Risk Assessment in Large Portfolios: why imposing the wrong constraints hurts☆

Daniele Bianchi[*,a], Carlos M. Carvalho[b]

[a] *Department of Finance, Bocconi University, Milan, Italy*
[b] *IROM, McCombs School of Business, University of Texas at Austin, USA*

Abstract

Jagannathan and Ma (2003) argue that with no-short sale constraints in place, the sample covariance matrix performs as well as more sophisticated covariance matrix estimates in reducing portfolio risk exposure. This is still argued considering more general convex gross-exposure constraints. Introducing a common Bayesian framework, we empirically explain why artificially constraining portfolio weights is suboptimal from a statistical decision theoretical point of view. Considering portfolio allocations as decision rules, we introduce a basic actual bayes risk function and check for admissibility of constraining weights with respect to a general dynamic Bayesian portfolio allocation model. We concentrate on global minimum variance portfolios. Based on simulation exercises and an empirical investigation on daily AMEX/NYSE returns, we provide evidence of inadmissibility/suboptimality of artificially imposed portfolio constraints in an expected utility framework, being therefore unconsistent with a standard von Neumann-Morgenstern rationality assumption.

Keywords: Risk Reduction, DLM, shrinkage estimators, portfolio allocation, no-short-sale, gross-exposure contraints, Large Portfolios, Expected Utility

1. Introduction

Portfolio theory and selection has been one of the fundamental theoretical development in finance. Since the seminal work in Markowitz (1952a) and Markowitz (1952b), portfolio selection in a mean-variance framework represents, by far, the most common formulation of portfolio choice problems. Markowitz portfolios, have had indeed a deep impact on financial economics literature, and more widely, is considered a milestone in modern finance. To implement these portfolios in practice, one has to estimate the mean and covariance matrix of asset returns. The majority of the portfolio choice literature falls under the heading of plug-in estimation. The econometrician estimates the parameters of the data generating process through sample moments and plugs them into either

☆This draft: May 18, 2011. We thank Peter Mueller, Stathis Tompaidis, and seminar participants at McCombs Business School. We are responsible for any errors and omissions.

*Corresponding author

an analytical or numerical solution to the investor's optimization problem. However, the implementation of mean-variance portfolios with sample moments estimates is notorious for producing extreme weights that highly fluctuate over time and poorly perform out-of-sample. The instability of portfolio weights is mainly due to estimation error, especially on the expected returns, as pointed out in Chopra and Ziemba (1993) and Merton (1980). For this reason, researchers have recently focused on the minimum-variance portfolios, which rely solely on the covariance structure, and supposedly perform better out-of-sample. Although disregard the mean estimation, Global Minimum Variance Portfolios still suffer with instability due to covariance estimation errors. This is more pronounced when the portfolio size N is large (see Fan et al. (2008)), getting close to the sample size T. Several solutions have been proposed in the literature to mitigate estimation error. Jagannathan and Ma (2003), propose to artificially non-negative constraint allocations to induce portfolio stability. Likewise, Fan et al. (2008), DeMiguel et al. (2009a) and Brodie et al. (2008) propose to use gross-exposure constraints essentially treating the portfolio allocation problem as a penalized regression. More generally, imposing artificial constraints is argued to help in stabilize the weights, reducing risk exposure, through covariance regularization, regardless the type of covariance estimator, as reported in Jagannathan and Ma (2003). However, according to Green and Hollifield (1992), these constraints are likely to be wrong in population and hence introduce specification error. Extreme negative and positive weights unlikely are due solely on estimation error, but depend on the unknown structure of the covariance matrix. Other techniques have been suggested to reduce sensitivity of Markowitz portfolios trying to get more stable portfolio weights. Jorion (1986) proposed a James-Stein estimator for the means, while Ledoit and Wolf (2004) proposed a shrinkage estimator of the covariance matrix towards either the indentity or a factor model covariance. More general Bayesian Portfolio allocation approaches are considered in Polson and Tew (2000), Frost and Savarino (1986), Jorion (1986), McCulloch and Rossi (1990), Pastor and Stambaugh (2000) and Pastor (2000). Bayesian portfolio optimization not only helps in reducing estimation error through the specification of the prior, but addresses also parameter uncertainty integrating out the state parameters. The portfolio allocation is then maximized with respect to the predictive distribution of the returns. Finally resampling base procedure are proposed in Michaud (1998) among the others, aimed to address parameter uncertainty ex-post in certainty equivalent framework. We focus on the Global Minimum Variance Portfolio [GMVP] just as in Jagannathan and Ma (2003), Brodie et al. (2008) and DeMiguel et al. (2009a) among the others. The goal of the paper is to empirically investigate the admissibility of constrained plug-in portfolio allocations, using sample covariance estimates, as optimal decision rules in an expected utility maximization framework. Essentially, we want to point out the suboptimality of imposing artificial portfolio constraints just for regularization purposes, arguing that careful covariance estimates are marginally more relevant for risk reduction, especially in large portfolios. In the portfolio allocation problem, the weights $\mathbf{w}_t$ are interpreted as outcomes of a decision process $\mathbf{w}_t = a(\theta_t)$, in which the action $a \in \mathcal{A}$ is the functional form of the weights. This matches the standard definition of *action* in the statistical decision theoretical literature, defined as $a : \theta_t \rightarrow \mathbf{w}_t$, i.e. a function mapping the state (the covariance) in an outcome (the weights). The *true* states are normally unobservable therefore the agent is assumed to choose among different actions whose consequences cannot be anticipated. In an expected utility framework this is managed assigning a quantitative value to the investor's utility function for each outcome $\mathbf{w}_t$ and a probability distribution $p(\theta)$. The key point is that constraining portfolio weights define an action tha can be compared to the others on the basis of the expected utility. Under the von Neumann-Morgenstern rationality assumption the representative investor is assumed to choose the action maximizes the expected value of the utility function given the distribution of the states. In the minimum variance framework this translates to rank the actions on the basis of portfolio risk exposure, then choosing the one minimizes it. We do that as a first step discriminating different portfolio decision rules on the basis of portfolio risk exposures. This classic horse race is extended defining a common Bayesian framework in which each of the portfolio decision rules is taken under the same predictive distribution, allowing to isolate the effect of the action on the portfolio out-of-sample performances. Based on simulation examples and a real dataset, we find that artificially constraining portfolio weights for risk reduction purposes, turns out to be a dominated strategy, both in terms of expected utility and risk exposure. From a statistical decision theory standpoint this translate in inadmissibility, meaning, contradicts the standard von Neumann-Morgenstern rationality assumption. Both simulation and empirical analysis point out the relevant of imposing dynamics and structure on the covariance estimates regardless artificial constraints.

The main contributions of the paper are the following: (1) we investigate in a standard expected utility maximization framework different portfolio strategies stressing the incompatibility of constraining weights with a standard rationality assumption in a decision process. We (2) extend the expected utility comparison defining a common Bayesian framework, isolating the effect of the action on a general risk function under a predictive distribution. Yet, (3) we propose a dynamic Bayesian covariance estimator with a sequentially defined shrinkage prior in the spirit of Ledoit and Wolf (2003).

The rest of the paper is organized as follows. In Section 2 we recall a general portfolio selection problem and the specific case of minimum variance framework. Section 3 reports a first set of simulation results comparing the different portfolio decision rules in an expected utility framework with a standard dynamic Bayesian covariance estimator as benchmark. Then Section 4 generalize the expected utility framework to a common Bayesian framewor by which we investigate constraining portfolio rules admissibility with on the basis of a standard decision theoretical inspired risk function. Section 5 and 6 report respectively the benchmark dynamic Bayesian covariance estimator with a sequential shrinkage prior specification and the simulation results relatively decision rules admissibility. Section 7 implements the common Bayesian framework analysis on a real daily dataset, while finally Section 8 report the concluding remarks.

2. The Portfolio Selection problem

2.1. A general expected utility formulation

Let us introduce a standard portfolio selection problem in which a representative investor aims to maximize the expected value of a general, continuosly differentiable, utility function. In the static decision problem, the investor aims to form a portfolio at $T+\tau$, with τ the investment horizon. Let us consider an N-dimensional vector of prices at each time t as $\mathbf{P}_t$ and suppose $\tau = 1$. The linear returns for the i_{th} asset are computed as $\mathbf{R}_t = (\mathbf{P}_{t+1}/\mathbf{P}_t) - 1$. Given a risk-free rate R_t^f the observed excess returns are $\mathbf{r}_t = \mathbf{R}_t - \mathbf{1}_N R_t^f$. The vector of relative weights $\mathbf{w} = \{w_1, ..., w_N\}$ is defined as

$$\mathbf{w} = \frac{diag(\mathbf{P}_t)\mathbf{c}}{\mathbf{c}'\mathbf{P}_t} \qquad \text{such that} \qquad \mathbf{r}_{p,t} = \mathbf{w}'\mathbf{r}_t \tag{1}$$

with $\mathbf{c} = \{c_1, ..., c_N\}$ the dollar amount invested in each stock. The excess returns are assumed to have a general distribution function $\mathbf{r}_t|\theta_t \sim p(\mathbf{r}_t|\theta_t)$, where $\theta_t \in \Theta$ being the unknown distribution parameters. Suppose the investor want to allocate the wealth on the N risky securities. From a decision theoretical point of view the vector of relative weights can be interpreted as the outcome of a decision process developed combining an action $a \in \mathcal{A}$ and the state of nature $\theta_t \in \Theta$, meaning $a : \theta_t \to \mathbf{w}_t$ such that $\mathbf{w}_t = a(\theta_t)$, (see Parmigiani and Inoue (2009) for more details). In the portfolio allocation problem the action is the functional form of the weights, the state of natures are the moments estimates and the outcome is the portfolio weights. The true states are unobservable so the investor is formally assumed to choose among different actions whose outcomes cannot be completely anticipated. Expected utility theory handles this choice by assigning a quantitative utility to each outcome $\mathbf{w}_t$ and probability distribution to θ_t. The representative investor is then assumed to satisfy the von Neumann-Morgenstern [NM] rationality assumption, selecting the action $a \in \mathcal{A}$ which maximizes the expected value of the resulting utility under the state probability distribution.

Most of the reference literature on portfolio allocation has been casted in a *certainty equivalence* framework, in which the underlying distribution is assumed to be completely identified, and the optimal portfolio is the solution of

$$\begin{aligned} \mathbf{w}_t^* &\in \arg\max E_{\mathbf{r}_t|\theta_t}\left[U\left(\mathbf{w}_t'\mathbf{r}_t\right)\right] \equiv \int_{\mathcal{R}} U\left(\mathbf{w}_t'\mathbf{r}_t\right) p\left(\mathbf{r}_t|\theta_t = \hat{\theta}_t\right) d\mathbf{r}_t \\ \text{s.t.} &\quad \mathbf{w}_t \in C\left(\mathbf{w}_t\right) = \left\{\mathbf{w}_t|\mathbf{w}_t'\iota = 1\right\} \end{aligned} \tag{2}$$

where $\iota = \{1, ..., 1\}'$ is an N-dimensional vector of ones, $C(\mathbf{w}_t)$ is the convex set of feasible portfolio choices and $\mathcal{R} = \mathbb{R}$ the space of returns. Finding the optimal solution to (2) is then traditionally defined as a two-steps procedure in which the econometrician firstly estimates $\hat{\theta}_t$ from the data, then solves (2) for θ_t given. This approach ignores estimation risk and parameter uncertainty (see Jorion (1986) and Klein and Bawa (1976) for more details). Estimation error is directly related to portfolio size, while parameter uncertainty increases as the amount of information (sample size) decreases. Yet, in general, state parameters are neither known, nor observable, implying that $p(\mathbf{r}_t|\hat{\theta}_t = \theta_t)$ is not completely specified.

Several solutions have been proposed in the literature to mitigate estimation error and parameter uncertainty. Jagannathan and Ma (2003), DeMiguel et al. (2009b), DeMiguel et al. (2009a) and Fan et al. (2008) among the others, proposed to modify (2) plugging artificial constraints solving the following

$$\begin{aligned} \mathbf{w}_t^* &\in \arg\max E_{\mathbf{r}_t|\theta_t}\left[U\left(\mathbf{w}_t'\mathbf{r}_t\right)\right] \equiv \int_{\mathcal{R}} U\left(\mathbf{w}_t'\mathbf{r}_t\right) p\left(\mathbf{r}_t|\theta_t = \hat{\theta}_t\right) d\mathbf{r}_t \\ \text{s.t.} &\quad \mathbf{w}_t \in C\left(\mathbf{w}_t\right) = \left\{\mathbf{w}_t|\mathbf{w}_t'\iota = 1, \|\mathbf{w}_t\|_1 \leq c\right\} \end{aligned} \tag{3}$$

The optimal allocation still comes from a two-steps procedure, however imposing $\|\mathbf{w}_t\|_1 \leq c$, helps in stabilize the portfolio weights through a regularization effect on the historical covariance matrix[1]. This reduces the effect of estimation error on portfolio risk exposure.

[1] The regularization effect comes from a reduction of the higher eigenvalues come from estimation error and an increasing of the lower eigenvalues of the covariance matrix due to sampling errors

Defining $c = 1$ we get the no-short sales constraints as in Jagannathan and Ma (2003) and DeMiguel et al. (2009a), while for other c we resort to the general gross-exposures in Fan et al. (2008). However (3) does not address parameter uncertainty since the whole procedure is still a plug-in approach. The regularization benefit just comes from a purely mathematical optimization argument.

On the other hand, parameter uncertainty, together with estimation risk, can be mitigated through a Bayesian framework in which θ_t is assumed to be a random quantity on itself. The quantative utility of the outcome $\mathbf{w}_t = a(\theta_t)$ is averaged over the predictive distribution of the excess returns as proposed in Jorion (1986), Frost and Savarino (1986), Klein and Bawa (1976), Pastor (2000) and Polson and Tew (2000) among the others. The optimal portfolio allocation problem becomes as follows

$$\begin{aligned} \mathbf{w}_t^* \quad &\in \quad \arg\max E_{\mathbf{r}_t}\left[U\left(\mathbf{w}_t'\mathbf{r}_{t+1}\right)\right] \equiv \int_{\mathcal{R}} U\left(\mathbf{w}_t'\mathbf{r}_{t+1}\right) p(\mathbf{r}_{t+1}|\mathbf{r}) d\mathbf{r}_{t+1} \\ \text{s.t.} \quad & \quad \mathbf{w}_t \in C\left(\mathbf{w}_t\right) = \left\{\mathbf{w}_t | \mathbf{w}_t'\iota = 1\right\} \end{aligned} \tag{4}$$

where

$$p(\mathbf{r}_{t+1}|\mathbf{r}) \propto \int_{\Theta} p(\mathbf{r}_{t+1}|\mathbf{r},\theta_t) p(\theta_t|\mathbf{r}) d\theta_t \qquad \text{with} \qquad p(\theta_t|\mathbf{r}) \propto p(\mathbf{r}|\theta_t) p(\theta_t) \tag{5}$$

with $\mathbf{r}$ the $(N \times T)$ matrix of returns upto time t, $p(\theta_t|\mathbf{r})$ the posterior distribution of the parameters and $p(\mathbf{r}_{t+1}|\mathbf{r},\theta_t)$ the likelihood. Essentially the predictive distribution averages the probability of a future observation over the posterior distribution of the states, such that the future returns depend juts on past returns. The Bayesian approach (4) helps in accounting for parameter uncertainty integrating out the states in (5) and reducing the estimation error through the prior $p(\theta_t)$ specification.

Other solutions proposed to both reduce estimation error and account for parameter uncertainty from a frequentist point of view are due to Michaud (1998) among the others. The bottom line is to resample the vector of relative weights, bootstrapping returns in a parametric setting. This procedure, however, involves solving the portfolio allocation first in plug-in framework, then averaging out the re-sampled weights. This means to address estimation error and uncertainty ex-post. In the Bayesian setting insetad, uncertainty is taken into account ex-ante, before solving the investor's optimization problem. A brief description of a re-sampling based approach is provided in the appendix. Next section focus on the minimum-variance portfolio just as in Jagannathan and Ma (2003), Fan et al. (2008) and DeMiguel et al. (2009a).

2.2. Minimum Variance Portfolio Selection

Let us consider the same risk-averse investor, with $\tau = 1$ as investment horizon, who must allocate funds between a portfolio of N risky assets. As before the relative weights represent the outcome of a decision process through the action a given the state of nature θ_t. Returns $\mathbf{R}_t$ and linear excess returns $\mathbf{r}_t$ are defined as in Section 2.1. Suppose $\mathbf{r}_t \sim N(\mu_t, \Sigma_t)$, with $\mu_t = \{\mu_{1,t}, ..., \mu_{n,t}\}$ and $\Sigma = \{\sigma_{i,j,t}\}$ respectively mean and covariances, such that $\theta_t = (\mu_t, \Sigma_t)$ [2]. The excess return of the portfolio $\mathbf{r}_{p,t} = \mathbf{w}_t'\mathbf{r}_t$ is normal with mean $\mu_{p,t} = \mathbf{w}_t'\mu_t$ and variance $\sigma_{p,t}^2 = \mathbf{w}_t'\Sigma_t\mathbf{w}_t$. Considering a quadratic utility function. , the standard, two-steps, portfolio allocation problem becomes

$$\begin{aligned} \mathbf{w}_t^* \quad &\in \quad \arg\max E_{\mathbf{r}_t|\theta}\left[U(\mathbf{w}_t'\mathbf{r}_t)\right] \equiv \int_{\mathcal{R}} U(\mathbf{w}_t'\mathbf{r}_t)p(\mathbf{r}_t|\theta_t)d\mathbf{Y}_t = \mathbf{w}_t'\mu_t - \frac{\gamma}{2}\mathbf{w}_t'\Sigma_t\mathbf{w}_t \\ \text{s.t.} \quad & \quad \mathbf{w}_t \in C(\mathbf{w}_t) = \left\{\mathbf{w}_t|\mathbf{w}_t'\iota = 1\right\} \end{aligned} \tag{6}$$

where $\gamma > 0$ is the risk aversion parameter. Let us consider zero-mean multivariate returns, i.e $\mathbf{Y}_t = \mathbf{r}_t - \mu$, such that $\mathbf{Y}_t|\Sigma_t \sim N(0, \Sigma_t)$, just as in Jagannathan and Ma (2003) and Fan et al. (2008) among the other. This is economically plausible for daily returns. Yet the aim is to isolate the comparison of decision rules on covariance estimates. The mean-variance portfolio allocation problem is rewritten as

$$\begin{aligned} \mathbf{w}_t^* \quad &\in \quad \arg\max E_{\mathbf{Y}_t|\theta_t}\left[U(\mathbf{w}_t'\mathbf{Y}_t)\right] \equiv \int_{\mathcal{Y}} U(\mathbf{w}_t'\mathbf{Y}_t)p(\mathbf{Y}_t|\theta_t)d\mathbf{Y}_t = -\frac{1}{2}\mathbf{w}_t'\Sigma_t\mathbf{w}_t \\ \text{s.t.} \quad & \quad \mathbf{w}_t \in C(\mathbf{w}_t) = \left\{\mathbf{w}_t|\mathbf{w}_t'\iota = 1\right\} \end{aligned} \tag{7}$$

Equation (7) represents the well known Global Minimum Variance Portfolio [GMVP]. The $\gamma > 0$ risk aversion parameter just shifts the optimal solution then disappear in the optimal programming. Just as in the general portfolio allocation problem, estimation error and paramater uncertainty play a key role in getting valuable out-of-sample results. The artificial constraining approach of Jagannathan and Ma (2003) translates to

$$\begin{aligned} \mathbf{w}_t^* \quad &\in \quad \arg\max E_{\mathbf{Y}_t|\theta_t}\left[U(\mathbf{w}_t'\mathbf{Y}_t)\right] \equiv \int_{\mathcal{Y}} U(\mathbf{w}_t'\mathbf{Y}_t)p(\mathbf{Y}_t|\theta_t)d\mathbf{Y}_t = -\frac{1}{2}\mathbf{w}_t'\hat{\Sigma}_t\mathbf{w}_t \\ \text{s.t.} \quad & \quad \mathbf{w}_t \in C(\mathbf{w}_t) = \left\{\mathbf{w}_t|\mathbf{w}_t'\iota = 1, \mathbf{w}_t \geq 0\right\} \end{aligned} \tag{8}$$

with $\hat{\Sigma}_t$ the standard historical covariance estimator at time t. The L1-norm constraints supposedly have a regularization effect on the covariance matrix. This is formalized in Proposition 1 in Jagannathan and Ma (2003) as follows:

[2] Here, $\sigma_{i,j,t}$ denotes the covariance between stock i_{th} and j_{th} at time t, and $\sigma_{i,i} > 0$.

Proposition 1 (Jagannathan and Ma (2003)). *Let $\hat{\Sigma}$ be the historical covariance estimator, and $\delta = (\delta_1, ...\delta_N)$, $\lambda = (\lambda_1, ..., \lambda_N)$ being respectively the lagrange multipliers for the nonnegativity constraints, i.e. $\mathbf{w} \geq 0$ and the upper bounds $\mathbf{w} \leq \overline{\mathbf{w}}$. Then plugging*

$$\tilde{\Sigma} = \hat{\Sigma} + \mathbf{J} \qquad \textit{with} \qquad \mathbf{J} = (\delta \mathbf{1}' + \mathbf{1}\delta') - (\lambda \mathbf{1}' + \mathbf{1}\lambda') \tag{9}$$

in an unconstrained mean-variance optimization is equivalent of using the historical $\hat{\Sigma}$ covariance estimator with the non-negativity and upper bounds.

Essentially, imposing portfolio constraints, regardless the covariance estimates, is argued to mitigate estimation error reducing portfolio risk exposure as a whole. On the other hand the Bayesian approach defined in (4) translates in the minimum-variance portfolio choice as

$$\begin{aligned} \mathbf{w}_t^* \quad &\in \quad \arg\max \int_{\mathcal{Y}} U(\mathbf{w}_t' \mathbf{Y}_{t+1}) p(\mathbf{Y}_{t+1}|\mathbf{Y}) d\mathbf{Y}_{t+1} = -\frac{1}{2} \mathbf{w}_t' \Sigma_{t+1|t} \mathbf{w}_t \\ \text{s.t.} \quad &\quad \mathbf{w}_t \in C(\mathbf{w}_t) = \left\{ \mathbf{w}_t | \mathbf{w}_t' \iota = 1 \right\} \end{aligned} \tag{10}$$

where $\Sigma_{t+1|t} = Var(\mathbf{Y}_{t+1}|D_t)$, is the covariance matrix under the predictive distribution. In the GMVP case the predictive is easily defined as

$$p(\mathbf{Y}_{t+1}|\mathbf{Y}) \propto \int_{\Sigma} p(\mathbf{Y}_{t+1}|\Sigma) p(\Sigma|\mathbf{Y}) d\Sigma \qquad \text{with} \qquad p(\Sigma|\mathbf{Y}) \propto p(\mathbf{Y}|\Sigma) p(\Sigma) \tag{11}$$

Given $\mathbf{Y}_t|\Sigma_t \sim N(0, \Sigma_t)$, it is quite easy to see that the key point is the definition of the prior for the covariance $p(\Sigma_t)$. Through $p(\Sigma_t)$ not only we define the predictive given the likelihood, but we can impose some exogenous identification structure, for instance from an asset pricing model, just as proposed in Pastor (2000) and Pastor and Stambaugh (2000) among the others. Finally, the resampling based optimal portfolio allocation, is briefly described in the appendix.

Given $\theta_t = \Sigma_t$, the outcome $\mathbf{w}_t = a(\Sigma_t)$ depends on the functional form of the weights, i.e. the action, and the covariance estimates, i.e. the state of nature. The von Neumann-Morgenstern rationality assumption in the Minimum-Variance framework assumes a rational investor chooses the pair functional form and covariance estimates to minimize the portfolio variance.

3. Expected Utility maximization: Constraining weights vs. Bayesian allocations.

Under the von Neumann-Morgenstern [NM] rationality assumption, the portfolio allocation problem, involves chosing the action, i.e. vector of weights, that maximizes the utility function averaged over state probabilities. Broadly speaking the NM rationality assumption allows us to *rank* the portfolio selection approaches, which provide the weights, based on their expected utility. In the minimum variance framework the representative investor is assumed to chosen the functional form and the covariance estimates that minimize portfolio risk exposure. The quantitative value of the expected utility is taken conditional to the action chosen and the covariance distribution, depending on the outcome $\hat{\mathbf{w}}_t = a(\hat{\Sigma}_t)$, with a the functional form of the weights and $\hat{\Sigma}_t$ the covariance estimate. In this section we point out that, in an expected utility maximization framework, constraining portfolio weights, is not "acting rationally" under the NM assumption. In other words, we provide some empirical findings that, the action (weights functional form) and the states (covariance estimates) implied by Jagannathan and Ma (2003), and its gross exposure generalizations, are not compatible with a rational expected utility maximization framework. This is done through two simulation examples, comparing the expected value of the utility function for each of the investigated portfolio selection strategies, reported below.

3.1. Simulation Design

The simulation framework generates a set of time-dependent covariance matrices. We used a RiskMetrics simulation methodology to impose heteroschedasticity getting a return dynamics closer to reality. Let us consider K simulation for T periods of N returns

$$\begin{aligned} \mathbf{\Sigma}^{(i)} &= \left(\Sigma_1^{(i)}, ..., \Sigma_T^{(i)}\right) \\ \mathbf{R}^{(i)} &= \left(R_1^{(i)}, ..., R_T^{(i)}\right) \qquad \text{for } i = 1, ..., K \end{aligned} \tag{12}$$

where $\Sigma_t^{(i)}$ is an $N \times N$ covariance matrix at time t and $R_t^{(i)}$ is an $N \times 1$ vector of returns. The i_{th} step is as follows

$$\begin{aligned} \Sigma_t^{(i)} &= \lambda \Sigma_{t-1}^{(i)} + (1-\lambda) R_t^{(i)} R_t^{(i)\prime} \\ R_t^{(i)} &= \Sigma_t^{(i)} \epsilon_t \qquad \text{with} \qquad \epsilon_t \sim NID\left(0, \mathbf{I}_K\right) \end{aligned} \tag{13}$$

such that

$$R_t^{(i)} \sim N\left(0, \Sigma_t^{(i)}\right) \tag{14}$$

with $\Sigma_0^{(i)}$ as initial value and $\lambda = 0.97$ for daily data. Notice $\Sigma_t^{(i)}$ differs across simulations. The benchmark index, i.e. the market, is constructed at every step t as an equally weighted average of the synthetic returns

$$R_{I,t}^{(i)} = \sum_{j=1}^{N} x_j R_{j,t}^{(i)},\ i = 1, ..., K \qquad \text{with} \qquad x_j = 1/N \tag{15}$$

The benchmark is used as target covariance structure in the Ledoit-Wolf shrinkage estimator, as well as to construct the factor model in (29). We simulate two years of daily data, $T = 504$ for $K = 20$ simulations. The initial $\Sigma_0^{(i)}$ is the historical covariance matrix of an N-dimensional randomly selected subset of stocks from the NYSE/AMEX. Some further description of the dataset is provided below. The simulation example is split in two parts: (a) we fix the portfolio and change the sample size, and (b) fixing the sample and changing the portfolio size. The aim is to investigate the effect of portfolio and sample size on the comparison of portfolio decision rules.

3.2. Portfolio strategies

In this section we describe the portfolio decision rules compared to constraining allocations. Since these models are familiar to most readers, we provide just a brief description of each one. The list of the model considered is reported in Table (1). The naive 1/N portfolio (ew) involves holding an equally weighted portfolio $\mathbf{w}_t^{ew} = 1/N$ at each time t. This strategy does not deal with neither any estimation nor optimization representing therefore a challenging benchmark in terms of, both estimation risk and parameter uncertainty.

[Insert Table 1 here]

The gross-exposure [GE] and the no-short sales strategies [JM] are respectively reported in Fan et al. (2008), DeMiguel et al. (2009a) and Jagannathan and Ma (2003). The JM action involves a constrained quadratic optimization, where the covariance estimate is from the usual MLE sample covariance estimation. The constraints are represented by non-negativity of portfolio weights and an upper bound diversification constraint. In the following simulation $\overline{\mathbf{w}} = 1$ meaning, we can invest max 100% of the portfolio in one asset. The GE portfolio decision rule is a generalization of the JM framework as suggested in DeMiguel et al. (2009a). Instead of no-short sales the GE involves some negative weights plugging an L_1-norm penalization on the minimum variance objective function. The purpose of the artificial constraints in Fan et al. (2008), DeMiguel et al. (2009a) and Jagannathan and Ma (2003), is stabilize the portfolio weights mitigating estimation risk through an artificial covariance regularization. The Ledoit-Wolf [LW] strategy is the unconstrained plug-in portfolio allocation in (7), plugging the shrinkage covariance estimates developed in Ledoit and Wolf (2003). The target covariance is the one-factor model covariance matrix. Yet, the RiskMetrics [RM] involves a closed form for the weights (the action) using an exponential weighted moving average estimates of the covariance matrix (the state parameter), which is then plugged in (7). The bootstrap portfolio [BU] allocation is inspired to Michaud (1998) generating a sample of weights plugging a resampled covariance in (7). Finally a simple benchmark Bayesian portfolio strategy is developed following a standard updating recursion. Let us suppose that the initial prior is

$$(\Sigma_0|D_0) \sim IW(b_0, \mathbf{S}_0) \tag{16}$$

with $D_t = (\mathbf{Y}_t, D_{t-1})$ the information set. The posterior at time *t-1* is defined as

- (a) posterior at time *t-1*:

$$(\Sigma_{t-1}|D_{t-1}) \sim IW(b_{t-1}, \mathbf{S}_{t-1}) \tag{17}$$

- (b) prior at time t is

$$(\Sigma_t|D_{t-1}) \sim IW(\delta b_t, \delta \mathbf{S}_{t-1}) \tag{18}$$

- (c) such that the posterior at time t becomes

$$(\Sigma_t|D_t) \sim IW(b_t, \mathbf{S}_t) \qquad \text{with} \qquad b_t = \delta b_{t-1} + 1,\ \mathbf{S}_t = \delta \mathbf{S}_{t-1} + \mathbf{Y}_t' \mathbf{Y}_t \tag{19}$$

- (d) given $\mathbf{Y}_t|\Sigma_t \sim N(0, \Sigma_t)$ the predictive $p(\mathbf{Y}_{t+1}|D_t)$ is

$$\mathbf{Y}_{t+1}|D_t \sim T(0, \mathbf{S}_t, b_t) \tag{20}$$

with $T(0, \mathbf{S}_t, b_t)$ a multivariate T-student distribution with $\mathbf{S}_t$ scale parameter and

b_t degrees of freedom.

From (20) we can define the predictive covariance matrix used in (4) as $\Sigma_{t+1|t} = var(\mathbf{Y}_{t+1}|D_t) = E(var(\mathbf{Y}_{t+1}|\Sigma)|D_t) + var(E(\mathbf{Y}_{t+1}|\Sigma)|D_t)$.

3.3. Simulation results: Out-of-sample wealth risk exposure minimization

In each simulation the risk exposure minimization portfolio performances are analyzed backtesting each of the portfolio strategies over six months of out-of-sample daily returns with daily rebalancing. Rolling sample estimation is used except for the Bayesian strategy and the RiskMetrics since they properly discount past information through a decay factor. Given a sample $M = 252$ we generate $T - M = 132$ out-of-sample portfolio returns, meaning, six months of trading day returns with daily rebalancing. We suppose the aim of the representative investor is to minimize the riskness of net wealth returns generated by each a self-financing minimum variance portfolio strategy. The realized gross-returns in the k_{th} simulation from the strategy $s = 1, ...S$ are defined as

$$\mathbf{r}^s_{k,t+1} = \mathbf{w}^{s\prime}_{k,t}\mathbf{r}_{k,t+1}, \quad t = M+1, ...T \tag{21}$$

with $\mathbf{r}_{k,t+1}$ the realized simulated asset returns at time *t+1*, and $\mathbf{w}^s_{k,t}$ the portfolio allocation from the strategy s, in simulation k, chosen ex-ante at time t. The wealth dynamics, net of the transaction costs tc, is written as

$$\mathbf{W}^s_{k,t+1} = \mathbf{W}^s_{k,t}(1 + \mathbf{r}^s_{k,t+1})\left(1 - tc \times \sum_{j=1}^{N} \left|w^s_{j,k,t+1} - w^s_{j,k,t^+}\right|\right) \tag{22}$$

where $w^s_{j,k,t+1}$ is the relative portfolio weight in asset j, at time *t+1*, under portfolio rule s, in simulation k, and w^s_{j,k,t^+} is the same weight right before rebalancing. For each simulation, the realized net return on wealth for strategy s is given by

$$\mathbf{r}^{s,\mathbf{W}}_{k,t+1} = \frac{\mathbf{W}^s_{k,t+1}}{\mathbf{W}^s_{k,t}} - 1 \tag{23}$$

The out-of-sample performance/riskness measure is essentially the standard deviation of (23) as a proxy of net return on wealth riskness. We set the proportional transaction costs tc equal to 50 basis point per transaction as reported in Balduzzi and Lynch (1999). In the first simulation M is fixed while portfolio sizes span from $N = 50$ to $N = 250$. We run $K = 20$ simulations. The results reported are the averages across the K simulations. Table (2) reports a first set of simulation results.

[Insert Table 2 here]

Risk Wealth represents the annualized standard deviation of (23). For $N = 50$ the BC rule reports a slightly lower risk exposure than the competing portfolio decision rules. The highest risk exposure corresponds to the naive EW strategy. This is consistent with the results in DeMiguel et al. (2009a). They showed that for a $N = 48$ dimensional portfolio the 1/N strategy underperforms most of the competitors in terms of risk exposure. The GE strategy reports the highest turnover. The same is true in the $N = 100$ case as well as the other portfolio sizes. The JM still reports the lowest turnover. However it suffers with lack of diversification since around 90% of the available assets are not used in the portfolio allocation. Yet, the BC reports the lowest wealth volatility, even though the JM is quite close. The latter, however, does not represents a financially reliable alternative since invests in 19 stocks out of 150. There is an evident lack of diversification. The same situation persists for $N = 200$ and is extreme for $N = 250$. Indeed the JM portfolio rule essentially invest in one stock without rebalancing. This is due to the enormous amount of estimation error in the input historical covariance matrix since $N/M \approx 1$, as suggested by Random Matrix theory standard arguments. With reference $N = 250$, the BC and the LW are almost equivalent. This is consistent with Ledoit and Wolf (2003), since when $N/M \approx 1$ is the situation where the shrinkage estimator is mainly justified. Recall that the BC does not imposed any particular on the covariance recursion. Interestgly the naive EW portfolio rule reports the lowest turnover. Yet, this is clearly due to the absence of estimation in the naive strategy. There is neither estimation error nor optimization algorithms, therefore, the portfolio weights resulting are more stable through rebalancing. This is consistent with DeMiguel et al. (2009a). Overall the BC reduces risk exposure (expected utility maximization) fairly better than the other strategies.

The second simulation example sets the portfolio size to $N = 100$ changing the insample length span from $M = 120$ to $M = 210$. We do not report the value for $M = 250$ since overlaps the aforementioned statistics for the first simulation example. Table (3) reports the relative results.

[Insert Table 3 here]

The JM still suffers with diversification issues leading essentially to financial irrelevant/corner portfolio allocations. The EW reports overall the lowest turnover. Yet, this is due to the absence of estimation and optimization characterizes the naive strategy, which leads to highly stable weights, consistently with the findings in DeMiguel et al. (2009a). Overall the BC reports the lowest risk exposure through insample sizes. However the LW portfolio rules is fairly comparable, delivering financially stable and diversified allocations. This suggests the relevance of imposing structure on the covariance estimation. Let us recall that the BC does not consider any particular structure in the covariance estimates. On the other hand the LW incorporates a very simple dynamics through the rolling sample estimation procedure.

4. Decision Rules, Admissibility and a common Bayesian framework

Section 4 shows that the action and covariance estimate implied by JM and GE are not generally preferred in an expected utility maximization framework. These strategies therefore are essentially incompatible with the standard von Neumann-Morgenstern rationality assumption, meaning, they are incompatible to an expected utility maximization framework since apparently constantly dominated by alternative, unconstrained, actions and covariance estimates. However, the quantitative value assigned to the utility from the outcome $\mathbf{w}_t = a(\Sigma)$, is averaged over different distribution assumptions for each decision rule. Indeed, while the predictive distribution is used in the Bayesian framework, the likelihood is the distribution under which we average the utility from the decision outcome for JM, GE and the others, i.e. Ledoit-Wolf, RiskMetrics. Therefore, one could argue the superior performance of the Bayesian approach in terms of out-of-sample wealth risk exposure, can be essentially driven by the forecasting property nested in the predictive distribution. This is not the case in the likelihood case since the two-steps standard approach assumes past information is all we need to forecast the future covariance structure. Then is not really understandable if the superior out-of-sample risk reduction comes from either avoiding artificial constraints or using the correct action and state estimates. Imposing no-short sales on the Bayesian portfolio does not solve the issue since (a) we assume there are not market frictions is terms of short-sales, and (2) the conceptual difference in the distribution used to average out the utility are still different. What we propose is a common Bayesian interpretation of the frequentist approaches used as alternative to the Bayesian one. By using the results in Jagannathan and Ma (2003) we map the constraining portfolio rules in the unconstrained case, defined a proper shrinking bayesian covariance estimator. The aim is to have the same action, functional form of the weights, that is the closed form, such that we can discriminate among different portfolio decision rules, which are reconduced to prior specification. Given the functional form of the weights we can interpret each prior specification as the recipe turning the data into the action, meaning the covariance estimates. In other words, defining a common bayesian interpretation of the JM and GE rules, other than the others, we can essentially compare among different decision rules. The purpose is then check for admissibility of each of them. This is done empirically discriminating decision rules on the basis of a proper risk function. We do not compare analytically the prior distributions since, given the likelihood, each posterior is optimal with respect to its prior. The predictive distribution is the probability of having a future observation under the likelihood averaged over the posterior distribution. Therefore, since the likelihood is given, and each posterior is optimal conditional on the prior, is hard to analytically investigate admissibility of the prior relative to JM rather than GE, with respect to a benchmark Bayesian estimator. We assign a quantitative score to each portfolio decision rule on the basis of a risk function $R(\tilde{\mathbf{w}}, \hat{\mathbf{w}})$, where $\tilde{\mathbf{w}}$ is the *Oracle* decision rule, while $\hat{\mathbf{w}}$ is the portfolio decision rule chosen by the representative investor. Generally speaking a decision rule $\mathbf{w}_0(\theta)$ is said to be inadmissible, if there exists another rule $\mathbf{w}_1(\theta)$ such that $R(\tilde{\mathbf{w}}, \mathbf{w}_1) \leq R(\tilde{\mathbf{w}}, \mathbf{w}_0)$ for all $\theta \in \Theta$, and $R(\tilde{\mathbf{w}}, \mathbf{w}_1) < R(\tilde{\mathbf{w}}, \mathbf{w}_0)$ for some $\theta \in \Theta$, where $R(\tilde{\mathbf{w}}, \hat{\mathbf{w}})$ is a general risk function. The risk function we refer to for the minimum variance portfolio decision rule is formalized in Proposition 2.

Proposition 2. *Let us consider zero-mean multivariate normal returns* $\mathbf{Y}_t|\Sigma \sim N(0, \Sigma)$, *such that* $\theta = \Sigma$, *and a standard quadratic mean-variance utility function. The action* $a \in \mathcal{A}$, *is a function mapping the data into the outcome* $\hat{\mathbf{w}}_t = a(\hat{\Sigma})$. *The regret loss function is defined as*

$$\begin{aligned} L(\tilde{\mathbf{w}}, \hat{\mathbf{w}}_t) = L_u(\tilde{\mathbf{w}}, \hat{\mathbf{w}}_t) - \inf_{\hat{\mathbf{w}}_t \in \mathcal{A}} L_u(\tilde{\mathbf{w}}, \hat{\mathbf{w}}_t) \;&\equiv\; \sup_{\mathbf{w}} U(\mathbf{w}'\mathbf{Y}_{t+1}) - U(\hat{\mathbf{w}}_t'\mathbf{Y}_{t+1}) \\ &= \; U(\tilde{\mathbf{w}}'\mathbf{Y}_{t+1}) - U(\hat{\mathbf{w}}_t'\mathbf{Y}_{t+1}) \end{aligned} \tag{24}$$

where $\tilde{\mathbf{w}} = a(\Sigma)$, *i.e. the* Oracle *optimal portfolio decision rule and* $\hat{\mathbf{w}}_t = a(\hat{\Sigma})$ *the portfolio decision rule at time* t *based on a covariance estimates. The associated risk function is defined as the expected value under the predictive distribution of the regret loss function*

$$\begin{aligned} R(\tilde{\mathbf{w}}, \hat{\mathbf{w}}_t) &= \int_{\mathcal{Y}} L(\tilde{\mathbf{w}}, \hat{\mathbf{w}}_t) p(\mathbf{Y}_{t+1}|\mathbf{Y}) d\mathbf{Y}_{t+1} \\ &\equiv U(\tilde{\mathbf{w}}'\mathbf{Y}_{t+1}) - \int_{\mathcal{Y}} U(\hat{\mathbf{w}}_t'\mathbf{Y}_{t+1}) p(\mathbf{Y}_{t+1}|\mathbf{Y}) d\mathbf{Y}_{t+1} \end{aligned} \tag{25}$$

The portfolio allocation is defined minimizing (25). *This is empirically equivalent to maximize the expected value of the utility function under the predictive distribution. A textbook proof is provided in the appendix.*

According to (25), choosing the optimal decision rule, minimizing $R(\tilde{\mathbf{w}}, \hat{\mathbf{w}}_t)$ to get the outcome $\hat{\mathbf{w}}_t$ turns out to be equivalent of maximizing the expected value of the utility function under the predictive distribution. Indeed, from a purely optimization point of view, $U(\tilde{\mathbf{w}}'\mathbf{Y}_{t+1})$ essentially clears away. However, the purpose is to discriminate among the decision rules out-of-sample. Therefore, even though the decision rules is chosen ex-ante, its admissibility must be evaluated ex-post. Just as in Pastor and Stambaugh (2000), Kandel and Stambaugh (1996) and McCulloch and Rossi (1990), we define a performance metrics we call the *Actual Bayes Risk* [ABR]. By using Proposition 2 and the assumption of short sales allowed, we could define the Actual Bayes Risk, plugging $\mathbf{w}_t^* = \mathbf{w}(\Sigma_{t+1|t}) = (\Sigma_{t+1|t}^{-1}\iota)/(\iota'\Sigma_{t+1|t}^{-1}\iota)$ in (25) as follows[3]

$$R(\tilde{\mathbf{w}}, \mathbf{w}_t^*) = U(\tilde{\mathbf{w}}'\mathbf{Y}_{t+1}) - U(\mathbf{w}_t^{*'}\mathbf{Y}_{t+1}) = \mathbf{w}_t^{*'}\Sigma_{t+1}\mathbf{w}_t^* - \frac{1}{\iota'\Sigma_{t+1}^{-1}\iota} \tag{26}$$

with $\tilde{\mathbf{w}}$ the unknown *Oracle* optimal portfolio allocation, $\iota = \{1, ..., 1\}$ an N-dimensional vector of ones and Σ_{t+1} the observed, ex-post, covariance matrix. In other words, (26) represents the difference in terms of realized utility between the oracle and the action taken under the risk minimization criteria in (25). We checked out-of-sample the admissibility comparing the value assigned to (26) relative to each of the outcomes $\hat{\mathbf{w}}_t$ coming from different decision rules. Economically speaking is the gain perceived by an investor who is forced to accept the $\hat{\mathbf{w}}(\hat{\Sigma})$ with respect to the true unobservable decision portfolio rule.

4.1. A Common Bayesian Framework

Proposition 2 shows that the optimal portfolio decision rule reduce to maximize the expected value of the utility function under the predictive distribution. We need to define common Bayesian framework, in order to have a predictive distribution to compute the expected utility in (25). Given the likelihood $\mathbf{Y}_t|\Sigma \sim N(0, \Sigma)$, the predictive density is a function of the prior $p(\Sigma)$. Each strategy can be reconduced to a differentr prior specification. From Proposition 1 we can argue that

Proposition 3. *Let $\tilde{\Sigma} = \hat{\Sigma} + \mathbf{J}$, be the Jagannathan and Ma (2003) regularized covariance estimator with $\hat{\Sigma}$ the historical covariance estimates. Suppose now that $p(\Sigma) = IW(\nu_0, \mathbf{J})$ and $\mathbf{Y}_t|\Sigma \sim N(0, \Sigma)$ such that, we can generally interpret $\tilde{\Sigma}$ as the posterior expectation*

[3]The order on the RHS is inverted since the utility function in the minimum variance case is negative and the inf is reached with Σ known.

of an inverse-wishart distribution such that

$$E[\Sigma|\mathbf{Y}] = \alpha\tilde{\mathbf{J}} + (1-\alpha)\hat{\Sigma} \tag{27}$$

with $\alpha \in (0,1)$*,* $\tilde{\mathbf{J}} = \mathbf{J}/(\nu_0 - N - 1)$*, and* α *decreasing in* T *such that* $\lim_{T\to\infty} E[\Sigma|D_t] \equiv \lim_{T\to\infty} \hat{\Sigma} = \Sigma$*, consistently with the shrinkage intensity in Ledoit and Wolf (2004).*

Proposition 3 allows the JM and GE being interpreted as different prior specifications, leading to different portfolio decision rules. In the JM case, for instance, the prior covariance matrix $\mathbf{J}$ comes from the lagrange multipliers of a standard quadratic programming algorithm imposing non-negativity constraints just as suggested in Jagannathan and Ma (2003). In our case the predictive is

$$p(\mathbf{Y}_{t+1}|\mathbf{Y}) = \int_{\mathcal{Y}} p(\mathbf{Y}_{t+1}|\mathbf{Y},\Sigma_t)p(\Sigma_t|\mathbf{Y})d\Sigma_t \quad \text{with} \quad p(\Sigma_t|\mathbf{Y}) = IW(\nu_0+T, \mathbf{J}+\mathbf{Y}'\mathbf{Y}) \tag{28}$$

Then given $\mathbf{Y}_{t+1}|\mathbf{Y},\Sigma_t \sim N(0,\Sigma_t)$ the predictive is $\mathbf{Y}_{t+1}|\mathbf{Y} \sim T(0,\Sigma_n,\nu_n)$ with $\Sigma_n = \mathbf{J}+\mathbf{Y}'\mathbf{Y}$ and $\nu_n = \nu_0 + T$. As aforementioned the specification of $\mathbf{J}$ in JM and GE comes from the lagrange multipliers as reported in Proposition 1. On the other hand, for the LW case, the prior covariance matrix is essentially the one factor covariance structure following Ledoit and Wolf (2003). The RiskMetrics case is reconduced to a standard sequential updating scheme of an inverse wishart distribution as in (20)-(16). We check for admissibility of constraining portfolio weights rules both through simulation examples and an empirical analysis on a real dataset. The simulation is run essentially to allow the Oracle utility being observable.

5. A Dynamic Bayesian shrinkage covariance estimator

A Natural approach of using financial models in decision making is through a bayesian framework. An asset pricing model can be used as a prior reference around which the decision maker makes the investment choice. The approach developed in this section refers to a classical linear asset pricing model to build a shrinkage prior combined with the historical information in the spirit of Ledoit and Wolf (2004) and Ledoit and Wolf (2003). The prior is updated using a standarda inverse-wishart updating scheme and is estimated through a Dynamic Linear Model [DLM]. The standard asset pricing model is from a *Seemingly Unrelated Regression* [SURE], defined through the quadruple $[\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t]$. They represents respectively, the $(1 \times L)$ covariates, the $(L \times L)$ state evolution matrix, the $(N \times N)$ observational variance matrix and finally the $((N \times L) \times (N \times L))$ evolution variance matrix, with K the number of factors and $L = K + 1$. It is sensible to assume that the intercepts and the slopes are correlated across the N stocks. Let us consider $\mathbf{Y}_t = \mathbf{r}_t - \mu$ with $\mathbf{r}_t = \mathbf{R}_t - \mathbf{1}_N r_f$ the returns in excess of the risk-free rate r_f, and $\mathbf{X}_t$ the K-dimensional vector of factor returns. The prior bulding model equations

$$\begin{aligned} \mathbf{Y}_t &= (\mathbf{F}_t \otimes \mathbf{I}_N)\,\theta_t + v_t \qquad & v_t \sim NID(0, \mathbf{V}_t) \\ \theta_t &= (\mathbf{G}_t \otimes \mathbf{I}_N)\,\theta_{t-1} + w_t \qquad & w_t \sim NID(0, \mathbf{W}_t) \end{aligned} \tag{29}$$

where $\mathbf{F}_t = [1, \mathbf{X}_t]$, $\mathbf{G}_t = \mathbf{I}_L$, $\theta = [\alpha', \beta']'$ and $\mathbf{W}_t$ a $((N \times L) \times (N \times L))$ block diagonal matrix. Yet, θ_t is an $(N \times L)$-dimensional state vector. Let us suppose $D_t = \{\mathbf{Y}_t, D_{t-1}\}$ and an initial prior at $t = 0$ as multivariate normal

$$(\theta_0 | D_0) \sim N(\mathbf{m}_0, \mathbf{C}_0)$$

for some hyperparameters $\mathbf{m}_0$ and $\mathbf{C}_0$. The predictive covariance matrix in (4) is found using the following filtering recursion:

- (a) Posterior at time $t-1$:
  For some mean $\mathbf{m}_{t-1}$ and covariance matrix $\mathbf{C}_{t-1}$

$$(\theta_{t-1} | D_{t-1}) \sim N(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}) \tag{30}$$

- (b) Prior at time t:

$$\begin{aligned} &(\theta_t | D_{t-1}) \sim N(\mathbf{a}_t, \mathbf{R}_t) \\ &\text{where} \quad \mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1} \quad \text{and} \quad \mathbf{R}_t = \mathbf{P}_t + \mathbf{W}_t \\ &\text{with} \quad \mathbf{P}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' \quad \text{and} \quad \mathbf{W}_t = (\delta^{-1} - 1)\mathbf{P}_t \end{aligned}$$

- (c) One-step forecast of the asset pricing model and the portfolio covariance prior

$$\begin{aligned} &(\mathbf{Y}_t | D_{t-1}) \sim N(\mathbf{f}_t, \mathbf{Q}_t) \\ &\text{where} \quad \mathbf{f}_t = \mathbf{F}_t \mathbf{a}_t \quad \text{and} \quad \mathbf{V}_t = \delta \mathbf{V}_{t-1} + diag(\mathbf{e}_{t-1} \mathbf{e}_{t-1}') \\ &\text{with} \quad \mathbf{Q}_t = \mathbf{F}_t \mathbf{R}_t \mathbf{F}_t' + \mathbf{V}_t \end{aligned}$$

Combining the information from (31) and the historical information we can build a shrinkage prior as

$$(\Sigma_t|D_{t-1}) \sim IW(\delta b_{t-1}, \delta(\omega \mathbf{S}_{t-1} + (1-\omega)\mathbf{Q}_t)) \quad \text{with} \quad \omega \in (0,1) \tag{31}$$

for some initial $\mathbf{S}_0$.

- (d) Posterior at time t for the asset pricing model and the covariance posterior

$$\begin{aligned}
&(\theta_t|D_t) \sim N(\mathbf{m}_t, \mathbf{C}_t) \\
&\text{with} \qquad \mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t\mathbf{e}_t \qquad \text{and} \qquad \mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t\mathbf{Q}_t\mathbf{A}_t' + \mathbf{V}_t \\
&\text{where} \qquad \mathbf{A}_t = \mathbf{R}_t\mathbf{F}_t\mathbf{Q}_t^{-1} \qquad \text{and} \qquad \mathbf{e}_t = \mathbf{Y}_t - \mathbf{f}_t
\end{aligned}$$

Therefore the posterior of the portfolio covariance estimator can be defined as

$$\begin{aligned}
&(\Sigma_t|D_t) \sim IW(b_t, \mathbf{S}_t) \\
&\text{with} \qquad \mathbf{S}_t = \delta(\omega\mathbf{S}_{t-1} + (1-\omega)\mathbf{Q}_t) + \mathbf{Y}_t\mathbf{Y}_t' \qquad b_t = \delta b_{t-1} + 1
\end{aligned}$$

where $\delta \in (0,1)$ represents a discount factor by which we impose time-varying dynamics in the system evolution. Given the likelihood $\mathbf{Y}_t|\Sigma_t \sim N(0, \Sigma_t)$, we can define the predictive by using (32) as

$$\mathbf{Y}_{t+1}|D_t \sim T(0, \mathbf{S}_t, b_t) \tag{32}$$

As we can see, (31) does not represent a double use of the data. Indeed, the predictive $\mathbf{Q}_t$ is essentially a function of the covariates at time $t-1$ plus the observation variance $\mathbf{V}_{t-1}$, while $\mathbf{S}_{t-1}$ is a function of $\mathbf{Y}_{t-1}$ on its own, regardless covariates. The same is true for (32). In other words, the agent intervene in the system modifying at each step t the prior according to the information coming from the equilibrium asset pricing model.

6. A simulation study of Admissibility: Constraining weights vs Bayesian shrinkage estimator.

The portfolio strategies implemented are those reported in Table (1). Instead of the classic Bayesian portfolio allocation BC we use the shrinkage version reported in Section 5. Rolling sample estimation is used except for the Bayesian strategy and the RiskMetrics methodology since they properly discount past information through a decay factor $\delta \in (0,1)$. In each of the $K = 20$ simulation results the out-of-sample performance metrics is the Actual Bayes Risk. The performances are analyzed backtesting each of the portfolio strategies over six months of out-of-sample daily returns with daily rebalancing. The out-of-sample period is $T - M = 132$ daily returns, with $M = 252$ fixed in the first set of simulation results, and spanning from $M = 100$ to $M = 210$ in the second set of simulations. The Actual Bayes Risk for the k_{th} simulation under the $s = 1, ..., S$ strategy is defined as

$$ABR_k^s = U(\tilde{\mathbf{w}}_k{}' \mathbf{Y}_{k,t+1}) - U(\mathbf{w}_{k,t}^{s*\prime} \mathbf{Y}_{k,t+1}) \tag{33}$$

where $\tilde{\mathbf{w}}_k$ is the *Oracle* decision rules in the k_{th} simulation, $\mathbf{w}_{k,t}^{s*}$ is the outcome of the s_{th} strategy on the k_{th} simulation and $\mathbf{Y}_{k,t+1}$ is the zero-mean simulated return. The results reported are the averages of (33) across the K simulations. The first simulation example runs different portfolio sizes for a fixed insample information set. This is done to point out the portfolio size effect on the decision rule admissibility, especially as the ratio $N/M \approx 1$. According to basic random matrix theory arguments this represents the situation in which noise dominates in the covariance estimates and a structured estimator is needed. Table (4) reports the results.

[Insert Table 4 here]

With reference to $N = 50$, the RM portfolio decision rule reaches the lowest level of ABR. Surprisingly the EW naive portfolio strategy is sensibly outperformed. This somehow contradicts, from a decision theoretical point of view, the results in DeMiguel et al. (2009b) in which a sensible superior performance of the $1/N$ strategy is well-documented. The DB is consistently outperformed by the others except for EW and GE. Let us recall that we set $\omega = 0.3$ in the DB computation, and that RM essentially represents the DB with $\omega = 1$, i.e. disregard the asset pricing model information. Therefore, the inferior performance for $N = 50$ is justified by the dominating relevance of historical information, given the low N/M ratio. Considering $\omega = 0.3$ we force the agent to consider the information from the asset pricing model which turn out to be irrelevant, even hurting, if a large amount of historical information is available. The big picture changes for $N = 100$, when the ratio N/M slighlty increases. Here the DB outcome has the lowest ABR, while all of the others portfolio decision rules increases the ABR except for the EW. The latter, although has the highest ABR, obviously marginally improves since do not involved neither estimation nor computation issues. The level of daily turnover and the number of long positions are fairly comparable across the strategies without sensible extremes. The same kind of results persists for $N = 150$ and $N = 200$. Surprisingly for $N = 250$ both the JM and the GE portfolio decision rules reaches the higher ABR. The reason is that when the ratio $N/M \approx 1$ the quadratic optimization programmin they are based on to get the lagrange multipliers do not find an acceptable solution, as we can also see from the very high level of daily Turnover. The DB still gets the lowest ABR, which is however, quite comparable with the RM. The latter however, involves five times Turnover, meaning, is a five times more costly portfolio strategy. Overall, both the JM and the GE, i.e. the constraining portfolio weights, seems not be admissible since constantly dominated, in terms if ABR, by the others. In this sense the RM and the DB reaches the lowest ABR. Let us recall they represents essentially the same Bayesian recursion with a different value of $\omega \in (0, 1)$ in (32). In order to have a clearer understading Figure (1) reports the Relative Economic Loss [REL], consistently with Jorion (1986). The latter is defined as the average across the K simulations of the relative change in the ABR as follows

$$REL_s^k = \frac{U(\tilde{\mathbf{w}}_k{}'\mathbf{Y}_{k,t+1}) - U(\mathbf{w}_{k,t}^{s*\prime}\mathbf{Y}_{k,t+1})}{U(\tilde{\mathbf{w}}_k{}'\mathbf{Y}_{k,t+1})} \tag{34}$$

It can be interpreted as the percentage economic loss perceived by an investor who is forced to accept the portfolio selection rule $\mathbf{w}_{k,t}^{s*}$ instead of the unknown Oracle $\tilde{\mathbf{w}}_k$. It is nonnegative by construction and, economically speaking, the lower is the better is the s_{th} portfolio decision rule. Yet, (34) is a positive transformation of (33), being useful to check for admissibility.

[Insert Figure 1 here]

As we can see from Figure (1), all of the portfolio decision rules converges as the ratio $N/M \to 0$, meaning the noise in the covariance estimator sensibly reduces. However, the DB strategy consistently outperform the others from $N = 250$ to $N = 100$ being essentially equivalent in $N = 50$ with respect to RM, LW and BU. This picture provides some further insight on the inadmissibility of constraining weights as “equivalently optimal” allocation rules.

The second simulation example runs for fixed portfolio size $N = 100$ changing the insample from $M = 130$ to $M = 210$. As in the first simulation example, the portfolio

ABRs are analyzed backtesting the decision rules over six months of out-of-sample daily returns with daily rebalancing. As further metrics we report the daily percentage Turnover and the number of Long position as diversification proxy. Table (5) reports the performance statistics.

[Insert Table 5 here]

With reference to $M = 130$, the DB reaches a sensibly lower ABR. Although has a fairly low Turnover, the EW strategy reaches an high ABR, as reported in the previous simulation example. Even though the DB outperforms the others, also the RM and the LW still performs pretty well. The results are qualitatively the same in $M = 170$. Here the DB reaches both a lower Turnover and the lowest ABR. The level of diversification is almost the same across the strategies. Finally for $M = 210$ both the JM and GE improves, together with BU. This is because the ratio N/M decreases, increasing the relevance of historical information in the covariance estimates since noise decreases. The DB portfolio decision rule still has the lowest ABR. As for the first simulation example Figure (2) reports the REL averaged across simulations as from (34).

[Insert Figure 2 here]

As we can see, the DB constantyl outperform the others portfolio decision rules in terms of PEL, being around 5 across sample sizes. The EW shows the highest REL meaning, the investor obliged to choose the $1/N$ strategy against the Oracle incures in the highest relative economic loss. The GE get a REL qualitatively as high as the EW portfolio decision rule being around 30 across sample sizes. With reference to JM, LW, BU and RM, they tend to converge after $M = 170$ the ratio N/M being closer to zero, reducing the noise in the covariance estimates. As a whole the DB shows the lowest Relative Economic Loss.

7. An Empirical Analysis on the NYSE/AMEX

Whether portfolio constraints helps or hurts for risk reduction in Large Portfolios is essentially an empirical question. In this section we investigate the effect of constraining weights as opposed to use alternative portfolio allocation strategies on a real dataset. Our goal is to compare out-of-sample performances on large portfolios for each of the decision rules reported in Table (1). The optimal outcome $\mathbf{w}_t = a(\Sigma_{t+1|t})$, for each decision rule is set in the common Bayesian framework developed in Section 6. However, on the real dataset, the *Oracle* is not assumed to be observable, therefore, the action taken by the representative agent essentially aims to maximize the expected value of the utility function under the predictive distribution. The latter can be defined for all of the portfolio strategies exploiting Proposition 3. The dataset consists of all the common stocks traded on NYSE and AMEX, with stock price greater than five dollars and non trivial market capitalization, keeping a total of 2246 shares. Data are taken from CRSP. In order to construct the linear asset pricing model we use a Value-Weighted Index from CRSP containing all the stocks in the NYSE/AMEX as proxy for the market portfolio. Yet, data are from CRSP, while the risk-free daily rate is taken from Ken French's website. The study period is from October 29, 2007 to December 31, 2010 with daily data. Our analysis relies on rolling sample estimation except for the Bayesian and the RiskMetrics approaches since they properyl discount past information through a decay factor $\delta \in (0, 1)$. To avoid any particular stock picking methodology the stock selection is made randomly for each of the portfolio size considered, then kept constant throughout out-of-sample. The portfolio sizes are $N = [50, 200, 400]$, and for each size several performance measures are considered. Given an insample period of two years of daily data $M = 504$ we generate $T - M = 132$ of out-of-sample portfolio returns, meaning six-month of daily returns, for each of the portfolio strategies reported in Table (1). The first set of metrics is based on the realized gross-return for each of the $s = 1, ...S$ strategies defined as

$$\mathbf{r}_{s,t} = \mathbf{w}'_{s,t}\mathbf{r}_{t+1}, \quad t = M + 1, ...T \tag{35}$$

with $\mathbf{r}_{t+1}$ the realized simulated asset returns at time *t+1*, and $\mathbf{w}_{s,t}$ the outcome from the s_{th} decision rule, chosen on the basis of expected utility maximization at time t. From the realized gross-return we can define the wealth dynamics, net of the transaction costs tc as

$$\mathbf{W}_{s,t+1} = \mathbf{W}_{s,t}(1 + \mathbf{r}_{s,t+1}) \left(1 - tc \times \sum_{j=1}^{N} \left| w^j_{s,t+1} - w^j_{s,t^+} \right| \right) \tag{36}$$

where $w^j_{s,t+1}$ the relative portfolio weight in asset j, at time *t+1*, under portfolio rule s, and w^j_{s,t^+} is the same weight right before rebalancing. The realized net return on wealth for strategy s is then given by

$$\mathbf{r}^{\mathbf{W}}_{s,t+1} = \frac{\mathbf{W}_{s,t+1}}{\mathbf{W}_{s,t}} - 1 \tag{37}$$

From (37) the first two performance out-of-sample metrics are the Expected Net Wealth as

$$\mu^{\mathbf{W}}_s = \frac{1}{T-M}\sum_{t=1}^{T-M} \mathbf{r}^{\mathbf{W}}_{s,t} \tag{38}$$

and the out-of-sample Wealth Risk Exposure defined as

$$\sigma^{\mathbf{W}}_s = \sqrt{\frac{1}{T-M}\sum_{t=1}^{T-M} \left(\mathbf{r}^{\mathbf{W}}_{s,t} - \mu^{\mathbf{W}}_s\right)^2} \tag{39}$$

Then the out-of-sample wealth sharpe ratio out-of-sample for the strategy s is simply defined as

$$SR^{\mathbf{W}}_s = \frac{\mu^{\mathbf{W}}_s}{\sigma^{\mathbf{W}}_s} \tag{40}$$

Now, (38), (39) and (40) represent the first set of performance measures to evaluate portfolio decision rules in Table (1). They are strictly related to economic performances of each strategy. The transaction costs tc are equal to 50 basis points as reported in Balduzzi and Lynch (1999). The second set of out-of-sample metrics is more related to portfolio stability and diversification. To get of sense of the amount of trading involved by each strategy we compute the Turnover as

$$TO_s = \frac{1}{T-M}\sum_{t=1}^{T-M}\sum_{j=1}^{N}\left(\left|w^j_{s,t+1} - w^j_{s,t^+}\right|\right) \tag{41}$$

This quantity can be interpeted as the average percentage of wealth traded in each period. We report the absolute value of turnover for each of the S strategies. Finally we report the number of highest and the lowest weights together with the number of long positions defined as

$$Long_s = \sum_{j=1}^{N} \mathbf{1}\left\{w^j_{s,t}\right\} \qquad \text{with} \qquad \mathbf{1}\left\{w^j_{s,t}\right\} = \begin{cases} 1 & \text{if } w^j_{s,t} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{42}$$

All of the performance measures are measured daily then annualized. The empirical results are reported in Table (6). With reference to $N = 50$ the first row reports the mean wealth return. Surprisingly all of the decision rules, except DB and RM , gets negative returns. In particular the $1/N$ naive portfolio strategy has a highly negative annual return. The highest positive return is from the DB strategy.

[Insert Table 6 here]

Let us recall that the portfolio selection is made randomly for each portfolio size and the stocks selected are the same across the different strategies. The RM reports the lowest wealth annualized risk exposure, even though the DB fairly behaves. The RM essentially impose dynamics on the covariance estimates, while in the DB we impose both dynamics and structure through the shrinkage like sequential prior specification depicted in Section 5. The outperformance of RM therefore provides some evidence on how dynamics dominates structure if the ratio N/M is appreciably low. This is theoretically confirmed considering that RM represents DB with $\omega = 1$, meaning, disregarding the sequential shrinking prior. All of the others decision rules underperform even though they report fairly good results. This is true for the LW, the JM and the BU case. The third row reports the Sharpe Ratio as computed in (40). As we can see, even though the DB is slightly riskier than the RM reports a fairly higher annualized sharpe ratio because of higher positive annualized returns. This can be better understand from Figure (3).

[Insert Figure 3 here]

The DB seems to be the only decision rule gets appreciably positive cumulative wealth out-of-sample. On the other hand the EW strategy, although performs quite well till $T - M = 80$ sensibly drops thereafter. The BU and the LW show almost the same path for the cumulative wealth, being however less than one for most of the testing sample. The out-of-sample Turnover, is almost the same across the decision rules except for DB and the EW. As a matter of fact, the $1/N$ strategy reports a fairly lower Turnover on average. This is in line with DeMiguel et al. (2009b) and DeMiguel et al. (2009a), in which the authors reported the $1/N$ strategy as the more stable decision rule with respect to other classical benchmark strategies. The higher stability of the naive strategy is essentially due to the fact that does not involve neither estimation nor optimization. The last three for $N = 50$ report respectively the *Min* and the *Max* portfolio weights together with the number of *Long* positions. Since the number of stocks considered is fairly low, *Long* does not differ sensibly across the different portfolio rules. On the other hand, while the EW by construction does not report short positions, all of the strategies report fairly extreme positions, in light of the portfolio size. The RM shows an average short minimum weight of -38% which is fairly extreme considering the agent could potentially invest in 50 stocks. The same is true for the average long positions across strategies which is almost always around 60%. However, DB shows a different characteristic. The portfolio composition report very small average minimum short positions and financially reasonable average long positions. This essentially provides evidences in support of the financial reliability of the DB as, not only the most profitable risk-adjusted strategy, but also the more financially reliable. The whole picture is almost the same for $N = 200$. The first row reports the expected net return on wealth as from (38). As for the previous case the DB decision rule shows the highest value. Again, both the RM and the DB are the only strategies reporting a positive average net return on wealth. The LW reports the lowest wealth risk exposure, together with the JM decision rule. Remember that they essentially both represent a shrinkage covariance estimation decreasing the high eigenvalues due to estimation error and increasing the low eigenvalues due to sampling error. The RM is outperformed by the DB meaning, as the ratio N/M increases getting close to one, the structure imposed by the shrinking prior starts to play a relevant role. The third row report the Sharpe Ratio. The DB reports the highest annualized Sharpe Ratio even though the annualized risk exposure is not the lowest. This is because the highest expected net return on wealth. Overall, yet, surprisingly all of the portfolio decision rules has a negative mean return on wealth except for RM and DB. This can be better understood looking at the out-of-sample path of the cumulative portfolio strategy returns reported in Figure (4).

[Insert Figure 4 here]

Still the DB reports a positive cumulative wealth dynamics considering the one dollar investment at the beginning of the testing period. The wealth dynamics showed by the naive $1/N$ strategy is fairly unstable reporting an high positive performance till $T-M = 90$ then dropping thereafter. The JM, together with the others, except RM and DB, reports a generally negative cumulative wealth considering the initial supposed investment. Still the LW and the BU have a testing period cumulative wealth very close to each other. With reference to Turnover, the RM turns out to increases marginally the daily amount of trading needed to get the net return on wealth. The DB still reports the lowest Turnover and the naive $1/N$ naive decision rule still have a fairly low level of Turnover. Yet, this depends on the absence of any estimation/optimization procedure involved in the $1/N$ strategy. The portfolio composition of the RM reports yet extreme average negative position with *Min* equal to -18%, which is fairly extreme considering the portfolio size. As for the $N = 50$ the DB reports the most financially plausible portfolio composition with a lowest short position below one percent and an highest average long position around 19%. The average mass of positive weights is fairly comparable across strategies, even though the DB seems to show some diversification benefit other than more financially relevant portfolio weights. The EW has only long positions by construction. Finally the picture becomes slightly more in favour of the DB portfolio rule for the $N = 400$ case. All of the portfolio strategies report a negative mean net return on wealth except for DB which get an annualized 1.1%. Yet, the DB portfolio decision rule has the lowest Wealth Risk Exposure as from (39). As for the other portfolio sizes, the JM and the LW have a fairly comparable level of risk exposure. They indeed represents a conceptually comparable shrinkage covariance estimator, as proved in Jagannathan and Ma (2003). This parallel behavior seems to support their results as a whole. Obviously, from the expected net return on wealth the DB is the only portfolio decision rule gets a positive annualized Sharpe Ratio, which is however fairly low being around 0.35. The level of Turnover marginallt increases for the DB strategy begin now almost in line with the $1/N$ naive portfolio decision rule. This, yet, is in line with DeMiguel et al. (2009a) and DeMiguel et al. (2009b). The rationale for the *Min* and *Max* positions is repeated also for $N = 400$. Indeed, the DB portfolio decision rule report very small average short positions as opposed to the fairly extreme negative positions of the other decision rules. The same is true for *Max*, since the DB reports fairly lower average long positions. Finally the number of *Long* positions marginally increases for the DB with respect to the others portfolio decision rules, meaning as the portfolio size increases, we get a marginally higher diversification benefit.

Summarizing the DB strategy seems to outperform as a whole the competing portfolio decision rules. This is especially true for Large Portfolios, i.e. $N = 400$. This findings essentially support the hypothesis that, when the ratio $N/M \to 1$, consider jointly dynamics and structure in the covariance estimates provides out-of-sample outperformances. Yet, the findings in the empirical section provide some useful insight on the suboptimality of using artificial portfolio constraints, which of course turn out to be useful especially for Large Portfolios (outperformance of JM on RM and EW), but still underperform a benchmark dynamic structured covariance estimator as proposed in DB.

8. Concluding Remarks

Jagannathan and Ma (2003) proposed constraining portfolio allocation rules as a way reduce estimation error/risk exposure through covariance regularization. The same is argued in DeMiguel et al. (2009b), Fan et al. (2008), DeMiguel et al. (2009a) and Brodie et al. (2008) among the others. They argued that more stable portfolio selection policies can be induced by imposing no-short sales, gross-exposure and upper bounds regardless covariance estimates. Interpreting portfolio weights as actions in a standard expected utility maximization framework, we first show the suboptimality of constraining weights with respect to a simple Bayesian benchmark rule. We generalize the expected utility framework to a common Bayesian setup to check for admissibility of portfolio decision rules in a standard statistical decision theory context, developing a Dynamic Bayesian covariance estimator as benchmark model. The latter combines a sequential shrinkage prior in the spirit of Ledoit and Wolf (2004) with a standard inverse wishart updating scheme. Based on simulation examples and an empirical analysis we show that artificially constrained allocations are suboptimal/inadmissible portfolio decision rules therefore incompatible with the standard von Neumann-Morgenstern rationality assumption in a general expected utility maximization framework.

Appendix

A. Proof of Proposition 1

Let us consider a general loss function as

$$L(\tilde{\mathbf{w}}, \hat{\mathbf{w}}) = L_u(\tilde{\mathbf{w}}, \hat{\mathbf{w}}) - \inf_{\hat{\mathbf{w}} \in \mathcal{A}} L_u(\tilde{\mathbf{w}}, \hat{\mathbf{w}}) \tag{43}$$

The posterior expected loss of the action $\hat{\mathbf{w}}(\theta)$, when the posterior distribution is $p(\theta|\mathbf{R})$ can be defined as

$$\rho(\tilde{\mathbf{w}}, \hat{\mathbf{w}}) = \int_{\Theta} L(\tilde{\mathbf{w}}, \hat{\mathbf{w}}) p(\theta|\mathbf{R}) d\theta \tag{44}$$

A (posterior) *Bayes Action* is then the one minimizes (44). Now integrating over the predictive distribution we get

$$R(\tilde{\mathbf{w}}, \hat{\mathbf{w}}) = \int_{\mathcal{R}} [\rho(\tilde{\mathbf{w}}, \hat{\mathbf{w}})] \, p(\tilde{\mathbf{R}}|\mathbf{R}) d\tilde{\mathbf{R}} = \int_{\mathcal{R}} \left[\int_{\Theta} L(\tilde{\mathbf{w}}, \hat{\mathbf{w}}) p(\theta|\mathbf{R}) d\theta \right] p(\tilde{\mathbf{R}}|\mathbf{R}) d\tilde{\mathbf{R}} \tag{45}$$

with $\tilde{\mathbf{R}}$ the unobserved future return. Now since $\rho(\tilde{\mathbf{w}}, \hat{\mathbf{w}}) \geq -\infty$ and all measures are finite, by Fubini's theorem we can invert the integrals such that

$$R(\tilde{\mathbf{w}}, \hat{\mathbf{w}}) = \int_{\Theta} \left[\int_{\mathcal{R}} L(\tilde{\mathbf{w}}, \hat{\mathbf{w}}) p(\tilde{\mathbf{R}}|\mathbf{R}) d\tilde{\mathbf{R}} \right] p(\theta|\mathbf{R}) d\theta \tag{46}$$

Now the intuitive recipe for finding a bayes strategy is to minimize the expected loss inside the brackets. Indeed, in doing this we minimize the integral outside the brackets finding the so called optimal bayes rule (see Parmigiani and Inoue (2009) for more details).

B. Proof of Proposition 3

Let us consider the inverse-wishart prior for Σ and an N-dimensional $\mathbf{Y}_t$ i.i.d and zero-mean normally distributed vector of returns at time t

$$\begin{aligned} p(\Sigma|\mathbf{J}, \nu_0) &\propto |\Sigma|^{-(\nu_0+N+1)/2} \exp\left\{ -\frac{1}{2} tr\left(\mathbf{J}\Sigma^{-1}\right) \right\} \\ p(\mathbf{Y}|\Sigma) &\propto |\Sigma|^{-T/2} \exp\left\{ -\frac{1}{2} tr\left(\Sigma^{-1}\mathbf{Y}'\mathbf{Y}\right) \right\} \end{aligned} \tag{47}$$

The posterior distribution is conjugate so is easily found as

$$p(\Sigma|\mathbf{Y}) \propto |\Sigma|^{-(\nu_n+N+1)/2} \exp\left\{ -\frac{1}{2} tr\left(\Sigma^{-1}\Lambda_n\right) \right\}, \quad \text{with} \quad \nu_n = \nu_0 + T, \Lambda_n = \mathbf{J} + \mathbf{Y}'\mathbf{Y} \tag{48}$$

such that the conditional mean turns out to be

$$E[\Sigma|\mathbf{Y}] = \frac{\Lambda_n}{\nu_n - N - 1} = \frac{\mathbf{J} + \mathbf{Y}'\mathbf{Y}}{\nu_0 + T - N - 1} = \alpha\tilde{\mathbf{J}} + (1-\alpha)\mathbf{S} \tag{49}$$

with

$$\alpha = \frac{\nu_0 - N - 1}{\nu_0 + T - N - 1} \in (0,1) \quad \text{and} \quad \tilde{\mathbf{J}} = \frac{\mathbf{J}}{\nu_0 - N - 1} \tag{50}$$

Taking the limits for $T \to \infty$ of (49) we get the proposition.

C. The Historical resampled allocation

In this section we explain the bootstrap procedure implemented to construct a robust portfolio allocation based on historical estimation. The aim is to define a portfolio allocation procedure comparable with the other portfolio allocation decision rules. Let us consider the optimal allocation function in (**??**), then under the assumption of short-sales we have the closed form solution in (**??**). Now, to define the optimal portfolio weights the resampling recipes has the following steps.

Step 1 Estimate the Covariance structure Σ_0 from the observed time series of returns. This is done by the usual MLE estimator

$$\hat{\Sigma} = \frac{1}{T-1}\sum_{t=1}^{T} R_t R_t' \tag{51}$$

with R_t the N-dimensional vector of returns at time t.

Step 2 Suppose the time series $\mathbf{R} \sim NID(0, \Sigma_0)$;

Step 3 Resample a large number Q of monte carlo scenarios from the distribution assumption in Step 2.

Step 4 Estimate Σ_q for $q = 1, ..., Q$ as in Step 1 from each of the $\mathbf{R}^q$ returns generated.

Step 5 Compute the global minimum variance portfolio weights $\mathbf{w}^q$ as in (**??**).

Step 6 Define the optimal resample portfolio as the average of the above allocations

$$\mathbf{w}^{mc} = \frac{1}{Q}\sum_{q=1}^{Q} \mathbf{w}^q, q = 1, ..., Q \tag{52}$$

This is not a probability model, but represents a fairly good practical standarad to deal with input parameters uncertainty. Yet, allows to use the historical estimation Σ_0 as a baseline input parameters.

References

Balduzzi, P. and Lynch, A., W. (1999). Transaction costs and predictability: some utility costs calculations. *Journal of Financial Economics*, (52):47–78.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., and Loris, I. (2008). Sparse and stable markowitz portfolios. *ECB Working Paper Series*, (936):1–22.

Chopra, V., K. and Ziemba, W., T. (1993). The effect of errors in means, variance and covariances on optimal portfolio choice. *Journal of Portfolio Management*, (Winter):6–11.

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

DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, (22):1916–1653.

Fan, J., Zhang, J., and Yu, K. (2008). Asset allocation and risk assessment with gross exposure constraints for vast portfolios. *Preprint, Princeton University*.

Frost, P., A. and Savarino, J., E. (1986). An empirical bayes approach to efficient portfolio selection. *The Journal of Financial and Quantitative Analysis*, 21(3):293–305.

Green, R., C. and Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified? *The Journal of Finance*, (47):1785–1809.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, (4):1651–1683.

Jorion, P. (1986). Bayes-stein estimation for portfolio analysis. *The Journal of Financial and Quantitative Analysis*, 21(3):279–292.

Kandel, S. and Stambaugh, R., F. (1996). On the predictability of stock returns: an asset allocation perspective. *Journal of Finance*, (51):385–424.

Klein, R., W. and Bawa, V., S. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, (3):215–231.

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

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

Markowitz, H., M. (1952a). Portfolio selection. *Journal of Finance*, 7:77–91.

Markowitz, H., M. (1952b). *Portfolio Selection: Efficient diversification of investments.* John Wiley & Sons, New York.

McCulloch, R. and Rossi, P., E. (1990). Posterior, predictive and utility-based approaches to testing the arbitrage pricing theory. *Journal of Financial Economics*, (28):7–38.

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

Michaud, R. (1998). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization.* New York: Oxford University Press.

Parmigiani, G. and Inoue, L. Y. T. (2009). *Decision Theory: Principles and Approaches.* John Wiley & Sons, New York.

Pastor, L. (2000). Portfolio selection and asset pricing models. *The Journal of Finance*, LV.

Pastor, L. and Stambaugh, R., F. (2000). Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, (56):335–381.

Polson, N., G. and Tew, V. B. (2000). Bayesian portfolio selection: An empirical analysis of the s&p 500 index 1970-1996. *Journal of Business, Economics and Statistics*, 18(2):164–173.

Table 1: List of portfolio decision rules considered

No.	Model	Abbreviation
Naive		
1.	1/N with rebalancing	EW
Portfolio Constraints		
2.	Shortsales constrained portfolio	JM
3.	Gross-Exposure portfolio $c = 2$	GE
Bayesian approach		
4.	Classical Bayesian portfolio allocation	BC
5.	Dynamic Bayesian with shrinking prior	DB
Alternative covariance estimators		
6.	Ledoit-Wolf with 1-factor covariance target	LW
7.	RiskMetrics	RM
Alternative robust portfolio rules		
8.	Bootstrap portfolio allocation	BU

Table 2: Expected Utility Maximization: Simulation Results - Insample Fixed

This table reports the risk exposure of minimum-variance portfolios using different allocations. The data are simulated with a time-varying covariance structure. The insample period is fixed to $M = 252$ daily observations, while portfolio size span from $N = 50$ to $N = 250$. The out-of-sample testing period is $T - M = 121$ trading days. The portfolio weights are daily rebalanced. *EW* represents the naive 1/N portfolio, *JM* and *GE* the no-short sales and the gross-exposure artificial constrained weights, *BC* a simple Bayesian portfolio strategy with time-varying covariance. Yet, *LW* and *RM* both reports an unconstrained GMVP plugging respectively a shrinkage covariance estimates towards a factor model and the usual exponential weighting moving average. Finally, *BU* is a resampling based unconstrained minimum-variance portfolio.

		Methodologies						
Size		RM	JM	GE	LW	BU	BC	EW
50	Risk Wealth	0.067990	0.060200	0.072780	0.062877	0.063797	0.056755	0.166028
	TurnOver	0.032587	0.006479	0.359226	0.020650	0.022060	0.019367	0.012118
	N. Long Pos.	20	10	18	18	18	18	50
100	Risk Wealth	0.057882	0.045158	0.051283	0.046298	0.048962	0.042798	0.141181
	TurnOver	0.045055	0.005214	0.439700	0.026849	0.030193	0.026561	0.012317
	N. Long Pos.	40	11	34	35	35	39	100
150	Risk Wealth	0.089913	0.053164	0.069232	0.057776	0.067641	0.052821	0.146090
	TurnOver	0.092401	0.007488	0.599743	0.043058	0.058818	0.038336	0.012537
	N. Long Pos.	69	19	37	65	67	66	150
200	Risk Wealth	0.099998	0.049277	0.119402	0.050160	0.078895	0.043762	0.138513
	TurnOver	0.149204	0.008329	0.932948	0.048565	0.099791	0.044383	0.012952
	N. Long Pos.	97	23	42	89	97	95	200
250	Risk Wealth	0.220221	0.297997	0.312212	0.040518	0.232431	0.039929	0.148290
	TurnOver	0.637470	0.000000	36.574793	0.042717	0.737727	0.045042	0.012710
	N. Long Pos.	124	1	16	105	124	118	250

Table 3: Expected Utility Maximization: Simulation Results - Portfolio Size Fixed

This table reports the risk exposure of minimum-variance portfolios using different allocations. The data are simulated with a time-varying covariance structure. The portfolio size is fixed to $N = 100$, while insample period changes spanning from $M = 120$ to $M = 210$. The out-of-sample testing period is $T - M = 121$ trading days. Portfolio weights are daily rebalanced. *EW* represents the naive 1/N portfolio, *JM* and *GE* the no-short sales and the gross-exposure artificial constrained weights, *BC* a simple Bayesian portfolio strategy with time-varying covariance. Yet, *LW* and *RM* both reports an unconstrained GMVP plugging respectively a shrinkage covariance estimates towards a factor model and the usual exponential weighting moving average. Finally, *BU* is a resampling based unconstrained minimum-variance portfolio.

		Methodologies						
Size		RM	JM	GE	LW	BU	BC	EW
	Risk Wealth	0.088811	0.049510	0.060835	0.047101	0.081881	0.046138	0.156424
120	TurnOver	0.074036	0.006686	0.479757	0.026652	0.067588	0.024985	0.012402
	N. Long Pos.	46	16	35	40	46	45	100
	Risk Wealth	0.072644	0.052958	0.062246	0.051857	0.063814	0.048408	0.162611
160	TurnOver	0.064289	0.006174	0.475588	0.028473	0.049362	0.027276	0.012108
	N. Long Pos.	44	15	34	39	42	44	100
	Risk Wealth	0.067737	0.046875	0.057602	0.048135	0.052614	0.046468	0.139600
210	TurnOver	0.052322	0.005500	0.459538	0.027385	0.035827	0.025708	0.012045
	N. Long Pos.	42	14	37	38	40	43	100

Table 4: Admissibility: Simulation results - Insample fixed

This table reports the value of the Actual Bayes Risk [ABR] reported in (26). The insample is fixed to $M = 252$, meaning, one year of daily trading data. The portfolio size span from $N = 50$ to $N = 250$. The portfolio is simulated out-of-sample for $T - M = 132$ daily observations, with daily rebalancing. The portfolio selection is made randomly and the results are averaged across the $K = 20$ simulations. The portfolio strategies are those reported in Table (1). *RM* represents the RiskMetrics methodology, *LW* is the shrinkage estimator with the one factor model covariance as target, *DB* the Dynamic Bayesian estimator defined in Section 5, *GE* represents the gross-exposure constrained portfolio with $c = 2$. *JM* is the noshort sales portfolio. Finally *BU* and *EW* are respectively the Historical Bootstrap plug-in portfolios and the $1/N$ naive portfolio strategy. The Table reports *ABR* annualized, the level of *Turnover* implied by the portfolio strategy and the number of *Long* positions in the portfolios.

		Methodologies						
Size		RM	JM	GE	LW	BU	DB	EW
	ABR	0.002332	0.002861	0.012462	0.002850	0.002888	0.003362	0.025271
50	TurnOver	0.023384	0.019660	0.019720	0.019593	0.019795	0.009713	0.012451
	N. Long Pos.	18	18	18	18	18	17	50
	ABR	0.003301	0.003349	0.011390	0.003243	0.003440	0.001702	0.019623
100	TurnOver	0.038739	0.035051	0.046377	0.034179	0.035333	0.012155	0.012733
	N. Long Pos.	41	39	41	39	39	39	100
	ABR	0.001520	0.001828	0.004560	0.001605	0.001846	0.000973	0.020191
150	TurnOver	0.041755	0.039056	0.039211	0.035322	0.039379	0.009238	0.012965
	N. Long Pos.	61	52	52	50	52	54	150
	ABR	0.003827	0.005082	0.008582	0.003470	0.005279	0.001358	0.019187
200	TurnOver	0.061898	0.072551	0.072870	0.053389	0.073255	0.012438	0.012429
	N. Long Pos.	91	88	88	83	88	88	200
	ABR	0.001755	0.036783	0.191680	0.001951	0.048301	0.001566	0.012975
250	TurnOver	0.054036	0.388994	0.388994	0.054629	0.455052	0.013165	0.012169
	N. Long Pos.	115	123	123	106	125	114	250

Table 5: Admissibility: Simulation results - Portfolio size fixed

This table reports the value of the Actual Bayes Risk [ABR] reported in (26). The portfolio size is fixed $N = 100$, while the insample information span from $M = 100$ to $M = 210$. The portfolio is simulated out-of-sample for $T - M = 132$ daily observations, with daily rebalancing. The portfolio selection is made randomly and the results are averaged across the $K = 20$ simulations. The portfolio strategies are those reported in Table (1). *RM* represents the RiskMetrics methodology, *LW* is the shrinkage estimator with the one factor model covariance as target, *DB* the Dynamic Bayesian estimator defined in Section 5, *GE* represents the gross-exposure constrained portfolio with $c = 2$. *JM* is the noshort sales portfolio. Finally *BU* and *EW* are respectively the Historical Bootstrap plug-in portfolios and the $1/N$ naive portfolio strategy. The Table reports *ABR* annualized, the level of *Turnover* implied by the portfolio strategy and the number of *Long* positions in the portfolios.

		Methodologies						
Size		RM	JM	GE	LW	BU	DB	EW
130	ABR	0.002258	0.004042	0.007596	0.002793	0.004209	0.001275	0.018520
	TurnOver	0.029525	0.050581	0.053169	0.038019	0.051433	0.010964	0.012733
	N. Long Pos.	44	40	40	38	40	38	130
170	ABR	0.002466	0.003146	0.009614	0.002803	0.003199	0.001663	0.019193
	TurnOver	0.034343	0.041511	0.052893	0.037870	0.041903	0.012150	0.012051
	N. Long Pos.	44	43	45	43	44	43	170
210	ABR	0.002765	0.002628	0.008771	0.002513	0.002658	0.002073	0.018657
	TurnOver	0.038760	0.035189	0.042637	0.033787	0.035546	0.012055	0.012254
	N. Long Pos.	43	41	42	40	41	31	210

Table 6: Empirical Results - NYSE/AMEX

This table reports the empirical results on the NYSE/AMEX stocks returns. The portfolio strategy headings are those reported on Table (1). The performances are analyzed backtesting each of the portfolio strategies on $T - M = 132$ trading days out-of-sample with daily rebalancing. The portfolio selection is made randomly and the size span from $N = 50$ to $N = 400$. Rolling sample estimation is used except for DB and RM since they properly discount insample information through a decay factor $\delta \in (0, 1)$. The insample consists of $M = 504$ two years of daily data. The out-of-sample performance metrics are reported in Section 7.

		Methodologies						
Size		RM	JM	GE	LW	BU	DB	EW
50	Mean	0.00119	-0.08040	-0.10786	-0.08040	-0.07822	0.03858	-0.13914
	Risk	0.06709	0.10260	0.20891	0.10251	0.10241	0.07347	0.18151
	Sharpe Ratio	0.01772	-0.78358	-0.51628	-0.78434	-0.76387	0.52520	-0.76658
	Turnover	0.02513	0.02420	0.02423	0.02400	0.02425	0.00744	0.01195
	Min	-0.38894	-0.17231	-0.17265	-0.16421	-0.17354	-0.00852	0.02000
	Max	0.46457	0.59909	0.59899	0.59799	0.59896	0.17221	0.02000
	Long	17	16	16	16	16	16	50
200	Mean	0.03063	-0.05204	-0.00850	-0.05124	-0.06067	0.04722	-0.06104
	Risk	0.05734	0.04279	0.06575	0.04269	0.04329	0.04954	0.19439
	Sharpe Ratio	0.53422	-1.21616	-0.12932	-1.20026	-1.40127	0.95237	-0.31400
	Turnover	0.04897	0.03376	0.03382	0.03349	0.03396	0.00776	0.01179
	Min	-0.18895	-0.05462	-0.05473	-0.05406	-0.05516	-0.00536	0.00500
	Max	0.49880	0.45460	0.45422	0.45428	0.45428	0.19035	0.00500
	Long	80	86	86	85	85	105	200
400	Mean	-0.14179	-0.27351	-0.42249	-0.27076	-0.23470	0.01164	-0.23645
	Risk	0.07553	0.07811	0.12195	0.07237	0.08049	0.03304	0.13660
	Sharpe Ratio	-1.87730	-3.50178	-3.46444	-3.74153	-2.91590	0.35244	-1.73093
	Turnover	0.07769	0.08126	0.08126	0.07266	0.08157	0.01164	0.01218
	Min	-0.09733	-0.13922	-0.13922	-0.10682	-0.13862	-0.00756	0.00250
	Max	0.37332	0.41630	0.41630	0.42406	0.41676	0.17959	0.00250
	Long	206	206	206	208	207	253	400

Figure 1: Relative Economic Loss insample information fixed $M = 252$

This figure reports the Relative Economic Loss defined in (34), obtained changing the portfolio size from $N = 50$ to $N = 250$, and fixing the insample length to $M = 252$. The out-of-sample information is six months of daily trading returns, i.e. $T - M = 132$.

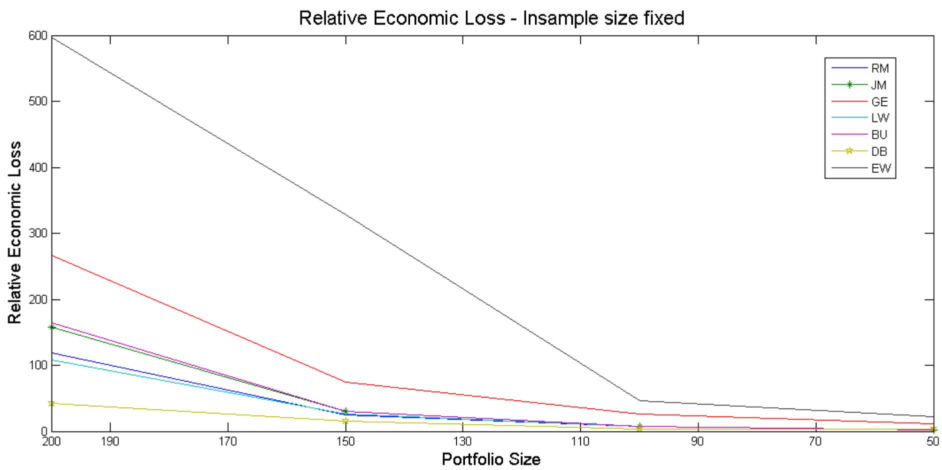

Figure 2: Relative Economic Loss portfolio size fixed $N = 100$

This figure reports the Relative Economic Loss defined in (34). The insample length span from $M = 130$ to $M = 210$, and the portfolio size is fixed at $N = 100$. The out-of-sample information is six months of daily trading returns, i.e. $T - M = 132$.

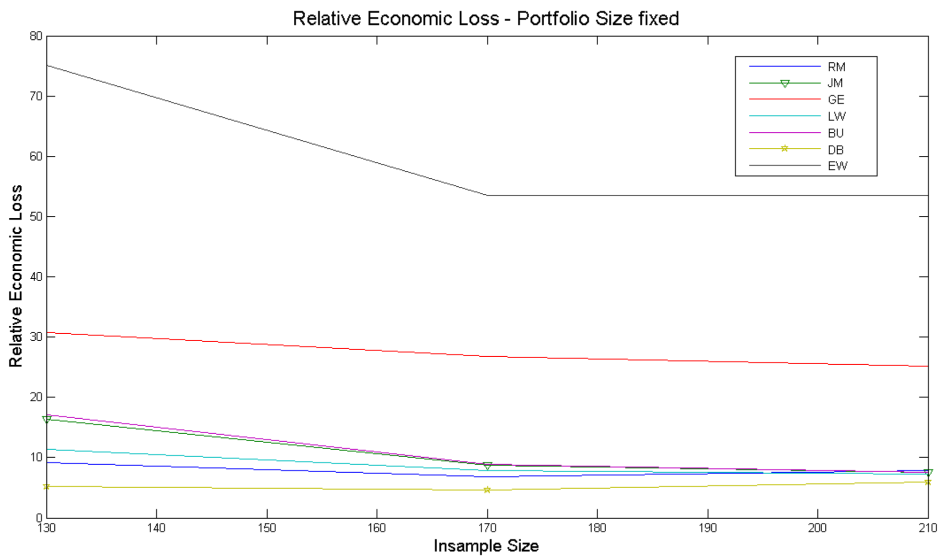

Figure 3: Cumulative Wealth out-of-sample $N = 50$

This figure reports the Cumulative portfolio wealth for the $N = 50$ portfolio size on a six-month daily tradin out-of-sample period. The portfolio is selected randomly at the beginning and kept constants across rebalancing to avoid any particular stock picking issue.

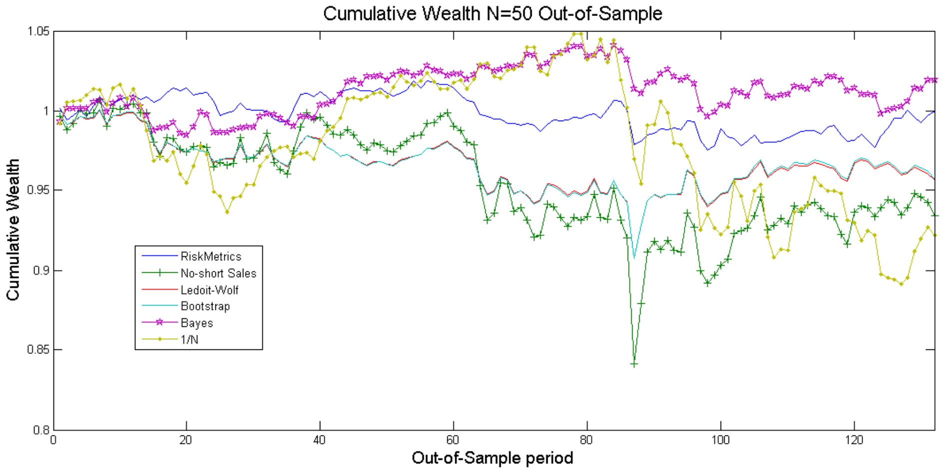

Figure 4: Cumulative Wealth out-of-sample $N = 200$

This figure reports the Cumulative portfolio wealth for the $N = 200$ portfolio size on a six-month daily tradin out-of-sample period. The portfolio is selected randomly at the beginning and kept constants across rebalancing to avoid any particular stock picking issue.