectations and factor loadings for the augmented factors.

		Panel A: CRSP Stock Universe													
		Mkt	SMB	HML	MOM	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7	PCA8	PCA9	PCA10
Expectation		6.17%	2.36%	4.24%	8.50%	1.66%	6.18%	3.32%	7.26%	13.10%	-9.06%	-7.54%	-2.92%	2.36%	6.28%
Volatility		15.93%	10.83%	10.35%	15.43%	85.96%	78.37%	74.80%	71.20%	79.64%	72.22%	76.82%	68.45%	68.01%	64.74%
Sharpe Ratio		0.388	0.218	0.409	0.551	0.019	0.079	0.044	0.102	0.164	-0.125	-0.098	-0.043	0.035	0.097
Exposures:															
Average		0.958	0.469	0.412	-0.081	0.006	0.002	0.005	-0.002	0.008	0.001	0.003	-0.015	0.009	0.006
Std Deviation		0.319	0.530	0.399	0.205	0.037	0.039	0.040	0.038	0.042	0.040	0.045	0.042	0.043	0.042
1%-Quantile		0.261	-0.490	-0.764	-0.746	-0.071	-0.091	-0.113	-0.109	-0.088	-0.092	-0.116	-0.126	-0.081	-0.102
10%-Quantile		0.553	-0.157	-0.055	-0.337	-0.023	-0.038	-0.044	-0.040	-0.032	-0.033	-0.042	-0.064	-0.035	-0.036
50%-Quantile		0.956	0.421	0.454	-0.050	0.002	0.000	0.009	0.000	0.001	-0.005	-0.001	-0.013	0.005	0.005
90%-Quantile		1.361	1.134	0.849	0.129	0.037	0.043	0.044	0.034	0.058	0.040	0.059	0.026	0.052	0.050
99%-Quantile		1.746	2.022	1.261	0.327	0.122	0.139	0.088	0.119	0.152	0.141	0.133	0.093	0.156	0.134
		Panel B: European Stock Universe													
		Mkt	SMB	HML	MOM	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7	PCA8	PCA9	PCA10
Expectation		7.79%	-1.00%	7.46%	12.70%	22.02%	-4.62%	59.87%	59.41%	9.95%	12.86%	-3.56%	0.02%	-3.67%	9.65%
Volatility		14.21%	8.23%	7.90%	13.19%	216%	203%	189%	152%	143%	137%	134%	132%	130%	128%
Sharpe Ratio		0.548	-0.122	0.944	0.963	0.102	-0.023	0.316	0.390	0.070	0.093	-0.027	0.000	-0.028	0.075
Exposures:															
Average		0.884	0.629	0.372	-0.026	-0.001	-0.008	-0.002	0.000	0.006	-0.003	0.001	0.000	0.004	0.006
Std Deviation		0.352	0.518	0.505	0.234	0.027	0.026	0.027	0.027	0.026	0.027	0.027	0.027	0.027	0.026
1%-Quantile		0.196	-0.608	-1.406	-0.643	-0.045	-0.047	-0.050	-0.067	-0.070	-0.069	-0.069	-0.064	-0.071	-0.056
10%-Quantile		0.453	0.025	-0.194	-0.325	-0.021	-0.032	-0.031	-0.032	-0.027	-0.035	-0.029	-0.031	-0.023	-0.024
50%-Quantile		0.859	0.612	0.443	-0.004	-0.003	-0.015	-0.008	0.000	0.008	-0.003	0.002	0.000	0.003	0.005
90%-Quantile		1.356	1.249	0.877	0.227	0.014	0.028	0.038	0.029	0.037	0.028	0.032	0.032	0.034	0.038
99%-Quantile		1.817	2.123	1.256	0.482	0.156	0.076	0.072	0.078	0.068	0.064	0.073	0.080	0.076	0.076
		Panel C: European Mutual Fund Universe													
		Mkt	SMB	HML	MOM	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7	PCA8	PCA9	PCA10
Expectation		7.79%	-1.00%	7.46%	12.70%	-1.98%	22.77%	-6.80%	-14.70%	-18.32%	2.36%	-5.98%	0.95%	22.31%	-8.81%
Volatility		14.21%	8.23%	7.90%	13.19%	106.97%	71.68%	77.96%	65.23%	53.37%	57.17%	47.04%	45.66%	38.48%	38.93%
Sharpe Ratio		0.548	-0.122	0.944	0.963	-0.018	0.318	-0.087	-0.225	-0.343	0.041	-0.127	0.021	0.580	-0.226
Exposures:															
Average		0.992	0.208	-0.050	-0.009	0.002	0.022	0.013	0.004	0.016	0.006	0.002	0.004	-0.004	0.011
Std Deviation		0.150	0.336	0.238	0.090	0.036	0.027	0.035	0.035	0.032	0.036	0.036	0.035	0.036	0.035
1%-Quantile		0.603	-0.247	-0.722	-0.220	-0.064	-0.038	-0.068	-0.077	-0.073	-0.080	-0.111	-0.086	-0.094	-0.058
10%-Quantile		0.819	-0.135	-0.298	-0.120	-0.056	-0.014	-0.040	-0.049	-0.015	-0.034	-0.028	-0.038	-0.043	-0.027
50%-Quantile		0.995	0.112	-0.067	-0.013	0.009	0.022	0.012	0.007	0.015	0.000	0.001	0.003	-0.005	0.009
90%-Quantile		1.193	0.735	0.268	0.099	0.040	0.053	0.061	0.031	0.040	0.056	0.034	0.042	0.042	0.049
99%-Quantile		1.309	1.186	0.514	0.247	0.083	0.103	0.076	0.126	0.133	0.100	0.133	0.110	0.081	0.128

Table A2: Simulated Expectations, Volatilities, and Sharpe Ratios from Portfolio Strategies

This table reports statistics on the expected return (Panel A), volatility (Panel B), and Sharpe Ratio (Panel C) realized by an investor with mean-variance preferences 1 and risk aversion parameter  $\gamma = 2$  when implementing the portfolio algorithms presented in section 4.3 based on simulated 60-month histories of returns. The reported performance averages across 400 simulations for the Restricted Four-Factor model and Augmented Factor model that includes an additional 10 principal components of returns as described in section 4.1. All portfolios implement the budget constraint that weights sum to unity, with the Long-Short portfolios constraining weights to be between -1000% and 1000% and the Long-Only portfolios constraining weights to be non-negative.

		Panel A: Simulated Average Expected Return From Portfolio Strategy														
		Restricted Factor Models					Augmented Factor Models									
		Long-Short Portfolios					Long-Short Portfolios									
$T = 60, \gamma = 2$		EU Funds	EU Stock	US Stock	EU Funds	EU Stock	US Stock	EU Funds	EU Stock	US Stock	EU Funds	EU Stock	US Stock	EU Funds	EU Stock	US Stock
$1/N$		7.05%	8.70%	7.05%	8.70%	8.05%	8.05%	8.70%	8.69%	8.28%	6.90%	8.69%	8.28%	6.90%	8.69%	8.28%
Subset		21.12%	17.70%	9.89%	8.13%	10.26%	8.50%	10.26%	17.61%	10.21%	26.86%	17.61%	10.21%	8.25%	10.21%	8.74%
GMV		3.71%	5.34%	3.65%	5.17%	6.07%	4.76%	6.07%	5.23%	3.77%	2.05%	5.23%	3.77%	3.79%	6.22%	4.97%
GMV-S		5.85%	7.34%	5.21%	6.25%	7.37%	5.65%	7.37%	6.99%	5.29%	4.84%	6.99%	5.29%	5.41%	7.24%	5.78%
GMV-SO		3.41%	5.13%	2.82%	4.68%	3.28%	2.90%	3.28%	4.89%	2.86%	1.77%	4.89%	2.86%	3.30%	3.49%	3.15%
Bayes-Stein		18.52%	113.69%	47.13%	5.22%	6.42%	5.14%	6.42%	110.69%	45.71%	12.64%	110.69%	45.71%	3.85%	6.43%	5.29%
BS-S		6.40%	10.23%	6.46%	6.34%	7.98%	6.10%	7.98%	9.87%	6.51%	5.42%	9.87%	6.51%	5.53%	7.87%	6.24%
BS-SO		4.60%	7.90%	3.92%	4.75%	3.85%	3.18%	3.85%	7.93%	3.92%	3.08%	7.93%	3.92%	3.37%	4.11%	3.42%
BS-LW		11.76%	11.22%	6.89%	7.73%	8.75%	8.20%	8.75%	10.70%	6.85%	11.46%	10.70%	6.85%	7.67%	8.86%	8.49%
DM-D		68.00%	45.86%	10.74%	8.23%	9.13%	7.75%	9.13%	45.27%	10.77%	69.02%	45.27%	10.77%	8.35%	9.42%	8.11%
DM-I		285.43%	548.12%	142.75%	8.89%	11.40%	9.49%	11.40%	535.86%	146.99%	309.28%	535.86%	146.99%	11.00%	11.52%	9.89%
DM-U																
		Panel B: Simulated Average Volatility from Portfolio Strategy														
		Restricted Factor Models					Augmented Factor Models									
		Long-Short Portfolios					Long-Short Portfolios									
$T = 60, \gamma = 2$		EU Funds	EU Stock	US Stock	EU Funds	EU Stock	US Stock	EU Funds	EU Stock	US Stock	EU Funds	EU Stock	US Stock	EU Funds	EU Stock	US Stock
$1/N$		13.94%	12.52%	16.63%	13.94%	12.52%	16.63%	14.07%	12.70%	16.70%	14.07%	12.70%	16.70%	14.07%	12.70%	16.70%
Subset		22.79%	17.02%	18.54%	13.86%	12.79%	16.76%	38.51%	21.07%	19.59%	14.68%	21.07%	19.59%	14.68%	13.36%	16.94%
GMV		4.26%	4.71%	5.32%	7.91%	8.52%	9.49%	6.33%	7.28%	6.03%	8.83%	7.28%	6.03%	8.83%	10.63%	9.86%
GMV-S		7.72%	7.51%	8.07%	9.48%	8.77%	9.63%	11.80%	9.44%	8.40%	10.29%	9.44%	8.40%	10.29%	10.09%	9.79%
GMV-SO		4.00%	3.85%	5.30%	8.14%	4.51%	6.05%	6.66%	7.14%	6.30%	8.72%	7.14%	6.30%	8.72%	7.35%	6.74%
Bayes-Stein		20.00%	134.89%	125.87%	7.94%	8.85%	10.13%	14.46%	127.01%	126.81%	8.85%	127.01%	126.81%	8.85%	10.70%	10.37%
BS-S		7.94%	10.11%	10.06%	9.55%	9.55%	10.43%	12.07%	11.67%	10.38%	10.41%	11.67%	10.38%	10.41%	10.66%	10.59%
BS-SO		4.63%	5.88%	6.95%	8.17%	4.71%	6.28%	7.05%	8.92%	7.79%	7.05%	8.92%	7.79%	7.58%	7.58%	6.94%
BS-LW		19.98%	13.98%	14.56%	14.14%	15.53%	18.10%	36.81%	18.85%	15.51%	14.84%	18.85%	15.51%	14.84%	16.54%	18.60%
DM-D		91.83%	63.73%	43.79%	12.97%	14.68%	16.83%	174.16%	87.31%	49.17%	14.20%	87.31%	49.17%	14.20%	15.91%	17.37%
DM-I		347.36%	623.49%	426.95%	16.48%	20.09%	24.23%	685.88%	859.83%	476.75%	17.54%	859.83%	476.75%	17.54%	21.09%	24.75%
DM-U																
		Panel C: Simulated Average Sharpe Ratio from Portfolio Strategy														
		Restricted Factor Models					Augmented Factor Models									
		Long-Short Portfolios					Long-Short Portfolios									
$T = 60, \gamma = 2$		EU Funds	EU Stock	US Stock	EU Funds	EU Stock	US Stock	EU Funds	EU Stock	US Stock	EU Funds	EU Stock	US Stock	EU Funds	EU Stock	US Stock
$1/N$		0.506	0.695	0.484	0.506	0.695	0.484	0.490	0.684	0.496	0.490	0.684	0.496	0.490	0.684	0.496
Subset		0.962	1.067	0.546	0.588	0.805	0.509	0.724	0.849	0.531	0.563	0.767	0.518	0.563	0.767	0.518
GMV		0.871	1.136	0.686	0.654	0.714	0.501	0.321	0.718	0.626	0.428	0.718	0.626	0.428	0.586	0.504
GMV-S		0.772	0.983	0.650	0.663	0.846	0.588	0.412	0.740	0.631	0.526	0.740	0.631	0.526	0.717	0.591
GMV-SO		0.855	1.333	0.533	0.576	0.727	0.479	0.265	0.684	0.455	0.377	0.684	0.455	0.377	0.474	0.467
GMV-LW		0.921	0.817	0.345	0.658	0.727	0.508	0.871	0.849	0.336	0.434	0.849	0.336	0.434	0.602	0.511
Bayes-Stein		0.821	1.031	0.652	0.667	0.842	0.587	0.450	0.844	0.633	0.531	0.844	0.633	0.531	0.739	0.591
BS-S		0.999	1.342	0.561	0.582	0.817	0.506	0.434	0.882	0.500	0.384	0.882	0.500	0.384	0.540	0.493
BS-SO		0.646	0.846	0.496	0.562	0.582	0.462	0.328	0.590	0.455	0.521	0.590	0.455	0.521	0.551	0.465
BS-SO		0.781	0.725	0.253	0.644	0.645	0.471	0.456	0.556	0.233	0.591	0.556	0.233	0.591	0.611	0.477
BS-LW		0.860	0.880	0.328	0.606	0.575	0.395	0.493	0.637	0.303	0.629	0.637	0.303	0.629	0.555	0.402
Data-Model																
DM-D																
DM-I																
DM-U																

Table A3: Simulated Certainty Equivalent Utility from Subset Portfolios of Variable Size

This table reports statistics on the Certainty Equivalent Utility (CEU) for an investor with mean-variance preferences 1 and risk aversion parameter  $\gamma = 2$  when implementing the portfolio algorithms presented in section 4.3 based on simulated 60-month histories of returns. All portfolios implement the budget constraint that weights sum to unity, with the Long-Short portfolios constraining weights to be between -1000% and 1000% and the Long-Only portfolios constraining weights to be non-negative. Panel A reports the average CEU across 500 simulations for the Restricted Four-Factor model and Augmented Factor model that includes an additional 10 principal components of returns as described in section 4.1. Panel B reports the simulated standard deviation of the CEU and Panel C reports the simulated frequency with which the strategy's CEU exceeds that of the 1/N portfolio.

		Panel A: Simulated Average Certainty Equivalent Utility by Subset Size									
		Restricted Factor Models					Augmented Factor Models				
		Long-Short Portfolios		Long-Only Portfolios		Long-Short Portfolios		Long-Only Portfolios		Long-Only Portfolios	
$T = 60, \gamma = 2$	Subset Size	EUFund	EUStock	EUFund	EUStock	EUFund	EUStock	EUFund	EUStock	EUFund	EUStock
1/N		5.11%	5.29%	5.11%	7.13%	5.29%	7.13%	4.92%	5.49%	4.92%	7.08%
$\tilde{N} = 2$		6.61%	5.57%	5.48%	7.90%	5.49%	8.07%	6.41%	5.72%	5.29%	7.71%
$\tilde{N} = 5$		10.51%	6.01%*	5.90%	8.41%	5.63%*	10.24%	9.35%*	6.12%*	5.76%	8.20%
$\tilde{N} = 10$		15.73%*	6.18%*	6.21%	8.62%	5.67%*	13.07%*	11.31%*	6.10%*	6.10%	8.42%
$\tilde{N} = 25$		21.56%*	-1.62%	6.61%	8.71%*	5.59%	11.08%*	-1.54%	-3.80%	6.54%	8.53%*
$\tilde{N} = 50$		-30.28%	-186.52%	6.89%*	8.62%*	5.38%	-210.84%	-98.82%	-216.32%	6.88%*	8.46%*
$\tilde{N} = 100$		-500.78%	-2430.77%	7.12%*	8.35%	4.92%	-3499.61%	-632.09%	-3613.87%	7.24%*	8.22%
		Panel B: Simulated Standard Deviation of Certainty Equivalent Utility by Subset Size									
		Restricted Factor Models					Augmented Factor Models				
		Long-Short Portfolios		Long-Only Portfolios		Long-Short Portfolios		Long-Only Portfolios		Long-Only Portfolios	
$T = 60, \gamma = 2$	Subset Size	EUFund	EUStock	EUFund	EUStock	EUFund	EUStock	EUFund	EUStock	EUFund	EUStock
$\tilde{N} = 2$		0.51%	0.28%	0.12%	0.15%	0.08%	0.24%	0.79%	0.12%	0.18%	0.14%
$\tilde{N} = 5$		1.86%	1.09%	0.26%	0.33%	0.19%	1.00%	3.59%	0.55%	0.34%	0.30%
$\tilde{N} = 10$		4.23%	2.54%	0.39%	0.45%	0.28%	2.57%	9.19%	1.59%	0.46%	0.43%
$\tilde{N} = 25$		13.87%	8.83%	0.56%	0.63%	0.43%	11.53%	30.41%	9.40%	0.66%	0.64%
$\tilde{N} = 50$		44.38%	59.99%	0.69%	0.81%	0.59%	98.28%	87.32%	88.51%	0.85%	0.85%
$\tilde{N} = 100$		193.82%	646.43%	0.82%	1.04%	0.81%	979.52%	297.20%	1003.16%	1.08%	1.15%
		Panel C: Frequency of Certainty Equivalent Utility > 1/N Certainty Equivalent Utility by Subset Size									
		Restricted Factor Models					Augmented Factor Models				
		Long-Short Portfolios		Long-Only Portfolios		Long-Short Portfolios		Long-Only Portfolios		Long-Only Portfolios	
$T = 60, \gamma = 2$	Subset Size	EUFund	EUStock	EUFund	EUStock	EUFund	EUStock	EUFund	EUStock	EUFund	EUStock
$\tilde{N} = 2$		100%*	100%*	100%*	100%*	100%*	100%*	96%*	98%*	98%*	98%*
$\tilde{N} = 5$		99%*	100%*	99%*	100%*	96%*	100%*	91%*	88%*	99%*	100%*
$\tilde{N} = 10$		98%	100%	99%*	100%*	90%	99%	83%	71%	99%*	100%
$\tilde{N} = 25$		90%	96%	99%	99%	78%	70%	47%	7%	99%	99%
$\tilde{N} = 50$		19%	0%	99%	95%	60%	0%	6%	0%	98%	94%
$\tilde{N} = 100$		0%	0%	98%	90%	34%	0%	0%	0%	98%	83%

Table A4: Simulated Expectations, Volatilities, and Sharpe Ratios from Subset Portfolios of Variable Size

This table reports statistics on the expected return (Panel A), volatility (Panel B), and Sharpe Ratio (Panel C) realized by an investor with mean-variance preferences 1 and risk aversion parameter  $\gamma = 2$  when implementing the portfolio algorithms presented in section 4.3 based on simulated 60-month histories of returns. The reported performance averages across 400 simulations for the Restricted Four-Factor model and Augmented Factor model that includes an additional 10 principal components of returns as described in section 4.1. All portfolios implement the budget constraint that weights sum to unity, with the Long-Short portfolios constraining weights to be between -1000% and 1000% and the Long-Only portfolios constraining weights to be non-negative.

Panel A: Simulated Average Expected Return by Subset Size												
T = 60, $\gamma = 2$	Restricted Factor Models			Long-Only Portfolios			Long-Short Portfolios			Augmented Factor Models		
	Subset Size	EUFund	USStock	EUFund	EUStock	USStock	EUFund	EUStock	USStock	EUFund	EUStock	USStock
$\tilde{N}=2$	7.05%	8.70%	8.05%	7.05%	8.70%	8.05%	6.90%	8.69%	8.28%	6.90%	8.69%	8.28%
$\tilde{N}=5$	8.49%	9.77%	7.97%	7.39%	9.38%	7.98%	8.66%	9.63%	8.16%	7.28%	9.27%	8.17%
$\tilde{N}=10$	13.03%	12.55%	8.47%	7.81%	9.95%	8.20%	15.00%	12.34%	8.68%	7.83%	9.85%	8.40%
$\tilde{N}=25$	21.21%	17.76%	9.93%	8.14%	10.27%	8.51%	26.99%	17.66%	10.22%	8.27%	10.22%	8.75%
$\tilde{N}=50$	50.48%	40.37%	17.20%	8.59%	10.55%	8.95%	66.36%	41.34%	17.83%	8.87%	10.58%	9.25%
$\tilde{N}=100$	109.32%*	127.49%*	47.50%*	8.94%*	10.65%*	9.24%*	134.59%*	131.07%*	49.27%*	9.35%*	10.77%*	9.62%*
	242.24%*	419.41%*	158.49%*	9.28%*	10.67%*	9.49%*	274.48%*	427.88%*	164.27%*	9.86%*	10.90%*	9.94%*

Panel B: Simulated Average Volatility by Subset Size												
T = 60, $\gamma = 2$	Restricted Factor Models			Long-Only Portfolios			Long-Short Portfolios			Augmented Factor Models		
	Subset Size	EUFund	USStock	EUFund	EUStock	USStock	EUFund	EUStock	USStock	EUFund	EUStock	USStock
$\tilde{N}=2$	13.65%	12.03%	15.48%	13.81%	12.16%	15.77%	14.91%	12.47%	15.57%	14.10%	12.49%	15.85%
$\tilde{N}=5$	15.49%	12.77%	15.41%	13.81%	12.37%	15.99%	23.12%	14.37%	15.70%	14.37%	12.84%	16.10%
$\tilde{N}=10$	22.68%	17.03%	18.58%	13.86%	12.80%	16.79%	38.34%	21.04%	19.47%	14.68%	13.36%	16.94%
$\tilde{N}=25$	52.52%	42.58%	42.09%	14.03%	13.50%	18.24%	80.34%	54.04%	45.10%	15.18%	14.23%	18.48%
$\tilde{N}=50$	116.38%	150.79%	150.98%	14.26%	14.19%	19.57%	149.88%	182.77%	160.72%	15.60%	15.09%	19.92%
$\tilde{N}=100$	269.77%	530.17%	577.38%	14.63%	15.18%	21.30%	296.72%	621.68%	609.81%	16.04%	16.28%	21.74%

Panel C: Simulated Average Sharpe Ratio by Subset Size												
T = 60, $\gamma = 2$	Restricted Factor Models			Long-Only Portfolios			Long-Short Portfolios			Augmented Factor Models		
	Subset Size	EUFund	USStock	EUFund	EUStock	USStock	EUFund	EUStock	USStock	EUFund	EUStock	USStock
$\tilde{N}=2$	0.625	0.814	0.516	0.535	0.771	0.506	0.584	0.773	0.524	0.517	0.743	0.516
$\tilde{N}=5$	0.874	1.002*	0.559**	0.566	0.806**	0.514**	0.669	0.866**	0.562**	0.546	0.769**	0.523**
$\tilde{N}=10$	0.971*	1.069**	0.547*	0.588	0.805*	0.509*	0.728	0.853*	0.535*	0.565	0.767*	0.518*
$\tilde{N}=25$	0.975**	0.952	0.403	0.615	0.785	0.493	0.845	0.774	0.390	0.587	0.748	0.503
$\tilde{N}=50$	0.941	0.840	0.308	0.630*	0.755	0.475	0.911*	0.720	0.300	0.602*	0.719	0.485
$\tilde{N}=100$	0.894	0.785	0.270	0.639**	0.707	0.448	0.933**	0.690	0.265	0.618**	0.677	0.460

Table A5: Simulated Certainty Equivalent Utility from Subset Portfolios of Variable Subset and Sample Size

This table reports statistics on the Certainty Equivalent Utility (CEU) for an investor with mean-variance preferences  $\lambda$  and risk aversion parameter  $\gamma = 2$  when implementing the portfolio algorithms presented in section 4.3 based on simulated histories of returns with variable lengths. All portfolios implement the budget constraint that weights sum to unity, with the Long-Short portfolios constraining weights to be between -1000% and 1000% and the Long-Only portfolios constraining weights to be non-negative. Panel A reports the average CEU across 500 simulations for the Restricted Four-Factor model and Augmented Factor model that includes an additional 10 principal components of returns as described in section 4.1. Panel B reports the simulated standard deviation of the CEU and Panel C reports the simulated frequency with which the strategy's CEU exceeds that of the  $1/N$  portfolio.

Panel A: Simulated Average Certainty Equivalent Utility by Subset and Sample Size												
Restricted Factor Models												
$T = 60, \gamma = 2$	Long-Short Portfolios			Long-Only Portfolios			Long-Short Portfolios			Long-Only Portfolios		
	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
$\hat{N} = 10$	11.97%	14.13%	3.74%	5.90%	8.14%	5.49%	-10.60%	9.30%	3.07%	5.72%	7.97%	5.66%
$T = 30$	15.73%	14.70%	6.18%	6.21%	8.62%	5.67%	11.31%	13.07%	6.10%	6.10%	8.42%	5.85%
$T = 60$	17.01%	14.80%	6.81%	6.61%	9.15%	5.85%	17.51%	13.82%	6.92%	6.47%	8.91%	6.03%
$T = 120$	17.35%	14.71%	7.08%	6.99%	9.64%	6.02%	19.96%	13.99%	7.15%	6.78%	9.38%	6.18%
$\hat{N} = 25$												
$T = 30$	-55.65%	-63.63%	-115.76%	6.19%	8.04%	5.31%	-213.57%	-181.88%	-138.78%	6.06%	7.90%	5.50%
$T = 60$	21.56%	21.50%	-1.62%	6.61%	8.71%	5.59%	-1.54%	11.08%	-3.80%	6.54%	8.53%	5.80%
$T = 120$	30.14%	24.61%	7.05%	7.17%	9.46%	5.90%	29.64%	21.70%	6.82%	7.08%	9.25%	6.10%
$T = 240$	31.55%	24.48%	8.93%	7.67%	10.15%	6.17%	38.94%	23.27%	8.84%	7.49%	9.95%	6.35%
$\hat{N} = 50$												
$T = 30$	-1.697%	-5.885%	-5.920%	6.40%	7.78%	4.99%	-2.055%	-9.876%	-6.804%	6.34%	7.70%	5.18%
$T = 60$	-30.28%	-104.89%	-186.52%	6.89%	8.62%	5.38%	-98.82%	-210.84%	-216.32%	6.88%	8.46%	5.61%
$T = 120$	29.34%	29.10%	-6.07%	7.55%	9.56%	5.80%	15.75%	18.26%	-8.27%	7.57%	9.38%	6.03%
$T = 240$	43.33%	35.51%	8.78%	8.12%	10.41%	6.18%	53.38%	33.66%	8.16%	8.05%	10.29%	6.39%

Panel B: Simulated Standard Deviation of Certainty Equivalent Utility by Subset and Sample Size												
Restricted Factor Models												
$T = 60, \gamma = 2$	Long-Short Portfolios			Long-Only Portfolios			Long-Short Portfolios			Long-Only Portfolios		
	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
$\hat{N} = 10$	8.19%	4.50%	3.46%	0.43%	0.56%	0.33%	22.83%	5.33%	3.75%	0.58%	0.53%	0.31%
$T = 30$	4.23%	2.54%	1.54%	0.39%	0.45%	0.28%	9.19%	2.57%	1.59%	0.46%	0.43%	0.26%
$T = 60$	2.40%	1.42%	0.80%	0.30%	0.33%	0.21%	4.37%	1.56%	0.84%	0.41%	0.37%	0.22%
$T = 120$	1.66%	1.00%	0.48%	0.23%	0.27%	0.17%	2.65%	1.01%	0.50%	0.34%	0.28%	0.17%
$\hat{N} = 25$												
$T = 30$	73.68%	58.15%	76.10%	0.63%	0.81%	0.51%	167.48%	107.91%	86.32%	0.81%	0.81%	0.49%
$T = 60$	13.87%	8.83%	8.56%	0.56%	0.63%	0.43%	30.41%	11.53%	9.40%	0.66%	0.64%	0.39%
$T = 120$	5.76%	3.51%	2.46%	0.43%	0.48%	0.32%	11.02%	4.37%	2.58%	0.59%	0.55%	0.31%
$T = 240$	3.34%	2.25%	1.22%	0.33%	0.37%	0.25%	5.11%	2.52%	1.29%	0.50%	0.40%	0.25%
$\hat{N} = 50$												
$T = 30$	1.315%	2.709%	3.244%	0.79%	1.06%	0.69%	1.446%	4.417%	3.669%	1.00%	1.06%	0.68%
$T = 60$	44.38%	59.99%	77.65%	0.69%	0.81%	0.59%	87.32%	98.28%	88.51%	0.85%	0.85%	0.51%
$T = 120$	11.83%	7.44%	7.26%	0.53%	0.60%	0.43%	28.67%	12.65%	7.74%	0.77%	0.73%	0.41%
$T = 240$	4.68%	3.64%	2.35%	0.40%	0.47%	0.34%	8.30%	4.82%	2.57%	0.67%	0.50%	0.33%

Panel C: Frequency of Certainty Equivalent Utility $> 1/N$ Certainty Equivalent Utility by Subset and Sample Size												
Restricted Factor Models												
$T = 60, \gamma = 2$	Long-Short Portfolios			Long-Only Portfolios			Long-Short Portfolios			Long-Only Portfolios		
	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
$\hat{N} = 10$	84%	96%	37%	97%	96%	73%	29%	72%	27%	90%	94%	73%
$T = 30$	98%	100%	75%	99%	100%	90%	83%	99%	71%	99%	100%	91%
$T = 60$	100%	100%	95%	100%	100%	100%	99%	100%	96%	100%	100%	100%
$T = 120$	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
$\hat{N} = 25$												
$T = 30$	17%	1%	0%	97%	87%	56%	2%	0%	0%	89%	85%	57%
$T = 60$	90%	96%	13%	99%	99%	78%	47%	70%	7%	99%	80%	80%
$T = 120$	100%	100%	79%	100%	100%	97%	98%	99%	77%	100%	99%	100%
$T = 240$	100%	100%	100%	100%	100%	100%	100%	100%	99%	100%	100%	100%
$\hat{N} = 50$												
$T = 30$	0%	0%	0%	95%	76%	38%	0%	0%	0%	88%	75%	37%
$T = 60$	19%	0%	0%	99%	95%	60%	6%	0%	0%	98%	94%	63%
$T = 120$	97%	99%	1%	100%	100%	88%	69%	86%	1%	100%	100%	90%
$T = 240$	100%	100%	92%	100%	100%	99%	100%	100%	87%	100%	100%	99%

**Table A6: Simulated Expectation, Volatility, & Sharpe Ratio of Subset Portfolios with Variable Subset & Sample Size**

This table reports statistics on the expected return (Panel A), volatility (Panel B), and Sharpe Ratio (Panel C) realized by an investor with mean-variance preferences 1 and risk aversion parameter  $\gamma = 2$  when implementing the portfolio algorithms presented in section 4.3 based on simulated histories of returns with variable lengths. The reported performance averages across 400 simulations for the Restricted Four-Factor model and Augmented Factor model that includes an additional 10 principal components of returns as described in section 4.1. All portfolios implement the budget constraint that weights sum to unity, with the Long-Short portfolios constraining weights to be between -1000% and 1000% and the Long-Only portfolios constraining weights to be non-negative.

		Panel A: Simulated Average Expected Return by Subset and Sample Size														
		Restricted Factor Models					Augmented Factor Models									
		Long-Short Portfolios					Long-Only Portfolios									
$T = 60, \gamma = 2$	$\hat{N} = 10$	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
	T=30	23.37%	20.02%	10.82%	7.88%	9.85%	8.50%	28.11%	19.99%	11.19%	7.88%	9.83%	8.74%	7.88%	9.83%	8.74%
	T=60	21.21%	17.76%	9.93%	8.14%	10.27%	8.51%	26.99%	17.66%	10.22%	8.27%	10.22%	8.75%	8.27%	10.22%	8.75%
	T=120	20.70%	17.09%	9.61%	8.49%	10.76%	8.51%	27.08%	16.92%	9.91%	8.62%	10.66%	8.75%	8.62%	10.66%	8.75%
	T=240	20.30%	16.72%	9.59%	8.84%	11.22%	8.58%	26.58%	16.50%	9.74%	8.90%	11.11%	8.77%	8.90%	11.11%	8.77%
$\hat{N} = 25$																
	T=30	78.27%	79.45%	31.15%	8.24%	9.97%	8.89%	89.68%	80.60%	32.50%	8.38%	10.04%	9.19%	8.38%	10.04%	9.19%
	T=60	50.48%	40.37%	17.20%	8.59%	10.55%	8.95%	66.36%	41.34%	17.83%	8.87%	10.58%	9.25%	8.87%	10.58%	9.25%
	T=120	44.43%	32.59%	14.25%	9.09%	11.25%	9.00%	64.26%	33.23%	14.79%	9.39%	11.22%	9.30%	9.39%	11.22%	9.30%
	T=240	40.78%	29.76%	13.47%	9.55%	11.90%	9.13%	59.74%	30.19%	13.64%	9.79%	11.87%	9.37%	9.79%	11.87%	9.37%
$\hat{N} = 50$																
	T=30	295.60%	502.73%	175.31%	8.53%	9.94%	9.14%	266.36%	493.70%	182.22%	8.80%	10.13%	9.50%	8.80%	10.13%	9.50%
	T=60	109.32%	127.49%	47.50%	8.94%	10.65%	9.24%	134.59%	131.07%	49.27%	9.35%	10.77%	9.62%	9.35%	10.77%	9.62%
	T=120	90.91%	67.01%	25.39%	9.53%	11.51%	9.35%	130.40%	70.61%	26.52%	10.04%	11.56%	9.70%	10.04%	11.56%	9.70%
	T=240	71.61%	51.79%	20.47%	10.05%	12.32%	9.51%	107.59%	54.28%	20.72%	10.51%	12.39%	9.82%	10.51%	12.39%	9.82%
		Panel B: Simulated Average Volatility by Subset and Sample Size														
		Restricted Factor Models					Augmented Factor Models									
		Long-Short Portfolios					Long-Only Portfolios									
$T = 60, \gamma = 2$	$\hat{N} = 10$	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
	T=30	32.25%	23.51%	25.25%	14.05%	13.03%	17.27%	59.76%	31.86%	27.20%	14.64%	13.58%	17.46%	14.64%	13.58%	17.46%
	T=60	22.68%	17.03%	18.58%	13.86%	12.80%	16.79%	38.34%	21.04%	19.47%	14.68%	13.36%	16.94%	14.68%	13.36%	16.94%
	T=120	18.86%	14.90%	16.29%	13.70%	12.66%	16.26%	30.24%	17.34%	16.87%	14.62%	13.22%	16.43%	14.62%	13.22%	16.43%
	T=240	17.01%	14.07%	15.61%	13.59%	12.57%	15.95%	25.46%	15.75%	15.84%	14.54%	13.10%	16.05%	14.54%	13.10%	16.05%
$\hat{N} = 25$																
	T=30	111.27%	116.98%	117.36%	14.27%	13.83%	18.82%	167.45%	158.53%	127.00%	15.15%	14.55%	19.10%	15.15%	14.55%	19.10%
	T=60	52.52%	42.58%	42.09%	14.03%	13.50%	18.24%	80.34%	54.04%	45.10%	15.18%	14.23%	18.48%	15.18%	14.23%	18.48%
	T=120	37.30%	27.84%	26.11%	13.83%	13.32%	17.54%	57.85%	33.38%	27.59%	15.14%	13.99%	17.80%	15.14%	13.99%	17.80%
	T=240	30.16%	22.79%	20.84%	13.68%	13.21%	17.14%	45.22%	26.07%	21.49%	15.11%	13.81%	17.30%	15.11%	13.81%	17.30%
$\hat{N} = 50$																
	T=30	425.89%	782.60%	758.65%	14.52%	14.63%	20.28%	460.10%	997.14%	813.08%	15.59%	15.49%	20.66%	15.59%	15.49%	20.66%
	T=60	116.38%	150.79%	150.98%	14.26%	14.19%	19.57%	149.88%	182.77%	160.72%	15.60%	15.09%	19.92%	15.60%	15.09%	19.92%
	T=120	77.81%	60.97%	55.38%	14.06%	13.92%	18.76%	105.73%	71.52%	58.31%	15.62%	14.70%	19.07%	15.62%	14.70%	19.07%
	T=240	52.86%	40.11%	33.80%	13.86%	13.77%	18.19%	73.12%	45.09%	35.06%	15.61%	14.44%	18.45%	15.61%	14.44%	18.45%
		Panel C: Simulated Average Sharpe Ratio by Subset and Sample Size														
		Restricted Factor Models					Augmented Factor Models									
		Long-Short Portfolios					Long-Only Portfolios									
$T = 60, \gamma = 2$	$\hat{N} = 10$	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
	T=30	0.768	0.877	0.436	0.563	0.760	0.495	0.415	0.641	0.415	0.540	0.728	0.503	0.540	0.728	0.503
	T=60	0.971	1.069	0.547	0.588	0.805	0.509	0.728	0.853	0.535	0.565	0.767	0.518	0.565	0.767	0.518
	T=120	1.124	1.169	0.603	0.621	0.853	0.621	0.918	0.990	0.597	0.809	0.534	0.534	0.809	0.534	0.534
	T=240	1.209	1.203	0.624	0.652	0.895	0.539	1.056	1.056	0.623	0.613	0.849	0.548	0.613	0.849	0.548
$\hat{N} = 25$																
	T=30	0.721	0.682	0.260	0.581	0.727	0.475	0.555	0.514	0.250	0.557	0.697	0.484	0.557	0.697	0.484
	T=60	0.975	0.952	0.403	0.615	0.785	0.493	0.845	0.774	0.390	0.587	0.748	0.503	0.587	0.748	0.503
	T=120	1.201	1.180	0.546	0.659	0.848	0.515	1.128	1.007	0.536	0.806	0.622	0.524	0.806	0.622	0.524
	T=240	1.357	1.314	0.655	0.700	0.904	0.535	1.328	1.167	0.641	0.649	0.862	0.544	0.649	0.862	0.544
$\hat{N} = 50$																
	T=30	0.702	0.641	0.226	0.592	0.686	0.454	0.596	0.498	0.218	0.569	0.662	0.463	0.569	0.662	0.463
	T=60	0.941	0.840	0.308	0.630	0.755	0.475	0.911	0.720	0.300	0.602	0.719	0.485	0.602	0.719	0.485
	T=120	1.167	1.097	0.452	0.680	0.831	0.501	1.242	0.990	0.450	0.645	0.792	0.511	0.645	0.792	0.511
	T=240	1.354	1.293	0.605	0.727	0.898	0.525	1.474	1.208	0.590	0.676	0.861	0.534	0.676	0.861	0.534

Table A7: Simulated Certainty Equivalent Utility from Subset Portfolios with Variable Risk Aversion ( $\gamma$ )

This table reports statistics on the Certainty Equivalent Utility (CEU) for an investor with mean-variance preferences 1 after observing  $T = 60$  months of returns and implementing subset portfolios of size  $\hat{N} = 10$  under variable values for the risk aversion parameter  $\gamma = 2$ . All portfolios implement the budget constraint that weights sum to unity, with the Long-Short portfolios constraining weights to be between -1000% and 1000% and the Long-Only portfolios constraining weights to be non-negative. Panel A reports the average CEU across 500 simulations for the Restricted Four-Factor model and Augmented Factor model that includes an additional 10 principal components of returns as described in section 4.1. Panel B reports the simulated standard deviation of the CEU and Panel C reports the simulated frequency with which the strategy's CEU exceeds that of the 1/N portfolio.

Panel A: Simulated Average Certainty Equivalent Utility by Risk Aversion ( $\gamma$ )												
Restricted Factor Models												
Long-Short Portfolios			Long-Only Portfolios			Long-Short Portfolios			Long-Only Portfolios			
1/N	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
Mean-Variance	5.11%	5.11%	5.29%	5.11%	5.29%	5.29%	4.92%	7.08%	5.49%	4.92%	7.08%	5.49%
$\gamma = 0.1$	-120,205%	-56,387%	-62,011%	6.18%	8.20%	5.29%	-423,657%	-109,817%	-75,623%	6.18%	8.15%	5.55%
$\gamma = 1$	18.70%	19.15%	4.15%	6.20%	8.55%	5.59%	-9.09%	13.68%	3.22%	6.15%	8.40%	5.80%
$\gamma = 2$	15.78%	14.70%	6.18%	6.21%	8.62%	5.67%	11.41%	13.07%	6.10%	6.10%	8.42%	5.85%
$\gamma = 5$	10.39%	10.62%	5.89%	6.15%	8.50%	5.63%	10.47%	10.01%	5.99%	5.93%	8.24%	5.78%
$\gamma = 10$	7.98%	9.03%	5.54%	6.01%	8.22%	5.49%	7.95%	8.54%	5.64%	5.68%	7.91%	5.62%
$\gamma = 25$	6.40%	8.02%	5.27%	5.72%	7.79%	5.29%	5.91%	7.55%	5.36%	5.28%	7.41%	5.41%
$\gamma = 50$	5.84%	7.67%	5.17%	5.53%	7.57%	5.21%	5.14%	7.20%	5.26%	5.04%	7.16%	5.33%

Panel B: Simulated Standard Deviation of Certainty Equivalent Utility by Risk Aversion ( $\gamma$ )												
Restricted Factor Models												
Long-Short Portfolios			Long-Only Portfolios			Long-Short Portfolios			Long-Only Portfolios			
1/N	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
Mean-Variance	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
$\gamma = 0.1$	78,560%	30,516%	45,712%	0.43%	0.62%	0.45%	247,555%	56,627%	56,848%	0.48%	0.59%	0.42%
$\gamma = 1$	10.96%	5.26%	4.67%	0.41%	0.52%	0.35%	34.00%	6.46%	5.29%	0.48%	0.50%	0.33%
$\gamma = 2$	4.26%	2.54%	1.54%	0.39%	0.45%	0.28%	9.42%	2.56%	1.59%	0.47%	0.43%	0.26%
$\gamma = 5$	1.66%	1.05%	0.55%	0.32%	0.35%	0.19%	2.68%	1.01%	0.55%	0.41%	0.34%	0.19%
$\gamma = 10$	0.87%	0.54%	0.29%	0.25%	0.27%	0.15%	1.33%	0.53%	0.29%	0.32%	0.28%	0.15%
$\gamma = 25$	0.40%	0.25%	0.14%	0.17%	0.17%	0.09%	0.64%	0.25%	0.15%	0.22%	0.19%	0.10%
$\gamma = 50$	0.27%	0.17%	0.10%	0.14%	0.13%	0.07%	0.46%	0.18%	0.11%	0.18%	0.15%	0.07%

Panel C: Frequency of Certainty Equivalent Utility > 1/N Certainty Equivalent Utility by Risk Aversion ( $\gamma$ )												
Restricted Factor Models												
Long-Short Portfolios			Long-Only Portfolios			Long-Short Portfolios			Long-Only Portfolios			
1/N	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
Mean-Variance	0%	0%	0%	99%	94%	54%	0%	0%	0%	99%	97%	59%
$\gamma = 0.1$	92%	98%	49%	99%	99%	80%	37%	87%	35%	99%	99%	82%
$\gamma = 2$	98%	100%	75%	99%	100%	90%	83%	99%	71%	99%	100%	91%
$\gamma = 5$	100%	100%	87%	100%	100%	96%	96%	100%	82%	99%	100%	95%
$\gamma = 10$	100%	100%	81%	100%	100%	93%	98%	99%	69%	99%	100%	80%
$\gamma = 25$	100%	100%	44%	100%	100%	52%	94%	97%	21%	93%	96%	24%
$\gamma = 50$	100%	100%	13%	100%	100%	13%	70%	77%	1%	76%	73%	1%

# Separately Presented Tables and Figures as They Appear in the Main Text

Table 1: Simulation Universe Cross-sectional Return Properties

This table reports the cross-sectional return properties and risk exposures for simulated asset universes. Panels A, B, and C report on the universe of US Stocks (1963-2015) from the CRSP database, European Stocks (1988-2008) from Datastream, and European Equity Mutual Funds (1988-2008) from the Lipper database, respectively. The Restricted Moments column reports the mean and volatility of returns calibrated to a restricted (zero-alpha) four-factor model where factor loadings are calculated by regressing assets' excess returns on benchmark factors, ignoring missing data, with idiosyncratic volatility equal to the residual standard deviation. The usual four benchmark factors (Mkt, SMB, HML, Mom) come from Ken French's website, with benchmark expected returns and covariances calibrated to historical sample moments. The Augmented Moment Model extracts ten latent priced factors from the security universe using principal components and calibrates security returns to a fourteen factor (four benchmarks + 10 PCAs). The column Alpha reports the mispricing in the expected return when the augmented model is fit to only the four observed benchmark factors. Appendix Table A1 reports on the expectations and factor loadings for the augmented factors.

Panel A: CRSP Stock Universe ( $N=1,063$ Assets)									
	Restricted Moments		Benchmark Factor Loadings				Augmented Moment Model		
	Mean	Vol	Mkt	SMB	HML	MOM	Alpha	Mean	Vol
Average	8.05%	38.92%	0.96	0.47	0.41	(0.08)	0.22%	8.28%	38.86%
Std Deviation	2.98%	15.48%	0.32	0.53	0.40	0.21	0.98%	3.09%	15.40%
1%-Quantile	1.35%	13.37%	0.27	(0.49)	(0.79)	(0.77)	-2.38%	1.46%	13.36%
10%-Quantile	4.42%	22.17%	0.56	(0.15)	(0.07)	(0.34)	-0.82%	4.54%	22.17%
50%-Quantile	7.96%	35.76%	0.96	0.43	0.45	(0.05)	0.20%	8.05%	35.75%
90%-Quantile	11.86%	57.93%	1.36	1.13	0.86	0.13	1.24%	12.25%	57.70%
99%-Quantile	15.55%	92.81%	1.75	1.97	1.25	0.31	3.17%	16.45%	92.43%
Panel B: European Stock Universe ( $N=1,357$ Assets)									
	Restricted Moments		Benchmark Factor Loadings				Augmented Moment Model		
	Mean	Vol	Mkt	SMB	HML	MOM	Alpha	Mean	Vol
Average	8.70%	33.58%	0.88	0.63	0.37	(0.03)	-0.02%	8.69%	33.59%
Std Deviation	4.49%	10.71%	0.35	0.52	0.50	0.23	3.02%	4.40%	10.71%
1%-Quantile	-4.62%	13.81%	0.20	(0.61)	(1.40)	(0.64)	-7.90%	-3.68%	13.80%
10%-Quantile	3.27%	22.29%	0.45	0.02	(0.20)	(0.32)	-3.22%	3.31%	22.27%
50%-Quantile	9.00%	31.71%	0.86	0.61	0.44	(0.00)	-0.21%	8.89%	31.73%
90%-Quantile	13.93%	48.46%	1.36	1.25	0.88	0.23	3.57%	13.95%	48.39%
99%-Quantile	18.40%	66.98%	1.82	2.12	1.26	0.49	9.66%	18.82%	67.01%
Panel C: European Mutual Fund Universe ( $N=878$ Assets)									
	Restricted Moments		Benchmark Factor Loadings				Augmented Moment Model		
	Mean	Vol	Mkt	SMB	HML	MOM	Alpha	Mean	Vol
Average	7.05%	17.09%	1.00	0.21	(0.05)	(0.01)	-0.14%	6.90%	17.00%
Std Deviation	1.87%	2.85%	0.16	0.34	0.24	0.09	1.83%	2.05%	2.73%
1%-Quantile	2.86%	12.41%	0.60	(0.24)	(0.72)	(0.22)	-3.77%	2.23%	12.56%
10%-Quantile	4.99%	13.99%	0.81	(0.13)	(0.30)	(0.12)	-2.34%	4.50%	14.09%
50%-Quantile	6.91%	16.63%	1.00	0.11	(0.07)	(0.01)	-0.31%	6.87%	16.52%
90%-Quantile	9.47%	20.98%	1.20	0.74	0.27	0.10	2.56%	9.36%	20.68%
99%-Quantile	12.29%	25.67%	1.32	1.19	0.52	0.25	5.00%	13.00%	25.27%

Table 2: Simulation Universe Benchmark Return Properties

This table reports the historical average, volatility, Sharpe Ratio, and Certainty Equivalent Utility for the benchmarks defining systematic factors in the US and European simulation universes. These four benchmark factors include market excess return (Mkt), small-minus-big capitalization (SMB), high-minus-low book-to-market ratio (HML), and winning-minus-losing (WML) momentum factors for the US and Europe taken from Ken French's website. The annualized Expected Return and Volatility is calculated from the sample mean and standard deviation of historical monthly returns, while the Certainty Equivalent (CE) utility is calculated based on the mean-variance utility function with no estimation error.

	Panel A: US Benchmark Factors				Panel B: Europe Benchmark Factors			
	Mkt	SMB	HML	MOM	Mkt	SMB	HML	MOM
Expected Return	6.17%	2.36%	4.24%	8.50%	7.79%	-1.00%	7.46%	12.70%
Volatility	15.93%	10.83%	10.35%	15.43%	14.21%	8.23%	7.90%	13.19%
Sharpe Ratio	0.39	0.22	0.41	0.55	0.55	(0.12)	0.94	0.96
CE Utility								
$\gamma = 1$	4.91%	1.77%	3.70%	7.30%	6.78%	-1.34%	7.14%	11.83%
$\gamma = 2$	3.64%	1.19%	3.16%	6.11%	5.77%	-1.68%	6.83%	10.96%
$\gamma = 4$	1.10%	0.01%	2.09%	3.73%	3.75%	-2.36%	6.21%	9.22%
$\gamma = 8$	-3.97%	-2.33%	-0.05%	-1.03%	-0.29%	-3.71%	4.96%	5.74%

Table 3: Simulation Universe Asset Optimized Portfolio Return Properties

This table reports the expected performance properties, including the annualized Expected Return, Volatility, and Sharpe Ratio, along with quantiles from the cross-sectional distribution of optimized weights for each of the six simulation universes. The optimized portfolios minimize variance (MinVar) and maximize Sharpe Ratio (MaxSR) allowing for arbitrary short positions (Unconstrained) and subject to no-shorting restrictions (Non-Negative), along with the 1/N portfolio for reference. Data for the universe of US Stocks (1963-2015) comes from the CRSP database, European Stocks (1988-2008) from Datastream, and European Equity Mutual Funds (1988-2008) from the Lipper database. The Restricted model of returns fits a zero-alpha four-factor model by regressing assets' excess returns on benchmark factors, with idiosyncratic volatility equal to the residual standard deviation. The four benchmark factors (Mkt, SMB, HML, Mom) come from Ken French's website, with factor expected returns and covariances equaling historical sample moments. The Augmented model extracts an additional ten latent priced factors from the security universe using principal components and calibrates security returns to a fourteen factor (four benchmarks + 10 PCAs).

Panel A.1: CRSP Stocks (Restricted)						Panel A.2: CRSP Stocks (Augmented)				
	1/N	Unconstrained		Non-Negative		1/N	Unconstrained		Non-Negative	
		MinVar	MaxSR	MinVar	MaxSR		MinVar	MaxSR	MinVar	MaxSR
Expectation	8.05%	0.58%	10.81%	2.74%	7.48%	8.28%	0.58%	14.37%	2.98%	8.15%
Volatility	16.63%	2.54%	10.92%	5.53%	9.40%	16.70%	2.82%	14.10%	6.06%	10.17%
Sharpe Ratio	0.484	0.230	0.990	0.495	0.796	0.496	0.204	1.019	0.491	0.801
Weight Distribution										
5%-Quantile	0.09%	-0.29%	-0.47%	0.00%	0.00%	0.09%	-0.33%	-0.75%	0.00%	0.00%
25%-Quantile	0.09%	-0.09%	-0.11%	0.00%	0.00%	0.09%	-0.10%	-0.24%	0.00%	0.00%
50%-Quantile	0.09%	0.01%	0.07%	0.00%	0.00%	0.09%	0.02%	0.09%	0.00%	0.00%
75%-Quantile	0.09%	0.16%	0.25%	0.00%	0.00%	0.09%	0.18%	0.37%	0.00%	0.00%
95%-Quantile	0.09%	0.65%	0.77%	0.00%	0.42%	0.09%	0.67%	0.99%	0.02%	0.56%
Panel B.1: European Stocks (Restricted)						Panel B.2: European Stocks (Augmented)				
	1/N	Unconstrained		Non-Negative		1/N	Unconstrained		Non-Negative	
		MinVar	MaxSR	MinVar	MaxSR		MinVar	MaxSR	MinVar	MaxSR
Expectation	8.70%	0.70%	16.01%	3.30%	10.88%	8.69%	0.54%	43.42%	3.17%	14.32%
Volatility	12.52%	2.04%	9.71%	4.01%	7.83%	12.70%	2.79%	24.94%	6.24%	11.19%
Sharpe Ratio	0.695	0.345	1.648	0.822	1.388	0.684	0.195	1.741	0.508	1.280
Weight Distribution										
5%-Quantile	0.07%	-0.21%	-0.27%	0.00%	0.00%	0.07%	-0.40%	-1.24%	0.00%	0.00%
25%-Quantile	0.07%	-0.07%	-0.07%	0.00%	0.00%	0.07%	-0.14%	-0.41%	0.00%	0.00%
50%-Quantile	0.07%	0.01%	0.03%	0.00%	0.00%	0.07%	0.03%	0.08%	0.00%	0.00%
75%-Quantile	0.07%	0.13%	0.20%	0.00%	0.00%	0.07%	0.18%	0.57%	0.00%	0.00%
95%-Quantile	0.07%	0.45%	0.55%	0.00%	0.52%	0.07%	0.52%	1.54%	0.02%	0.20%
Panel C.1: European Mutual Funds (Restricted)						Panel C.2: European Mutual Funds (Augmented)				
	1/N	Unconstrained		Non-Negative		1/N	Unconstrained		Non-Negative	
		MinVar	MaxSR	MinVar	MaxSR		MinVar	MaxSR	MinVar	MaxSR
Expectation	7.05%	0.49%	24.39%	4.48%	10.33%	6.90%	0.17%	179.40%	2.99%	13.27%
Volatility	13.94%	2.09%	14.85%	7.18%	11.01%	14.07%	2.83%	92.75%	8.11%	14.79%
Sharpe Ratio	0.506	0.232	1.643	0.624	0.938	0.490	0.059	1.934	0.369	0.897
Weight Distribution										
5%-Quantile	0.11%	-1.26%	-2.16%	0.00%	0.00%	0.11%	-2.95%	-28.17%	0.00%	0.00%
25%-Quantile	0.11%	-0.49%	-0.90%	0.00%	0.00%	0.11%	-1.02%	-10.11%	0.00%	0.00%
50%-Quantile	0.11%	0.02%	0.01%	0.00%	0.00%	0.11%	0.12%	0.11%	0.00%	0.00%
75%-Quantile	0.11%	0.72%	1.03%	0.00%	0.00%	0.11%	1.15%	9.82%	0.00%	0.00%
95%-Quantile	0.11%	1.63%	2.77%	0.00%	0.00%	0.11%	3.37%	28.26%	0.00%	0.00%

Table 4: Simulated Certainty Equivalent Utility from Portfolio Strategies

This table reports statistics on the Certainty Equivalent Utility (CEU) for an investor with mean-variance preferences 1 and risk aversion parameter  $\gamma = 2$  when implementing the portfolio algorithms presented in section 4.3 based on simulated 60-month histories of returns. All portfolios implement the budget constraint that weights sum to unity, with the Long-Short portfolios constraining weights to be between -1000% and 1000% and the Long-Only portfolios constraining weights to be non-negative. Panel A reports the average CEU across 500 simulations for the Restricted Four-Factor model and Augmented Factor model that includes an additional 10 principal components of returns as described in section 4.1. Panel B reports the simulated standard deviation of the CEU and Panel C reports the simulated frequency with which the strategy's CEU exceeds that of the 1/N portfolio.

		Panel A: Simulated Average Certainty Equivalent Utility From Portfolio Strategy												
		Restricted Factor Models					Augmented Factor Models							
		Long-Short Portfolios		Long-Only Portfolios			Long-Short Portfolios		Long-Only Portfolios					
$T = 60, \gamma = 2$		EUFund	EUStock	EUFund	EUStock	USStock	EUFund	EUStock	EUFund	EUStock	USStock	EUFund	EUStock	USStock
1/N		5.11%	7.13%	5.11%	7.13%	5.29%*	4.92%	7.08%	4.92%	7.08%	5.49%*	4.92%	7.08%*	5.49%*
Subset		14.80%**	14.77%**	6.23%**	6.22%	8.64%**	9.52%**	12.70%**	9.52%**	12.70%**	6.18%**	6.09%	8.39%**	5.86%**
GMV		3.55%	5.12%	3.36%	4.61%	3.89%	1.81%	4.66%	3.41%	3.10%	3.41%	3.10%	5.12%	3.93%
GMV-SO		5.27%	6.82%	4.54%	5.35%	4.71%	3.37%	6.09%	4.54%	4.35%	4.54%	4.35%	6.24%	4.79%
GMV-LW		3.25%	4.96%	2.54%	4.09%	2.54%	1.48%	4.35%	2.45%	2.63%	2.45%	2.63%	2.93%	2.68%
Bayes-Stein														
BS-S		14.47%*	-69.65%	-131.36%	4.66%	4.11%	10.70%**	-76.62%	-132.00%	3.15%	5.32%	3.15%	5.32%	4.14%
BS-SO		5.80%	9.24%*	5.40%*	5.43%	5.01%	3.92%	8.44%*	5.37%	4.44%	6.75%	4.44%	6.75%	5.07%
BS-LW		4.41%	7.55%	3.43%	4.17%	2.79%	2.78%	7.09%	3.31%	2.70%	3.49%	2.70%	3.49%	2.91%
Data-Model														
DM-D		6.98%	9.11%	4.53%	5.74%	4.95%	-4.10%	7.14%	4.04%	5.30%	5.93%	5.30%	4.80%	4.80%
DM-I		-26.72%	1.91%	-10.77%	6.64%*	4.83%	-327.04%	-41.18%	-14.31%	6.32%*	6.72%	6.32%*	4.82%	4.82%
DM-U		-1,065%	-3,459%	-1,745%	7.28%**	3.59%	-5,894%	-8,040%	-2,221%	7.75%**	7.02%	7.75%**	7.02%	3.73%
		Panel B: Simulated Standard Deviation of Certainty Equivalent Utility												
		Restricted Factor Models					Augmented Factor Models							
		Long-Short Portfolios		Long-Only Portfolios			Long-Short Portfolios		Long-Only Portfolios					
$T = 60, \gamma = 2$		EUFund	EUStock	EUFund	EUStock	USStock	EUFund	EUStock	EUFund	EUStock	USStock	EUFund	EUStock	USStock
1/N		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Subset		4.03%	2.59%	1.48%	0.40%	0.29%	8.99%	2.61%	1.44%	0.43%	0.42%	0.43%	0.25%	0.25%
GMV		0.45%	0.39%	0.25%	0.43%	0.44%	0.80%	0.57%	0.27%	0.57%	0.46%	0.57%	0.46%	0.46%
GMV-SO		0.70%	0.41%	0.30%	0.51%	0.37%	1.10%	0.59%	0.36%	0.53%	0.66%	0.53%	0.37%	0.37%
GMV-LW		0.47%	0.35%	0.23%	0.53%	0.20%	0.82%	0.58%	0.24%	0.57%	0.45%	0.57%	0.19%	0.19%
Bayes-Stein														
BS-S		4.20%	88.45%	137.50%	0.43%	0.47%	3.61%	91.17%	130.47%	0.56%	0.83%	0.56%	0.49%	0.49%
BS-SO		0.74%	1.27%	0.70%	0.51%	0.38%	1.12%	1.66%	0.70%	0.54%	0.61%	0.54%	0.43%	0.43%
BS-LW		0.65%	1.44%	0.73%	0.51%	0.28%	1.07%	1.68%	0.64%	0.57%	0.64%	0.57%	0.27%	0.27%
Data-Model														
DM-D		4.78%	3.58%	2.17%	1.15%	0.86%	12.26%	3.85%	2.67%	1.44%	1.91%	1.44%	0.91%	0.91%
DM-I		74.68%	21.95%	16.44%	1.02%	0.84%	391.63%	62.80%	16.46%	1.22%	2.07%	1.22%	0.91%	0.91%
DM-U		935.60%	1661.28%	834.40%	1.30%	1.37%	5486.10%	4236.92%	1260.57%	1.93%	2.45%	1.93%	1.10%	1.10%
		Panel C: Frequency of Certainty Equivalent Utility > 1/N Certainty Equivalent Utility												
		Restricted Factor Models					Augmented Factor Models							
		Long-Short Portfolios		Long-Only Portfolios			Long-Short Portfolios		Long-Only Portfolios					
$T = 60, \gamma = 2$		EUFund	EUStock	EUFund	EUStock	USStock	EUFund	EUStock	EUFund	EUStock	USStock	EUFund	EUStock	USStock
Subset		96%*	99%**	75%**	99%**	90%**	84%*	99%**	74%**	100%**	100%**	100%**	93%**	93%**
GMV		0%	0%	0%	13%	0%	0%	0%	0%	0%	1%	0%	1%	1%
GMV-SO		60%	20%	0%	68%	18%	8%	3%	0%	16%	8%	16%	1%	1%
GMV-LW		0%	0%	0%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Bayes-Stein														
BS-S		99%**	4%	0%	16%	0%	95%**	14%	0%	0%	1%	0%	1%	1%
BS-SO		84%	95%*	53%*	71%	21%	21%	84%*	44%*	18%	34%	18%	16%	16%
BS-LW		16%	58%	0%	4%	0%	3%	45%	0%	0%	0%	0%	0%	0%
Data-Model														
DM-D		65%	76%	41%	69%	39%*	24%	59%	31%	63%	34%	63%	23%	23%
DM-I		33%	51%	4%	94%	31%	1%	15%	1%	88%	48%	88%	28%*	28%*
DM-U		0%	0%	0%	95%**	8%	0%	0%	0%	94%*	53%*	94%*	3%	3%

\*\*-Highest ranked strategy

Table 5: Backtest Evaluation of Long-Only Portfolio Strategies

This table reports performance statistics from a backtest of portfolio strategies implemented by an investor with mean-variance preferences  $\mathbf{1}$  and risk aversion parameter  $\gamma = 2$  using the portfolio algorithms presented in section 4.3 based on rolling-window of 60-months of historical returns. All portfolios constrain weights to be non-negative and less than 10% and sum to unity. The geometric mean, arithmetic mean, and volatility are calculated for excess returns on the portfolio and annualized from monthly returns. Beta-Mkt, Beta-SMB, Beta-HML, and Beta-WML reflect the portfolio factor loadings, Alpha represents the annualized expected return mispricing for the portfolio, and Alpha t-Stat reports the t-Statistic for evaluating the significance of the mispricing. 1/N Monthly and 1/N Trail 12 Mth report the frequency with which the strategy's monthly return and trailing twelve month return, respectively, exceeded that of the 1/N rule. Mkt Monthly and Mkt Trail 12 Mth report the same for performance relative to the Market portfolio return. Turnover represents the annual average turnover for the strategy. Panel A reports results for the US Stock Universe with out-of-sample returns from 1968-2015, Panel B reports the results for the European Stock Universe without of sample returns from 1993-2008, and Panel C reports results for the European Mutual Fund Universe with out-of-sample returns from 1993-2008.

		Panel A: US Stocks Long Only Portfolio										Panel B: EU Stocks Long Only Portfolio										Panel C: EU Funds Long Only Portfolio															
		Subset Portfolios					Minimum Variance					Data and Model					Empirical Bayes																				
		$\hat{N} = 10$	$\hat{N} = 100$	GMV-S	GMV-SO	GMV-LW	DM-D	DM-I	DM-U	BS-S	BS-SO	BS-LW	$\hat{N} = 10$	$\hat{N} = 100$	GMV-S	GMV-SO	GMV-LW	DM-D	DM-I	DM-U	BS-S	BS-SO	BS-LW	$\hat{N} = 10$	$\hat{N} = 100$	GMV-S	GMV-SO	GMV-LW	DM-D	DM-I	DM-U	BS-S	BS-SO	BS-LW			
1968-2015, $T = 60$ , $\gamma = 2$	1/N	8.23%	9.89%**	6.07%	6.06%	5.18%	7.48%	7.96%	9.88%**	6.35%	6.35%	5.40%	1.54	1.70	1.41	1.79	1.60	1.60	-0.99	-1.37%	-0.61	0.91	1.08	1.34	1.54	1.70	1.41	1.79	1.60	1.60	-0.99	-1.37%	-0.61	0.91	1.08	1.34	
	Geometric Mean	8.19%	9.89%**	6.07%	6.06%	5.18%	7.48%	7.96%	9.88%**	6.35%	6.35%	5.40%	1.01	1.15	0.50	0.40	0.36	1.05	1.01	1.20	0.58	0.50	0.41	1.01	1.15	0.50	0.40	0.36	1.05	1.01	1.20	0.58	0.50	0.41			
	Arithmetic Mean	9.25%	12.60%**	6.35%	6.29%	5.40%	7.65%	10.05%	13.48%**	6.74%	6.74%	5.65%	0.07	0.15	0.07	0.07	0.00	0.25	0.22	0.39	0.06	0.03	-0.01	0.07	0.15	0.07	0.07	0.00	0.25	0.22	0.39	0.06	0.03	-0.01			
	Volatility	16.27%	24.79%	9.32%	8.82%	8.29%	21.60%	21.42%	28.02%	10.61%	10.03%	8.71%	0.19	-0.55	0.19	0.21	0.22	0.23	0.12	0.12	-0.61	0.19	0.28	0.25	0.22	-0.10	-0.55	0.19	0.21	0.22	0.23	0.12	0.12	0.12	0.28	0.25	
	Sharpe Ratio	0.569	0.529	0.682*	0.713**	0.652	0.447	0.469	0.481	0.636	0.666	0.649	0.07	0.22	0.07	0.07	0.04	0.33	0.36	0.30	0.09	0.10	0.05	0.05	0.05	-0.05	0.15	0.07	0.07	0.04	0.33	0.36	0.30	0.09	0.10	0.05	
	Alpha	0.80%	2.63%**	1.31%	1.90%	1.68%	-2.21%	-1.37%	2.82%**	0.94%	1.22%	1.41%	0.80%	2.63%**	1.31%	1.90%	1.68%	-2.21%	-1.37%	2.82%**	0.94%	1.22%	1.41%	0.80%	2.63%**	1.31%	1.90%	1.68%	-2.21%	-1.37%	2.82%**	0.94%	1.22%	1.41%			
	Alpha t-Stat	1.54	1.70	1.41	1.79	1.60	-0.99	-1.37%	-0.61	0.91	1.08	1.34	1.54	1.70	1.41	1.79	1.60	-0.99	-1.37%	-0.61	0.91	1.08	1.34	1.54	1.70	1.41	1.79	1.60	-0.99	-1.37%	-0.61	0.91	1.08	1.34			
	Beta-Mkt	1.01	1.04	1.04	1.04	1.04	1.05	1.01	1.01	1.20	0.58	0.41	1.01	1.04	1.04	1.04	1.04	1.05	1.01	1.20	0.58	0.50	0.41	1.01	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04		
	Beta-SMB	0.25	0.26	0.07	0.07	0.00	0.25	0.22	0.39	0.06	0.03	-0.01	0.07	0.15	0.07	0.07	0.00	0.25	0.22	0.39	0.06	0.03	-0.01	0.07	0.15	0.07	0.07	0.00	0.25	0.22	0.39	0.06	0.03	-0.01			
	Beta-HML	0.22	-0.10	0.19	0.21	0.22	0.23	0.12	0.12	-0.61	0.19	0.28	0.22	-0.10	-0.55	0.19	0.21	0.22	0.12	0.12	-0.61	0.19	0.28	0.25	0.22	-0.10	-0.55	0.19	0.21	0.22	0.12	0.12	0.12	0.28	0.25		
	Beta-MOM	-0.05	0.15	0.07	0.07	0.04	0.33	0.36	0.30	0.09	0.10	0.05	0.07	0.22	0.07	0.07	0.04	0.33	0.36	0.30	0.09	0.10	0.05	0.07	0.22	0.07	0.07	0.04	0.33	0.36	0.30	0.09	0.10	0.05			
	1/N Monthly	0%	53%**	48%	45%	45%	49%	50%	53%	49%	47%	46%	0%	53%**	48%	45%	45%	49%	50%	53%	49%	47%	46%	46%	0%	53%**	48%	45%	45%	49%	50%	53%	49%	47%	46%		
	1/N Trail 12 Mth	0%	50%	36%	38%	34%	59%**	59%**	57%	50%	40%	36%	0%	50%	36%	34%	59%**	59%**	57%	50%	40%	36%	36%	36%	0%	50%	36%	34%	59%**	59%**	57%	50%	40%	36%			
	Mkt Monthly	55%	55%**	48%	45%	46%	50%	52%	55%**	50%	48%	46%	55%	55%**	48%	45%	46%	50%	52%	55%**	50%	48%	46%	46%	46%	55%	55%**	48%	45%	46%	50%	52%	55%**	50%	48%	46%	
	Mkt Trail 12 Mth	62%**	59%	41%	43%	37%	59%	62%**	60%	45%	39%	39%	62%**	59%	41%	43%	37%	59%	62%**	60%	45%	39%	39%	39%	62%**	59%	41%	43%	37%	59%	62%**	60%	45%	39%	39%		
	Turnover	29%	147%	246%	168%	147%	287%	293%	238%	283%	180%	157%	29%	147%	246%	168%	147%	287%	293%	238%	283%	180%	157%	29%	147%	246%	168%	147%	287%	293%	238%	283%	180%	157%			
Panel B: EU Stocks Long Only Portfolio																																					
1988-2008, $T = 60$ , $\gamma = 2$	1/N	16.67%	18.91%**	15.25%	11.35%	7.90%	16.39%	14.20%	20.02%**	17.79%	14.97%	9.02%	1.18	1.018	1.509	1.326	0.992	0.877	0.766	0.975	10.14%**	5.97%	1.20%	1.18	1.018	1.509	1.326	0.992	0.877	0.766	0.975	10.14%**	5.97%	1.20%			
	Geometric Mean	16.49%	18.91%**	15.25%	11.35%	7.90%	16.39%	14.20%	20.02%**	17.79%	14.97%	9.02%	1.03	1.18	0.55	0.40	0.33	1.11	1.10	1.20	0.61	0.51	0.37	1.03	1.18	0.55	0.40	0.33	1.11	1.10	1.20	0.61	0.51	0.37			
	Arithmetic Mean	13.96%	15.15%	14.74%	11.15%	8.01%	17.20%	15.35%	20.63%**	17.08%	14.47%	9.02%	0.05	-0.23	0.05	0.26	0.29	0.13	0.11	-0.21	0.03	0.27	0.36	0.05	-0.23	0.05	0.26	0.29	0.13	0.11	-0.21	0.03	0.27	0.36			
	Volatility	13.96%	15.15%	14.74%	11.15%	8.01%	17.20%	15.35%	20.63%**	17.08%	14.47%	9.02%	0.09	0.29	0.09	0.08	0.06	0.35	0.37	0.38	0.10	0.11	0.08	0.09	0.29	0.09	0.08	0.06	0.35	0.37	0.38	0.10	0.11	0.08			
	Sharpe Ratio	1.181	1.118	1.509	1.326	0.992	0.877	0.766	0.975	1.548**	1.518*	1.060	1.181	1.118	1.509	1.326	0.992	0.877	0.766	0.975	1.548**	1.518*	1.060	1.181	1.118	1.509	1.326	0.992	0.877	0.766	0.975	1.548**	1.518*	1.060			
	Alpha	5.42%	7.35%**	8.20%**	3.99%	1.39%	1.59%	1.59%	7.18%	10.14%**	5.97%	1.20%	5.42%	7.35%**	8.20%**	3.99%	1.39%	1.59%	1.59%	7.18%	10.14%**	5.97%	1.20%	5.42%	7.35%**	8.20%**	3.99%	1.39%	1.59%	1.59%	7.18%	10.14%**	5.97%	1.20%			
	Alpha t-Stat	7.31	5.11	4.74	2.17	0.74	0.46	0.46	-0.07	3.14	3.14	0.63	7.31	5.11	4.74	2.17	0.74	0.46	-0.07	3.14	3.14	0.63	3.14	0.63	7.31	5.11	4.74	2.17	0.74	0.46	-0.07	3.14	3.14	0.63			
	Beta-Mkt	0.97	1.03	1.03	1.03	1.03	1.11	1.10	1.20	0.61	0.51	0.37	0.97	1.03	1.03	1.03	1.03	1.11	1.10	1.20	0.61	0.51	0.37	0.97	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03			
	Beta-SMB	0.47	0.51	0.33	0.23	0.23	0.61	0.68	0.67	0.36	0.26	0.24	0.47	0.51	0.33	0.23	0.23	0.61	0.68	0.67	0.36	0.26	0.24	0.47	0.51	0.33	0.23	0.23	0.61	0.68	0.67	0.36	0.26	0.24			
	Beta-HML	0.34	0.00	0.05	0.26	0.29	0.13	0.11	-0.21	0.03	0.27	0.36	0.34	0.00	0.05	0.26	0.29	0.13	0.11	-0.21	0.03	0.27	0.36	0.34	0.00	0.05	0.26	0.29	0.13	0.11	-0.21	0.03	0.27	0.36			
	Beta-MOM	-0.05	0.19	0.09	0.08	0.06	0.35	0.37	0.38	0.10	0.11	0.08	-0.05	0.19	0.09	0.08	0.06	0.35	0.37	0.38	0.10	0.11	0.08	-0.05	0.19	0.09	0.08	0.06	0.35	0.37	0.38	0.10	0.11	0.08			
	1/N Monthly	0%	53%**	46%	40%	35%	49%	47%	53%**	49%	44%	38%	0%	53%**	46%	40%	35%	49%	47%	53%**	49%	44%	38%	38%	0%	53%**	46%	40%	35%	49%	47%	53%**	49%	44%	38%		
	1/N Trail 12 Mth	0%	50%**	43%	30%	22%	47%	45%	59%	48%	40%	24%	0%	50%**	43%	30%	22%	47%	45%	59%	48%	40%	24%	24%	0%	50%**	43%	30%	22%	47%	45%	59%	48%	40%	24%		
	Mkt Monthly	71%**	66%**	66%**	44%	43%	59%	57%	77%	57%	50%	45%	71%**	66%**	66%**	44%	43%	59%	57%	77%	57%	50%	45%	45%	71%**	66%**	66%**	44%	43%	59%	57%	77%	57%	50%	45%		
	Mkt Trail 12 Mth	90%**	96%**	66%**	48%	32%	75%	74%	81%	62%	33%	33%	90%**	96%**	66%**	48%	32%	75%	74%	81%	62%	33%	33%	33%	90%**	96%**	66%**	48%	32%	75%	74%	81%	62%	33%			
	Turnover	28%	140%	281%	165%	132%	286%	284%	237%	326%	180%	146%	28%	140%	281%	165%	132%	286%	284%	237%	326%	180%	146%	28%	140%	281%	165%	132%	286%	284%	237%	326%	180%	146%			
Panel C: EU Funds Long Only Portfolio																																					
1988-2008, $T = 60$ , $\gamma = 2$	1/N	6.24%	9.96%**	6.36%	5.77%	5.84%	8.45%	10.81%**	9.71%**	6.45%	5.73%	5.88%	6.24%	9.96%**	6.36%	5.77%	5.84%	8.45%	10.81%**	9.71%**	6.45%	5.73%	5.88%	6.24%	9.96%**	6.36%	5.77%	5.84%	8.45%	10.81%**	9.71%**	6.45%					

Table 6: Subset Portfolio Simulated Expectations, Volatilities, and Sharpe Ratios: Variable Subset and Sample Size

This table reports statistics on the expected return (Panel A), volatility (Panel B), and Sharpe Ratio (Panel C) realized by an investor with mean-variance preferences 1 after observing  $T = 60$  months of returns and implementing subset portfolios of size  $\hat{N} = 10$  under variable values for the risk aversion parameter  $\gamma = 2$ . The reported performance averages across 400 simulations for the Restricted Four-Factor model and Augmented Factor model that includes an additional 10 principal components of returns as described in section 4.1. All portfolios implement the budget constraint that weights sum to unity, with the Long-Short portfolios constraining weights to be between -1000% and 1000% and the Long-Only portfolios constraining weights to be non-negative.

		Panel A: Simulated Average Expected Return by Risk Aversion ( $\gamma$ )											
		Restricted Factor Models					Augmented Factor Models						
		Long-Short Portfolios		Long-Only Portfolios		Long-Short Portfolios		Long-Only Portfolios		Long-Only Portfolios			
$T = 60, \hat{N} = 10$	$1/N$	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
Mean-Variance		7.05%	8.70%	8.05%	7.05%	8.70%	8.05%	6.90%	8.69%	8.28%	6.90%	8.69%	8.28%
$\gamma = 0.1$		2661.83%	1920.65%	737.57%	8.25%	10.18%	9.25%	3570.26%	1951.76%	773.87%	8.52%	10.33%	9.65%
$\gamma = 1$		36.43%	27.34%	13.59%	8.20%	10.34%	8.87%	49.24%	27.43%	14.07%	8.40%	10.36%	9.16%
$\gamma = 2$		21.32%	17.76%	9.93%	8.14%	10.27%	8.51%	27.42%	17.66%	10.22%	8.27%	10.22%	8.75%
$\gamma = 5$		12.26%	12.00%	7.73%	7.91%	9.89%	7.81%	14.31%	11.80%	7.91%	7.88%	9.78%	7.99%
$\gamma = 10$		9.23%	10.09%	7.00%	7.61%	9.41%	7.24%	9.95%	9.85%	7.14%	7.44%	9.25%	7.40%
$\gamma = 25$		7.42%	8.93%	6.56%	7.15%	8.79%	6.76%	7.33%	8.67%	6.68%	6.85%	8.59%	6.90%
$\gamma = 50$		6.82%	8.55%	6.41%	6.91%	8.51%	6.60%	6.45%	8.28%	6.52%	6.55%	8.29%	6.74%
		Panel B: Simulated Average Volatility of Return by Risk Aversion ( $\gamma$ )											
		Restricted Factor Models					Augmented Factor Models						
		Long-Short Portfolios		Long-Only Portfolios		Long-Short Portfolios		Long-Only Portfolios		Long-Only Portfolios			
$T = 60, \hat{N} = 10$	$1/N$	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
Mean-Variance		13.94%	12.52%	16.63%	13.94%	12.52%	16.63%	14.07%	12.70%	16.70%	14.07%	12.70%	16.70%
$\gamma = 0.1$		3364.43%	2349.64%	2390.62%	14.38%	14.06%	19.85%	6304.21%	3250.47%	2629.84%	15.24%	14.71%	20.19%
$\gamma = 1$		40.59%	27.82%	29.24%	14.12%	13.34%	18.04%	73.51%	36.19%	31.30%	14.96%	13.94%	18.26%
$\gamma = 2$		22.78%	17.03%	18.58%	13.86%	12.80%	16.79%	38.71%	21.04%	19.47%	14.68%	13.36%	16.94%
$\gamma = 5$		13.38%	11.58%	13.32%	13.24%	11.77%	14.70%	19.26%	13.29%	13.63%	13.95%	12.38%	14.79%
$\gamma = 10$		11.07%	10.22%	11.97%	12.61%	10.89%	13.23%	14.03%	11.41%	12.16%	13.24%	11.58%	13.31%
$\gamma = 25$		10.11%	9.56%	11.30%	11.94%	10.02%	12.10%	11.88%	10.61%	11.44%	12.52%	10.85%	12.20%
$\gamma = 50$		9.87%	9.38%	11.10%	11.72%	9.72%	11.78%	11.44%	10.41%	11.24%	12.29%	10.61%	11.88%
		Panel C: Simulated Average Sharpe Ratio for Return by Risk Aversion ( $\gamma$ )											
		Restricted Factor Models					Augmented Factor Models						
		Long-Short Portfolios		Long-Only Portfolios		Long-Short Portfolios		Long-Only Portfolios		Long-Only Portfolios			
$T = 60, \hat{N} = 10$	$1/N$	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock	EU Fund	EU Stock	US Stock
Mean-Variance		0.506	0.695	0.484	0.506	0.695	0.484	0.490	0.684	0.496	0.490	0.684	0.496
$\gamma = 0.1$		0.832	0.827	0.303	0.575	0.727	0.468	0.592	0.616	0.288	0.561	0.706	0.480
$\gamma = 1$		0.934	1.003	0.468	0.583	0.778	0.493	0.697	0.775	0.452	0.563	0.746	0.504
$\gamma = 2$		0.972	1.069	0.588	0.547	0.805	0.509	0.733	0.853	0.535	0.565	0.767	0.518
$\gamma = 5$		0.944	1.054	0.590	0.599	0.842	0.533	0.756	0.893	0.589	0.566	0.792	0.542
$\gamma = 10$		0.847	0.994	0.588	0.604	0.866	0.548	0.713	0.864	0.590	0.563	0.800	0.557
$\gamma = 25$		0.738	0.936	0.581	0.599	0.879	0.559	0.617	0.818	0.584	0.547	0.792	0.566
$\gamma = 50$		0.693	0.913	0.578	0.589	0.876	0.561	0.564	0.796	0.581	0.533	0.781	0.567

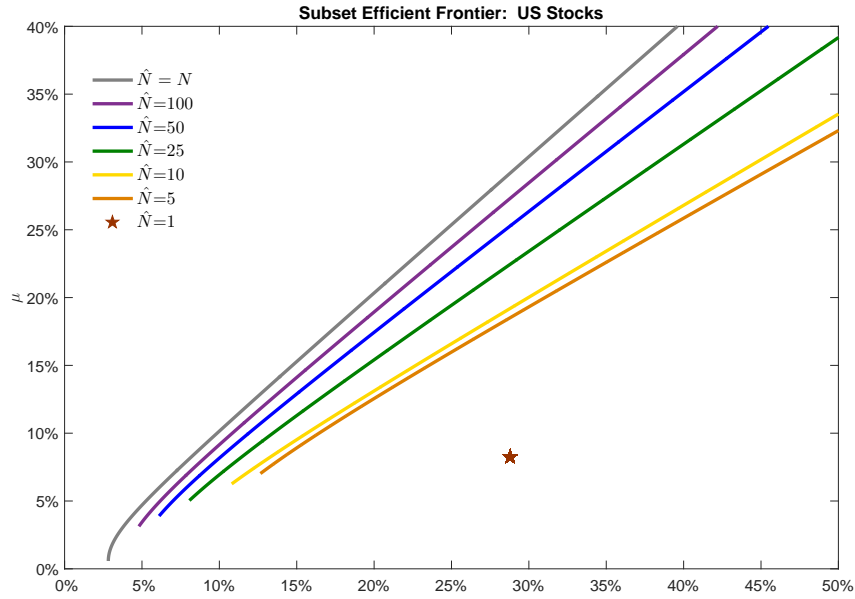


Figure 1: Population Complete Subset Efficient Frontier for Simulated US Stock Data  
This figure presents the mean-variance tradeoffs for investors implementing complete subset portfolios with access to population means and variances calibrated to a simulation universe representing  $N = 1,063$  US Stocks. For each subset, investors minimize portfolio variance allowing for short sales subject to the expected return target and complete subset portfolios aggregate individual subsets with equal weights across subsets. The  $1/N$  rule appears as a point representing the complete population subset portfolio of size one. As the subset size grows, the complete subset efficient frontier expands and converges to the efficient frontier generated by including all securities in the portfolio.

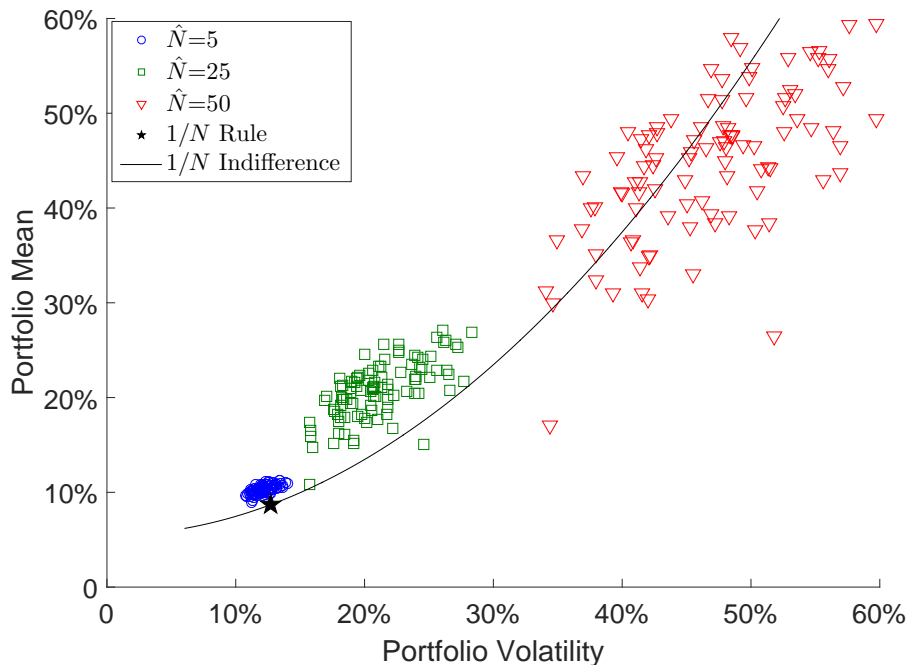


Figure 2: Subset Size, Sampling Error in Weights, and Portfolio Performance

This figure presents the mean-variance tradeoffs for investors choosing the subset size when implementing complete subset portfolios securities with population means and variances calibrated to a universe of  $N = 1,063$  US Stocks. The star represents the mean and volatility for the  $1/N$  rule with the black indifference curve reflecting the combinations of means and variances that lead to indifference for an investor with mean-variance preferences and risk aversion parameter  $\gamma = 4$ . From 100 simulation samples with 100 months' returns each, the blue circles represent the true means and variances from implementing the estimated complete subset portfolio weights with  $\hat{N} = 5$  and the green squares and red triangles represent the same with  $\hat{N} = 25$  and  $\hat{N} = 50$ , respectively. Though the performance of the strategy with  $\hat{N} = 5$  always lies above and to the left of the  $1/N$  rule's indifference curve, as the subset size becomes large, estimated portfolio weights, and consequent performance, become noisier and leads to potentially bad outcomes for investors.

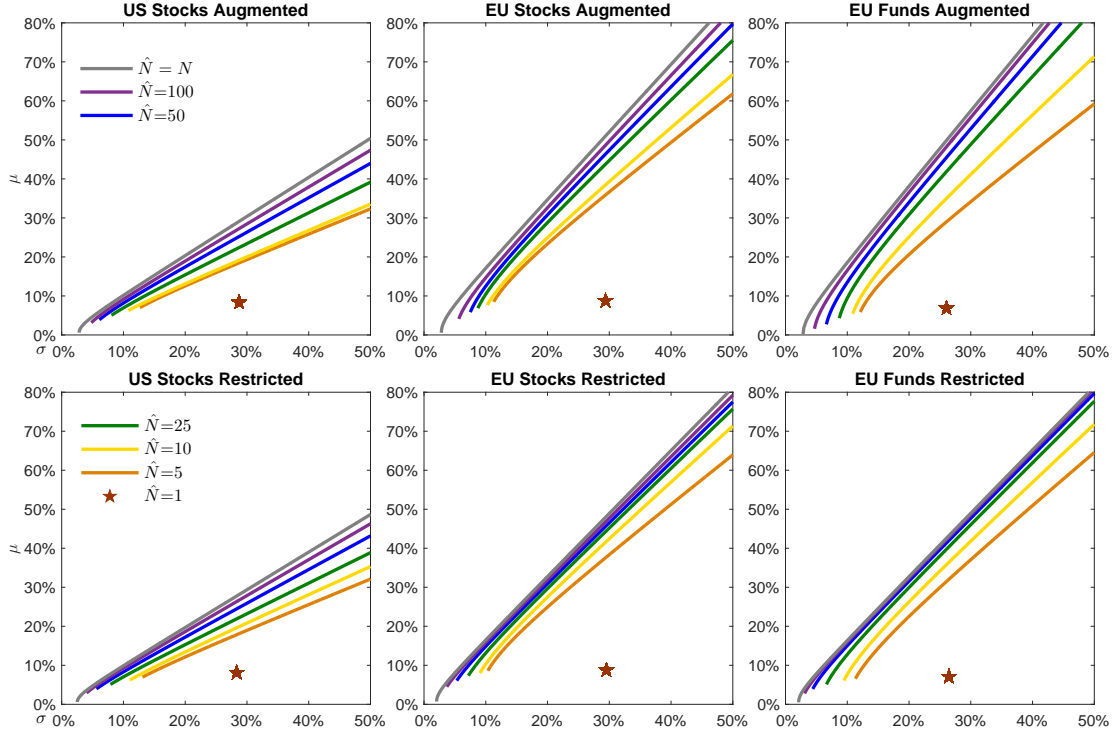
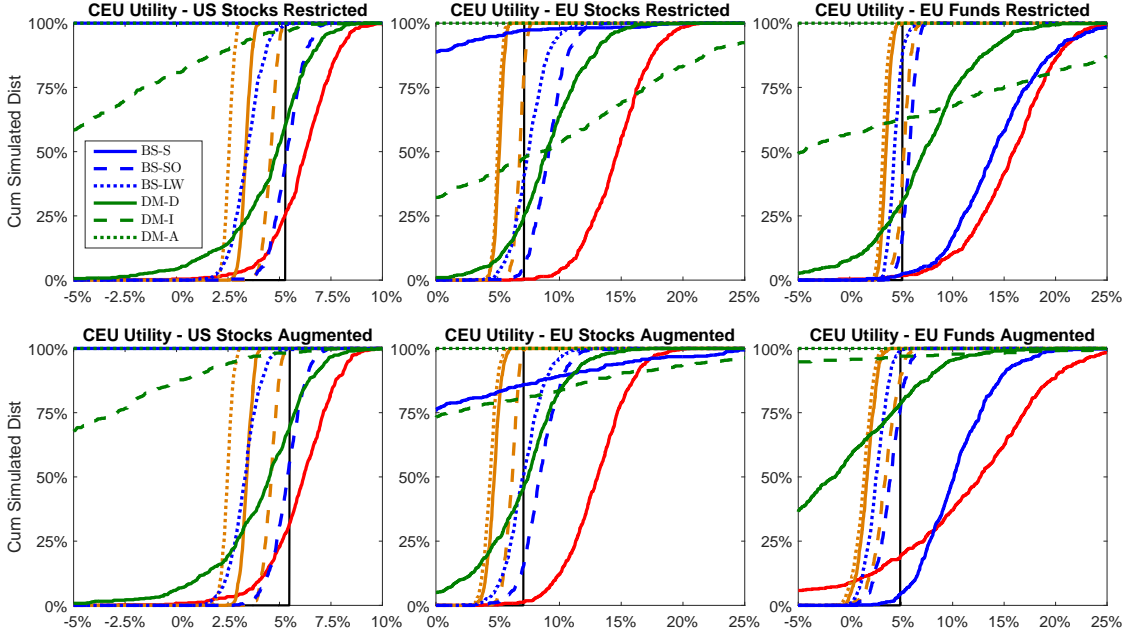
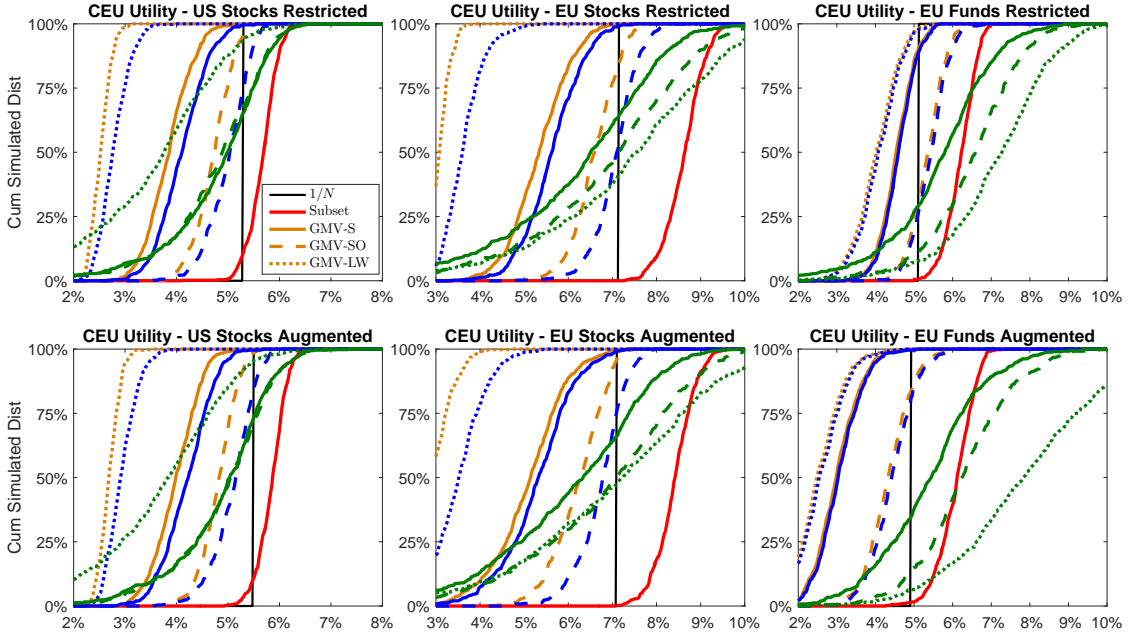


Figure 3: Population Subset Efficient Frontier for Simulated Asset Universes

This population subset efficient frontier characterizes the tradeoff between the expected return on a portfolio  $\mu$  and the portfolio's volatility  $\sigma$  when return generating process moments are known by the investor. The star represents the mean and variance of the  $1/N$  portfolio, corresponding to subset portfolios of size one. Different colors correspond to different subset portfolio sizes. Data for the universe of US Stocks (1963-2015) comes from the CRSP database, European Stocks (1988-2008) from Datastream, and European Equity Mutual Funds (1988-2008) from the Lipper database. The Restricted model of returns fits a zero-alpha four-factor model by regressing assets' excess returns on benchmark factors, with idiosyncratic volatility equal to the residual standard deviation. The four benchmark factors (Mkt, SMB, HML, Mom) come from Ken French's website, with factor expected returns and covariances equaling historical sample moments. The Augmented model extracts an additional ten latent priced factors from the security universe using principal components and calibrates security returns to a fourteen factor (four benchmarks + 10 PCAs).



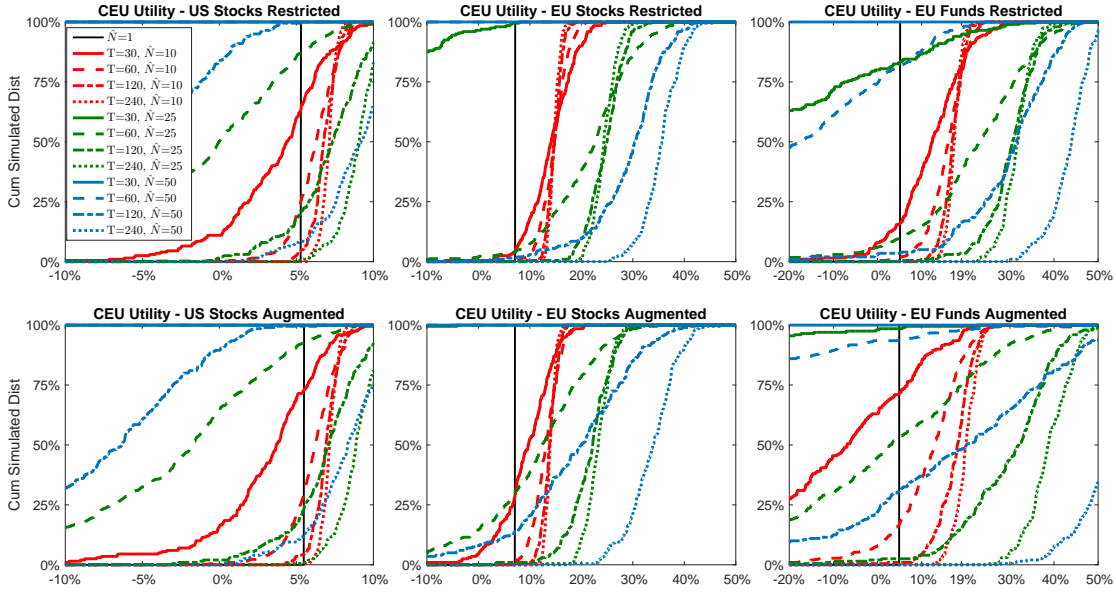
Panel A: Long-Short Portfolios



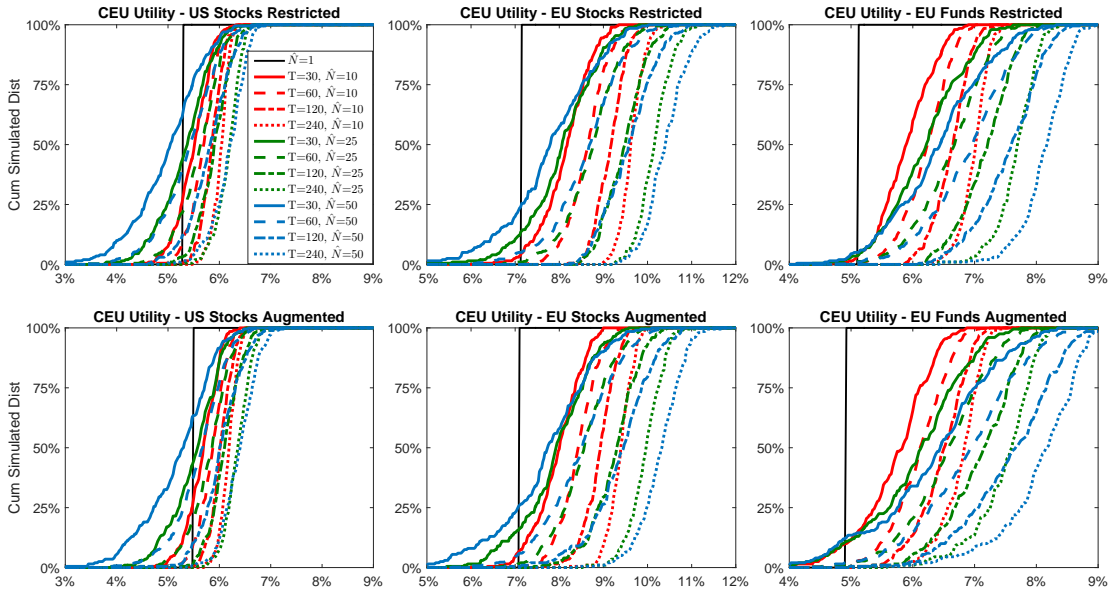
Panel B: Long-Only Portfolios

Figure 4: Simulated Distribution of Portfolio Certainty Equivalent Utilities,  $T = 60, \gamma = 2$

This figure plots the distribution of Certainty Equivalent Utility (CEU) for an investor with mean-variance preferences (1) and risk aversion parameter  $\gamma = 2$  when implementing the portfolio algorithms presented in section 4.3 based on 500 simulated 60-month histories of returns. All portfolios implement the budget constraint that weights sum to unity, with the Long-Short portfolios constraining weights to be between -1,000% and 1,000% and the Long-Only portfolios constraining weights to be non-negative. The black line represents the CEU of the 1/N rule while the solid red line represents the cumulated distribution for the simulated CEU of Subset portfolios with size  $\hat{N} = 10$ . The blue lines represent the performance of Empirical Bayesian estimators, with the different dashed presentations corresponding to different covariance matrices. The green lines characterize that of the Data and Model estimators, with the different dashed presentations corresponding to different prior beliefs.



Panel A: Long-Short Portfolios



Panel B: Long-Only Portfolios

Figure 5: Simulated Distribution of Portfolio Certainty Equivalent Utilities when Varying Subset Size and Sample History

This figure plots the distribution across 400 simulations of Certainty Equivalent Utilities (CEU) realized by an investor with mean-variance objective from Equation (1) parameterized by risk aversion coefficient  $\gamma = 2$  access to  $T \in \{30, 60, 120, 240\}$  months of normally-distributed return data. Portfolio weights are calculated by the subset portfolio algorithm using sample estimates for asset expectations and covariances with subset sizes  $\dot{N} \in \{1, 10, 25, 50\}$  both allowing for long and short positions (Panel A) and restricting weights to be non-negative (Panel B). The six subplots within each panel relate to the US Stock, European Stock, and European Mutual Fund Universes calibrated according to the Restricted four-factor model and the PCA Augmented fourteen factor models presented in section 4.1.