The Effect of Estimation in High-dimensional Portfolios

Axel Gandy[1] and Luitgard A. M. Veraart[2] [*]
[1]Imperial College London, [2]London School of Ecconomics

Abstract

We study the effect of estimated model parameters in investment strategies on expected log-utility of terminal wealth. The market consists of a riskless bond and a potentially vast number of risky stocks modelled as geometric Brownian motions. The well-known optimal Merton strategy depends on unknown parameters and thus cannot be used in practice. We consider the expected utility of several estimated strategies when the number of risky assets gets large. We suggest strategies which are less affected by estimation errors and demonstrate their performance in a real data example. Strategies in which the investment proportions satisfy an L_1-constraint are less affected by estimation effects.

KEY WORDS: optimal investment, continuous time, estimation effects, lasso, shrinkage, vast portfolios

1 Introduction

We consider an investor who seeks to maximise expected logarithmic utility of terminal wealth. We assume the same setup as in Merton (1971), i.e. the financial market consists of a riskless bond and a potentially vast number of risky stocks modelled as geometric Brownian motions. The optimal strategy in this setup, derived in Merton (1971), depends on the unknown model parameters, the vector of drifts and the volatility matrix. To use this optimal strategy, these parameters need to be estimated.

How does estimation influence the expected utility, in particular if the market consists of many stocks? We answer this question for various types of investment strategies.

The present paper contains several new results. First, we give analytic results for the expected utility of several strategies that use estimated parameters. Second, we provide new explicit optimal strategies for L_1-restricted portfolio optimisation problems. Third, we analyse the limiting behaviour of the expected utility when the number of available assets goes to infinity. Fourth, via theoretical considerations, simulations, and an empirical study, we discuss which strategies are less affected by estimation errors if a vast number of assets is available.

Parameter uncertainty is a well-known problem in portfolio selection, see e.g. Merton (1971); Detemple (1986, 1991); Dothan and Feldman (1986); Gennotte (1986). The problem of parameter uncertainty in the static Markowitz (1952, 1959) situation is for example considered by Bai et al. (2009). It is popular to use a Bayesian approach in continuous-time portfolio selection, which allows the use of filtering

[*]Address correspondence to Luitgard A. M. Veraart, London School of Economics, Department of Mathematics, Houghton Street, London WC2A 2AE, UK, l.veraart@lse.ac.uk

methods to account for parameter estimation, see e.g. Karatzas and Zhao (2001); Rogers (2001); Rieder and Bäuerle (2005); Garlappi et al. (2007). All of the above references do not consider the effect when the number of stocks d goes to infinity.

In this paper we consider continuous-time portfolio optimisation when the number of assets gets large/goes to infinity. We assume that the parameters are estimated on a set of past data and plugged into the various strategies. We do not use any Bayesian/filtering approach. We consider this problem in a Merton market. To our knowledge, this has not been discussed in the literature. The effect of letting the number of available assets tend to infinity has only been studied in the Markowitz framework (Pesaran and Zaffaroni, 2008).

Our analysis is based on a simple setup with logarithmic utility function and constant market coefficients. In this setup, we can show the key points about estimation effects: they must not be ignored and they can be reduced by using constrained strategies. There are more general portfolio choice problems with observable time-dependent stochastic model parameters, see e.g. Ocone and Karatzas (1991); Detemple et al. (2003). In these more general setups, time-dependent parameters would have to be estimated instead of mere constants. This should make the estimation effects even more prominent.

When letting the number d of assets tend to infinity we always consider the following example: The drift of the assets are realisation of independent and identically distributed random variables. The volatility matrix is such that, conditionally on those drifts, the log returns of different assets have a fixed constant correlation. In this setup, the optimal Merton-strategy with known parameters leads to a linearly increasing expected utility in the number d of available assets (Example 2.1).

In our theoretical derivations (Sections 3-6), but not in our simulation study (Section 7), we assume that the volatility matrix is known. This is reasonable because volatility is in principle much easier to estimate, particularly, when high-frequency data are available.

We assume that the mean returns are unknown. It is well-known that estimating the expected return is a very difficult task. As (Merton, 1980, p. 4) said: "...one might say that to attempt to estimate the expected return on the market is to embark on a fool's errand." For the expected return, sampling at higher frequencies does not improve the estimates. Instead, a very long time horizon is needed for an accurate estimation, see e.g. Merton (1980); Rogers (2001). Still, the expected return is one of the main input parameters in our and almost all portfolio selection models, and therefore we have to tackle this problem.

We begin our study of the effect of estimation on expected utility in Section 3 by considering the optimal Merton strategy with a plug-in estimator of the expected return. Our analytic results show that this plug-in estimator has a detrimental effect: when we send the number of stocks d to infinity then the expected utility goes to $-\infty$ in realistic cases.

In Section 4, we investigate whether this can be prevented by improving the strategy through James-Stein type shrinkage (Stein, 1956; James and Stein, 1961). This shrinkage improves an estimator for a multivariate quantity by shrinking it towards a specific target, resulting in a biased estimator with smaller mean square error. This can be interpreted as an empirical Bayes approach (Efron and Morris, 1975). The use of James-Stein-type estimators for the mean return and/or the covariance matrix has been advocated by e.g. Jorion (1986); Gruber (1998); Ledoit and Wolf (2004); Golosnoy and Okhrin (2009).

We apply James-Stein type shrinkage directly to the strategy rather than the

drift or the volatility. To do this one has to pick a strategy one is shrinking towards. When shrinking towards the optimal strategy, the James-Stein strategy is (almost) as good as the optimal strategy - except for a fixed penalty not depending on the number d of available stocks, see (4.2). In practice, when the optimal strategy is unknown, one has a linear loss of expected utility as the number d of available stocks goes to infinity, see (4.3).

Next, in Sections 5 and 6, we investigate whether restricting the set of investment strategies suitably improves the situation. Generally, restricting the set of strategies cannot lead to higher expected utility if the parameters are known. If those have to be estimated, however, it will turn out that restricting the set of strategies can improve the expected utility since the resulting strategies are less affected by estimation errors and more robust.

In the Markowitz portfolio selection model, Jagannathan and Ma (2003) show that restricting the strategies by a no-short-selling constraint improves the portfolio performance. The no-short-selling constraint is a special case of an L_1-constraint on the portfolio weights, i.e. the sum of the absolute values of the proportions of wealth invested in the risky assets is required to be bounded above by a predetermined constant. Fan et al. (2009) and Brodie et al. (2009) consider such a general L_1-constraint in the Markowitz framework. They find that the no-short-selling portfolio is usually too restrictive and can be improved by allowing some short-positions such that the sum of the absolute value of the portfolio weights still stays below a previously fixed level. The advantage of the L_1-constraint is that it induces sparsity into the portfolio, i.e. it does not invest in all assets. This is reasonable when the number of assets gets large, particularly in the presence of transaction costs and market frictions. Compared to optimal subset selection, which is an (NP-) hard problem, there are computational benefits in using an L_1-constraint, as efficient algorithms to solve L_1-constrained optimisation problems have been developed (Efron et al., 2004). In addition to sparsity, an L_1-constraint avoids the accumulation of estimation errors (Fan et al., 2009). There is empirical evidence that these L_1-constrained strategies even outperform an evenly balanced portfolio which is very rarely the case for a strategy obtained from solving an optimisation model (DeMiguel et al., 2007). None of the above papers consider the expected utility as $d \to \infty$, nor do they work in a Merton setup.

In Section 5, we investigate the effect of an L_1-constraint in the Merton context. A major advantage of an L_1-constraint is, as we will show, that the expected utility is bounded from below as $d \to \infty$.

We also provide analytic results for the optimal strategies with L_1-constraint and known parameters for a specific volatility matrix (Theorem 5.2, Corollary 5.3). We see that the L_1-constrained strategies are sparse, i.e. do not invest in all the assets.

We study how quickly this optimal strategy diversifies when plug-in estimators for the expected return are used. By using results from extreme value statistics, we obtain analytic results for $d \to \infty$ (Theorems 5.4, 5.5). We find that the diversification is very slow, i.e. the strategy invests only in few assets. This makes the resulting strategy less robust against changes in the market.

In Section 6, we consider two further L_1-constrained strategies: the $1/d$-strategy, that invests an equal amount in all available stocks, and a modification of the $1/d$-strategy, that only invests in the stocks with the most extreme returns. We call this strategy EWE-strategy (Equal Weighting of Extreme stocks). Essentially, the EWE-strategy is designed to have a higher return than the $1/d$-strategy whilst retaining its diversification and robustness properties. We find that the EWE-strategy performs

particularly well in our simulations as well as in the application to real data.

The theoretical as well as the out-of-sample utility of the strategies in examples are illustrated in Section 7. Section 8 contains a discussion and Section 9 contains conclusions. Proofs can be found in the appendix.

2 Model Definition and Classical Solution

2.1 The Financial Market and the Investor's Objective

We are working in a market with $d + 1$ assets, one bond and d stocks. We assume the standard log-Gaussian dynamics

$$dS_0(t) = S_0(t)rdt, \quad S_0(0) = 1,$$

$$dS_i(t) = S_i(t)\left(\mu_i dt + \sum_{j=1}^{d}\sigma_{ij}dW_j(t)\right), \quad S_i(0) > 0, \quad i = 1,\ldots,d,$$

where $r > 0$ is the constant interest rate, $W = (W_1,\ldots,W_d)^\top$ is a standard d-variate Brownian motion, $\mu = (\mu_1,\ldots,\mu_d)^\top$ is the constant drift and $\sigma = (\sigma_{ij})_{1\leq i,j\leq d}$ is the constant $d \times d$-volatility matrix. We assume that σ is of full rank. Explicitly,

$$S_i(t) = S_i(0)\exp\left(\left(\mu_i - \frac{1}{2}\sum_{j=1}^{d}\sigma_{ij}^2\right)t + \sum_{j=1}^{d}\sigma_{ij}W_j(t)\right), \quad i = 1,\ldots,d.$$

We assume that in this market an investor seeks to maximise

$$V(\pi|\mu) := \mathbb{E}[\log(X_T)],$$

where $T > 0$ is some fixed time, X_t denotes their wealth at time t, which satisfies

$$dX_t = \sum_{i=0}^{d}\pi_i(t)X_t\frac{dS_i(t)}{S_i(t)}$$

and $\pi_i(t)$ denotes the fraction of the wealth invested in the ith asset at time t. Hence, $\sum_{i=0}^{d}\pi_i(t) = 1$ for all t. We assume that X_0 is a constant. We use the notation $V(\pi|\mu)$ to emphasize that we will later allow the true μ to be random and that we will consider the above model conditionally on μ.

2.2 The Merton Solution

Next, we briefly derive the classical Merton (1971) strategy which is optimal among all strategies π that are adapted to the filtration $\mathcal{F}_t = \sigma(W(s), s \leq t), t \geq 0$ and that are sufficiently integrable such that their stochastic integrals with respect to W are martingales. Using $\pi_0(t) = 1 - \sum_{i=1}^{d}\pi_i(t)$ and setting $\pi = (\pi_1,\ldots,\pi_d)^\top$ we get

$$\frac{dX_t}{X_t} = \left(r + \pi^\top(t)(\mu - r1)\right)dt + \pi^\top(t)\sigma dW(t),$$

where $1 = (1,\ldots,1)^\top \in \mathbb{R}^d$. Hence,

$$X_T = X_0\exp\left(\int_0^T\left(r + \pi^\top(t)(\mu - r1) - \frac{1}{2}\pi(t)^\top\Sigma\pi(t)\right)dt + \int_0^T\pi(t)\sigma dW(t)\right),$$

where $\Sigma = \sigma\sigma^\top$. Therefore, the expected log-utility is

$$V(\pi|\mu) = \mathbb{E}[\log(X_T)] = \log(X_0) + rT + \mathbb{E}\left[\int_0^T \left(\pi(t)^\top(\mu - r1) - \frac{1}{2}\pi(t)^\top\Sigma\pi(t)\right)dt\right].$$

Hence the optimal π is constant in time. For any time-constant strategy π,

$$(2.1) \qquad V(\pi|\mu) = \log(X_0) + rT + T\mathbb{E}\left[\pi^\top(\mu - r1) - \frac{1}{2}\pi^\top\Sigma\pi\right].$$

The term in brackets can be written as $-\frac{1}{2}\|\sigma^\top\pi - \sigma^{-1}(\mu - r1)\|_2^2 + c$, where c does not depend on π and $\|\cdot\|_2$ is the Euclidean norm. Thus obtaining the optimal strategy requires the solution of a classic quadratic optimisation problem. If there are no constraints on π then the optimal strategy is

$$(2.2) \qquad\qquad\qquad \pi^* = \Sigma^{-1}(\mu - r1).$$

The corresponding expected utility is

$$(2.3) \qquad V(\pi^*|\mu) = \log(X_0) + rT + \frac{T}{2}[(\mu - r1)^\top\Sigma^{-1}(\mu - r1)].$$

2.3 Expected Utility as $d \to \infty$

In the following example, we consider the effect of increasing the number of assets on the expected utility. We assume that all model parameters are known and that the covariance matrix has a very simple structure. Particularly, we assume that all stocks have the same volatility and all pairs of stocks have the same correlation.

In the example we let μ be random and consider the model described in Sections 2.1, 2.2 conditionally on μ. We denote the utility obtained by integrating over μ by

$$V(\pi) = \mathbb{E}(V(\pi|\mu)).$$

Letting μ be random allows us to easily specify properties of μ as $d \to \infty$. $V(\pi)$ can be interpreted as average behaviour of the strategy over several different "truths".

Example 2.1 (Utility of the Merton strategy as $d \to \infty$).
Suppose $\Sigma = \eta^2\left[(1 - \rho)I + \rho 11^\top\right]$, for some $\eta^2 > 0$, $0 \le \rho < 1$, where $I \in \mathbb{R}^{d \times d}$ is the identity matrix and $1 = (1, \ldots, 1)^\top \in \mathbb{R}^d$. Furthermore, suppose that $\mu_1, \ldots, \mu_d$ are i.i.d. with $\mathbb{E}(\mu_i^2) < \infty$. Then, as we show in Appendix B,

$$(2.4) \qquad V(\pi^*) = \log(X_0) + rT + Td\frac{(1 + \rho(d - 2))\mathbb{V}ar[\mu_1] + (1 - \rho)(\mathbb{E}\mu_1 - r)^2}{2\eta^2(1 - \rho)(1 + \rho(d - 1))}.$$

Hence, if $\rho > 0$,

$$V(\pi^*) = \begin{cases} d\frac{T}{2\eta^2(1-\rho)}\mathbb{V}ar(\mu_1) + o(d), & \text{if } \mathbb{V}ar(\mu_1) > 0, \\ \log(X_0) + rT + \frac{T[\mathbb{E}\mu_1 - r]^2}{2\eta^2\rho} + o(1), & \text{if } \mathbb{V}ar(\mu_1) = 0, \end{cases}$$

where $o(d)$ and $o(1)$ are meant as $d \to \infty$. Thus $V(\pi^*)$ is linearly increasing in d unless $\mathbb{V}ar(\mu_1) = 0$. The latter case, when all stocks have the same non-random return, is obviously not very realistic.

If $\rho = 0$ then $V(\pi^*) = \log(X_0) + rT + \frac{Td}{2\eta^2}\left[\mathbb{V}ar(\mu_1) + [\mathbb{E}\mu_1 - r]^2\right]$ and hence,

$$V(\pi^*) = \begin{cases} d\frac{T}{2\eta^2}\left(\mathbb{V}ar(\mu_1) + [\mathbb{E}\mu_1 - r]^2\right) + o(d), & \text{if } \mathbb{P}(\mu_1 \ne r) > 0, \\ \log(X_0) + rT, & \text{else.} \end{cases}$$

Thus, unless all stocks have the same return as the risk-free bond, the utility will increase linearly with d.

Remark 2.2. How does the utility $V(\pi^*)$ depend on the correlation ρ in the previous example? One can show that $\frac{\partial}{\partial\rho}V(\pi^*)|_{\rho=0} = \frac{T}{2\eta^2}d(1-d)[\mathbb{E}\mu_1 - r]^2 < 0$ for $d > 1$. Furthermore, $\frac{\partial}{\partial\rho}V(\pi^*)$ has two roots. One that is clearly negative for large d, and the other one is

$$\frac{-[\mathbb{E}\mu_1 - r]^2 - \mathbb{V}ar(\mu_1) + \sqrt{\mathbb{V}ar(\mu_1)^2 + [\mathbb{E}\mu_1 - r]^2\mathbb{V}ar(\mu_1)\,d}}{(d-2)\,\mathbb{V}ar(\mu_1) - [\mathbb{E}\mu_1 - r]^2}.$$

For large d this is between 0 and 1. Hence, for large d, $V(\pi^*)$ is initially decreasing in ρ and increasing afterwards.

We have seen in the example that when parameters are known and not all the stocks have the same return, the expected utility is a linearly increasing function in the number of stocks. We will see in the next section that this is no longer the case when parameters have to be estimated.

3 Plug-in Merton Strategy

3.1 General Plug-in Strategies

In general the drift parameter μ and the volatility matrix σ will be unknown and need to be estimated from empirical data. In this section we consider the case where σ is known but μ is unknown and an estimator $\hat{\mu}$ is plugged into the optimal strategy π^*, i.e. we consider the strategy $\hat{\pi} = \Sigma^{-1}(\hat{\mu} - r1)$.

If $\hat{\pi} \sim N(m_0, V_0^2)$, for some $m_0 \in \mathbb{R}^d$ and $V_0^2 \in \mathbb{R}^{d \times d}$ then the expected utility is

$$V(\hat{\pi}|\mu) = \log(X_0) + rT + T\mathbb{E}[\hat{\pi}^\top(\mu - r1) - \frac{1}{2}\hat{\pi}^\top\Sigma\hat{\pi}]$$

$$= \log(X_0) + rT + T\left(m_0^\top(\mu - r1) - \frac{1}{2}[\text{tr}(\Sigma V_0^2) + m_0^\top\Sigma m_0]\right),$$

where tr denotes the trace operator. If, in addition, $\hat{\pi}$ is unbiased for the optimal strategy, i.e. $m_0 = \Sigma^{-1}(\mu - r1)$ then, using (2.3),

$$V(\hat{\pi}|\mu) = V(\pi^*|\mu) - \frac{T}{2}\text{tr}(\Sigma V_0^2).$$

Suppose we have observed the value S of the stocks over some past time interval $[-t_{est}, 0]$ for some $t_{est} > 0$. Then for $i = 1, \ldots, d$,

$$(3.1) \qquad \hat{\mu}_i = \frac{\log(S_i(0)) - \log(S_i(-t_{est}))}{t_{est}} + \frac{1}{2}\sum_{j=1}^{d}\sigma_{ij}^2$$

is an unbiased estimator of μ_i. For this estimator, $V_0^2 = \Sigma^{-1}/t_{est}$. Thus in this case

$$(3.2) \qquad V(\hat{\pi}|\mu) = V(\pi^*|\mu) - d\frac{T}{2t_{est}}.$$

Thus the loss through estimation is linear in the number of available assets. Moreover, it is increasing linearly in the time horizon and decreasing in the length of the estimation period t_{est}.

Remark 3.1. Can we do better if we update the estimator at each time point? Consider the the time-dependent plug-in strategy $\hat{\pi}(t) = \Sigma^{-1}(\hat{\mu}(t) - r1)$ with

$$\hat{\mu}_i(t) = \frac{\log(S_i(t)) - \log(S_i(-t_{est}))}{t + t_{est}} + \frac{1}{2}\sum_{j=1}^{d}\sigma_{ij}^2.$$

The expected utility of this strategy can be shown to be

$$V(\hat{\pi}(\cdot)|\mu) = V(\pi^*|\mu) - \frac{d}{2}\log\left(1 + \frac{T}{t_{est}}\right).$$

Comparing this to (3.2), we see that the loss through estimation is still linear in d — only the constant changes. The continuous updating of the estimator will lead to appreciable improvements only if the length T of the investment period is large compared to the length t_{est} of the initial estimation period.

3.2 Expected Utility of Plug-in Strategy as $d \to \infty$

Example 3.2. What happens to $V(\hat{\pi})$ as $d \to \infty$? Consider the same setup as in Example 2.1 with $\rho > 0$. Suppose we use the estimator in (3.1). Then

$$V(\hat{\pi}) = \mathbb{E}(V(\hat{\pi}|\mu)) = \mathbb{E}(V(\pi^*|\mu)) - \frac{Td}{2t_{est}} = d\frac{T}{2}\left(\frac{\mathbb{V}ar(\mu_1)}{\eta^2(1-\rho)} - \frac{1}{t_{est}}\right) + o(d)$$

$$\to \begin{cases} \infty, & \mathbb{V}ar(\mu_1) > \frac{\eta^2(1-\rho)}{t_{est}} \\ -\infty, & \mathbb{V}ar(\mu_1) < \frac{\eta^2(1-\rho)}{t_{est}} \end{cases} \quad (d \to \infty).$$

In particular if all stocks are the same ($\mathbb{V}ar(\mu_1) = 0$) then the limiting utility is $-\infty$.

Since we assume a special structure of Σ and (3.1), the covariance matrix of the estimator $\hat{\mu}$ is

$$\Sigma/t_{est} = \frac{\eta^2(1-\rho)}{t_{est}}I + \frac{\eta^2\rho}{t_{est}}11^\top.$$

Hence, the limiting behaviour of $V(\hat{\pi})$ depends on how $\mathbb{V}ar(\mu_1)$ relates to the idiosyncratic part of $\mathbb{V}ar(\hat{\mu}_1)$ (which is $\frac{\eta^2(1-\rho)}{t_{est}}$). In practice, we would usually expect the variance of μ_1 to be smaller than the idiosyncratic part of the variance of $\hat{\mu}_1$, i.e. $\mathbb{V}ar(\mu_1) < \frac{\eta^2(1-\rho)}{t_{est}}$, in which case the expected utility goes to $-\infty$.

4 James-Stein-type Shrinkage of the Strategy

In this section, we consider James-Stein-type (JS) shrinkage of the Merton plug-in strategy towards some given fixed strategy π^0. More precisely, we consider the strategy

$$\hat{\pi}^{JS,\pi^0} = \left(1 - \frac{a}{(\hat{\pi} - \pi^0)^\top\Sigma(\hat{\pi} - \pi^0)}\right)(\hat{\pi} - \pi^0) + \pi^0,$$

where $\hat{\pi} = \Sigma^{-1}(\hat{\mu} - r1)$ and $\pi^0 \in \mathbb{R}^d$, $a > 0$ are fixed constants.

4.1 JS-Strategy

The following theorem is the transformation of the classical result about the loss of the James-Stein estimator (James and Stein, 1961) to our situation.

Theorem 4.1. *If $\hat{\mu} \sim N(\mu, \Sigma/t_{est})$ then*

$$V(\hat{\pi}^{JS,\pi^0}|\mu) = V(\hat{\pi}|\mu) + \frac{T}{2}a\left[2\frac{d-2}{t_{est}} - a\right]\mathbb{E}\left[\frac{t_{est}}{d-2+2K}\right],$$

where $K \sim Poisson(\lambda)$, $\lambda = (\pi^* - \pi^0)^\top \Sigma(\pi^* - \pi^0)/2$ and $\hat{\pi}$ is the Merton plug-in strategy. $V(\hat{\pi}^{JS,\pi^0}|\mu)$ is maximised for $a = (d-2)/t_{est}$, giving

$$(4.1) \qquad V(\hat{\pi}^{JS,\pi^0}|\mu) = V(\hat{\pi}|\mu) + \frac{T}{2}\frac{(d-2)^2}{t_{est}}\mathbb{E}\left[\frac{1}{d-2+2K}\right].$$

A proof can be found in Appendix C.

The above theorem shows that the James-Stein strategy strictly dominates the Merton plug-in strategy $\hat{\pi}$ for $0 < a < 2(d-2)/t_{est}$.

If $\lambda = 0$, i.e. one is shrinking towards the optimal strategy π^*, then using the optimal $a = (d-2)/t_{est}$ and the estimator given in (3.1) we get

$$(4.2) \qquad V(\hat{\pi}^{JS,\pi^*}|\mu) = V(\hat{\pi}|\mu) + \frac{T(d-2)}{2t_{est}} = V(\pi^*|\mu) - \frac{T}{t_{est}}.$$

Thus, if one is shrinking towards the optimal strategy then the James-Stein strategy is (almost) as good as the optimal strategy - except for a fixed penalty that does not depend on the number d of available stocks.

Remark 4.2. The classical James-Stein estimator can also be interpreted as an empirical Bayes approach, see Efron and Morris (1975). In our context, the empirical Bayes approach would be as follows. We would assume a normal prior on π^* with expectation π^0. The prior variance of π^* would be estimated using the observed $\hat{\pi}$ (this is the empirical Bayes part). Then the posterior mean of π^* will coincide with the JS estimator $\hat{\pi}^{JS,\pi^0}$.

4.2 Expected Utility of JS-Strategy as $d \to \infty$

In the following, we consider the far more realistic case of not shrinking towards the correct strategy. We use the optimal a and shrink towards $\pi^0 = \frac{\beta}{d}1$ for some $\beta \in \mathbb{R}$. We assume that $\mu_1, \ldots, \mu_d$ are i.i.d. normally distributed with $\mathbb{P}(\mu_1 \neq r) > 0$ and $\mathbb{V}ar(\mu_1) < \infty$ and that $\Sigma = \eta^2 \left[(1-\rho)I + \rho 11^\top\right]$.

By Lemma C.1, $\mathbb{E}\left[\frac{1}{d-2+2K}\right] = \mathbb{E}(\mathbb{E}\left[\frac{1}{d-2+2K}|\lambda\right]) = \mathbb{E}\left[\frac{1}{d-2+2\lambda}\right] + O(\frac{1}{d^2})$. Then,

$$\lambda = \frac{1}{2}(\pi^* - \pi^0)^\top \Sigma(\pi^* - \pi^0) = \frac{1}{2}(\mu - r1)^\top \Sigma^{-1}(\mu - r1) + o(d)$$

almost surely, because $(\pi^0)^\top \Sigma \pi^* = (\pi^0)^\top(\mu - r1) = \frac{\beta}{d}\sum_{i=1}^d (\mu_i - r) \to \beta\mathbb{E}(\mu_1 - r)$ almost surely and $(\pi^0)^\top \Sigma \pi^0 = \frac{\beta^2}{d^2}\eta^2[d(1-\rho) + \rho d^2] = O(1)$. Using Lemma A.1 and the strong law of large numbers,

$$\lambda = d\frac{1}{2}\frac{1}{\eta^2(1-\rho)}\left(\frac{1}{d}(\mu - r1)^\top(\mu - r1) - \frac{\rho d}{1-\rho+\rho d}\left(\frac{1}{d}1^\top(\mu - r1)\right)^2\right) + o(d)$$

$$= \frac{1}{2}\alpha d + o(d),$$

where $\alpha = \frac{1}{\eta^2(1-\rho)}\mathbb{V}ar(\mu_1)$. By the bounded convergence theorem (for $d \geq 3$: $d/(d-2+2\lambda) \leq d/(d-2) \leq 3$),

$$d\mathbb{E}\left[\frac{1}{d-2+2\lambda}\right] = \mathbb{E}\left[\frac{d}{(1+\alpha)d+o(d)}\right] = \mathbb{E}\left[\frac{1}{(1+\alpha)+o(1)}\right] \to \frac{1}{1+\alpha}.$$

Hence, the difference between the expected utility of the James-Stein strategy and the plug-in Merton strategy is

$$V(\hat{\pi}^{JS,\pi^0}) - V(\hat{\pi}) = \mathbb{E}(V(\hat{\pi}^{JS,\pi^0}|\mu)) - \mathbb{E}(V(\hat{\pi}|\mu)) = \frac{T(d-2)^2}{2(1+\alpha)dt_{est}} + o(d)$$

$$= d\frac{T}{2(1+\alpha)t_{est}} + o(d).$$

We see that the expected utility from the plug-in Merton strategy can be increased linearly in the number d of stocks through James-Stein shrinkage. The gain due to the James-Stein strategy compared to the plug-in strategy is increasing in the time horizon and decreasing in the length of the estimation interval t_{est}.

Next, we consider the difference between the expected utility of a James-Stein strategy and the expected utility of an informed investor that uses the optimal strategy. Using (3.2),

$$V(\hat{\pi}^{JS,\pi^0}) - V(\pi^*) = \mathbb{E}(V(\hat{\pi}^{JS,\pi^0}|\mu)) - \mathbb{E}(V(\pi^*|\mu))$$

(4.3)
$$= -\left(1 - \frac{1}{1+\alpha}\right)\frac{Td}{2t_{est}} + o(d)$$

$$= -\frac{\alpha}{1+\alpha}\frac{Td}{2t_{est}} + o(d) = -\frac{\mathbb{V}ar(\mu_1)}{\eta^2(1-\rho) + \mathbb{V}ar(\mu_1)}\frac{Td}{2t_{est}} + o(d).$$

Comparing this to (3.2), we see that the loss in expected utility due to using the James-Stein strategy rather than the optimal strategy is $\frac{\alpha}{1+\alpha}$ times smaller than the loss due to using the plug-in strategy rather than the optimal stratey.

How well does the James-Stein strategy perform in absolute terms? If $\rho > 0$ and $\mathbb{V}ar(\mu_1) > 0$, combining (4.3) with Example 2.1 gives

$$V(\hat{\pi}^{JS,\pi^0}) = \mathbb{E}(V(\hat{\pi}^{JS,\pi^0}|\mu)) = d\frac{T}{2}\left(\frac{\mathbb{V}ar(\mu_1)}{\eta^2(1-\rho)} - \frac{\mathbb{V}ar(\mu_1)}{(\eta^2(1-\rho) + \mathbb{V}ar(\mu_1))\,t_{est}}\right) + o(d)$$

$$= d\frac{T}{2\eta^2(1-\rho)}\mathbb{V}ar(\mu_1)\left(1 - \frac{1}{\left(1+\frac{\mathbb{V}ar(\mu_1)}{\eta^2(1-\rho)}\right)t_{est}}\right) + o(d).$$

In particular, if $t_{est} \geq 1$ then $V(\hat{\pi}^{JS,\pi^0}) \to \infty$, i.e. the James-Stein strategy does not only perform better than the plug-in Merton strategy but the expected utility even converges to $+\infty$ as in the situation when the parameters are known.

If $\rho = 0$ and $\mathbb{P}(\mu_1 \neq r) > 0$ then

$$V(\hat{\pi}^{JS,\pi^0}) = d\frac{T}{2\eta^2}\left\{\mathbb{V}ar(\mu_1)\left(1 - \frac{1}{\left(1+\frac{\mathbb{V}ar(\mu_1)}{\eta^2}\right)t_{est}}\right) + [\mathbb{E}(\mu_1 - r)]^2\right\} + o(d).$$

Again $t_{est} \geq 1$ is sufficient for $V(\hat{\pi}^{JS,\pi^0}) \to \infty$.

5 L_1-restricted Strategy — Lasso

In this section we consider the effect of bounding the L_1-norm of the investment weights in the stocks. More precisely, we suppose that $\hat{\pi}$ satisfies $\|\hat{\pi}\|_1 = \sum_{i=1}^{d} |\hat{\pi}_i| \leq c$ for a constant $c \geq 0$. This bound does not include a bound on the investment in the risk-free asset. In a regression context this kind of constrained optimisation is called 'lasso' (Tibshirani, 1996).

The expected utility with this bound is

$$V(\hat{\pi}|\mu) = \log(X_0) + rT + T\mathbb{E}\left[\hat{\pi}^\top(\mu - r1) - \frac{1}{2}\hat{\pi}^\top\Sigma\hat{\pi}\right]$$

$$\geq \log(X_0) + rT - T\mathbb{E}\left\{c\max_i |\mu_i - r| + \frac{c^2}{2}\max_{i,j}|\Sigma_{ij}|\right\}.$$

Therefore, if $\max_i |\mu_i - r|$ and $\max_{i,j}|\Sigma_{ij}|$ are bounded then the utility of an L_1-restricted strategy cannot deteriorate to $-\infty$ as $d \to \infty$.

Remark 5.1. Tibshirani (1996, Section 5) showed that Lasso estimates can be understood as Bayes posterior mode under independent double exponential priors for the unknown parameters in the linear model (assuming diagonal covariance matrix).

In the following we first clarify the structure of the optimal L_1-constrained investment strategies for known parameters and a specific covariance matrix Σ. After that we analyse the effect of plugging estimators of the drift μ into this strategy. Particularly, we are interested in how the strategy diversifies as the number of assets gets large.

5.1   The Optimal Strategy with Known Drift

We consider the optimisation problem for a special structure of the covariance matrix, i.e. we assume $\Sigma = \sigma\sigma^\top = \eta^2(\rho11^\top + (1-\rho)I)$ for constants $\eta > 0$ and $1 \geq \rho \geq 0$. As we have remarked after (2.1) maximising $V(\pi|\mu)$ is equivalent to minimsing $\|\sigma^\top\pi - \sigma^{-1}(\mu - r1)\|_2^2$ which will be used in the following theorem.

Theorem 5.2. *Suppose $\Sigma = \sigma\sigma^\top = \eta^2(\rho11^\top + (1-\rho)I)$. Consider the optimisation problem*

$$(5.1) \qquad \begin{cases} \|\sigma^\top\pi - \sigma^{-1}(\mu - r1)\|_2^2 \to \min \\ \|\pi\|_1 \leq c \end{cases}$$

The unconstrained solution is

$$\pi^\dagger = \frac{1}{\eta^2(1-\rho)}\left((\mu - r1) - \frac{\rho}{1-\rho+\rho d}11^\top(\mu - r1)\right),$$

which is a solution to the constrained optimisation problem (5.1) if $\|\pi^\dagger\|_1 \leq c$. Otherwise, the unique solution to (5.1) is

$$\pi_i^* = \begin{cases} \frac{1}{\eta^2(1-\rho)}(\mu_i - r - a^+), & \text{if } \mu_i - r > a^+, \\ \frac{1}{\eta^2(1-\rho)}(\mu_i - r - a^-), & \text{if } \mu_i - r < a^-, \\ 0, & \text{otherwise,} \end{cases}$$

where $a^+ \geq a^-$ are solutions to the equations

$$a^+ + a^- = 2\eta^2\rho 1^\top\pi^*, \quad \|\pi^*\|_1 = c.$$

Such a^+ and a^- always exist and are unique.

A proof for this theorem is given in Appendix D. The main use of the above theorem is the clarification of the structure of the solution. If $\rho = 0$ then $a^+ = -a^-$ and we get the following simpler solution. In a regression context this simpler solution has been mentioned in (Tibshirani, 1996, Section 2.2).

Corollary 5.3. *Suppose* $\Sigma = \eta^2 I$ *and* $\mu \in \mathbb{R}^d$ *with* $|\mu_1 - r| > |\mu_2 - r| > \ldots > |\mu_d - r|$. *Then* $\pi^\dagger = \frac{1}{\eta^2}(\mu - r1)$ *is a solution to (5.1) if* $\|\pi^\dagger\|_1 \leq c$. *Otherwise, the unique solution to (5.1) is*

$$\pi^* := \frac{1}{\eta^2}(\operatorname{sgn}(\mu_1 - r)(|\mu_1 - r| - a), \ldots, \operatorname{sgn}(\mu_k - r)(|\mu_k - r| - a), 0, \ldots, 0)^\top,$$

where sgn *is the sign function,*

$$k = \min\left\{l \in \{1, \ldots, d\} : c \leq \frac{1}{\eta^2}\sum_{i=1}^{l}(|\mu_i - r| - |\mu_{l+1} - r|)\right\},$$

$a = \frac{1}{k}\left[\eta^2 c - \sum_{i=1}^{k}|\mu_i - r|\right]$ *and* $\mu_{d+1} = r$.

5.2   Optimal Strategy with Estimated Drift

We now analyse the optimal L_1-constrained strategy with plug-in estimator $\hat{\mu}$ of the drift vector μ, that is we plug $\hat{\mu}$ into (5.1). In particular, we determine the probability of investing in a fixed number of stocks when the number of stocks d in the market goes to infinity. We find that the optimal strategy does not diversify quickly.

We consider the independence case $\Sigma = \eta^2 I$ first. In this case we can get some explicit results using results from extreme value theory.

Assume that $\mu_1, \ldots, \mu_d$ are i.i.d. normally distributed and that conditionally on those, each μ_i can be estimated by an independent normally distributed estimator $\hat{\mu}_i$. Thus, $\hat{\mu}_1, \ldots, \hat{\mu}_d$ are independent and identically distributed.

Theorem 5.4. *Suppose* $\Sigma = \eta^2 I$ *and suppose we use the* L_1-*constrained optimal strategy with plug-in estimators* $\hat{\mu}_i$ *for* μ_i, *where* $\hat{\mu}_1, \ldots, \hat{\mu}_d$ *are independent and identically normally distributed. We use the threshold* $c = \alpha c_d$, *where* $c_d > 0$ *are norming constants specified in Theorem E.1 for a folded normal distribution* $FN(\mathbb{E}(\hat{\mu}_1 - r), \mathbb{V}ar(\hat{\mu}_1))$. *Then*

$$\#\{i : \pi_i^* \neq 0\} \overset{\mathcal{L}}{\to} K + 1 \quad (d \to \infty),$$

where $\overset{\mathcal{L}}{\to}$ *denotes convergence in distribution and* K *is a Poisson distribution with expected value* $\alpha\eta^2$.

Figure 1 contains some plots of the constants c_d.

One of the main ideas for the proof of Theorem 5.4, provided in Appendix E, is that the random variables $Z_i := |\hat{\mu}_i - r|$, $i = 1, \ldots, d$ are i.i.d. and follow a folded normal distribution. By Corollary 5.3, the number of stocks one invests in depends on the spacings of the upper order statistics of Z_i which can be analysed via extreme value theory. The norming constants are the scaling constants for extreme values of a folded normal distribution, see Theorem E.1.

Motivated by the above result, we compare, crudely via their means, the approximation $\#\{i : \pi_i^* \neq 0\} \approx K + 1$, where K is a Poisson$(\eta^2 c/c_d)$ distribution, to
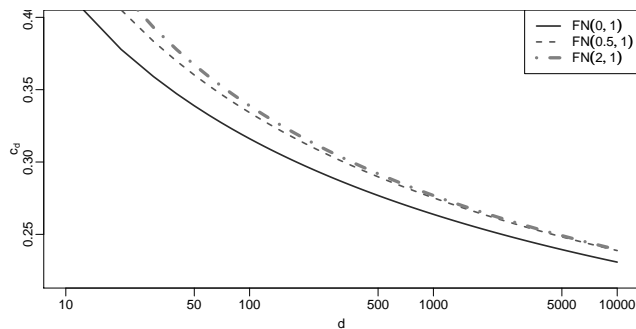
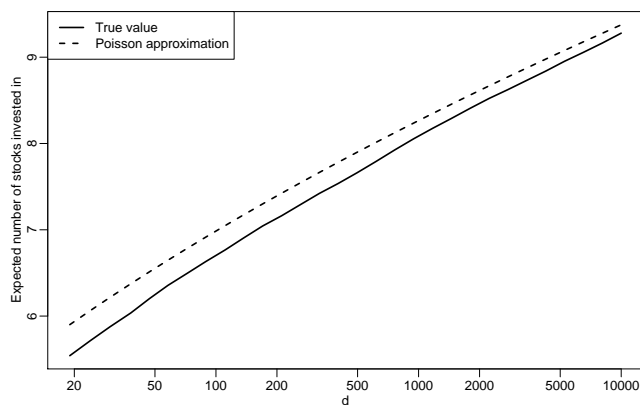Figure 1: Plot of the constants c_d in Theorem 5.4, see also Theorem E.1.

Figure 2: Mean number of stocks the L_1-strategy is investing in when $\rho = 0$, $\eta = 1$, $\hat{\mu}_1, \ldots, \hat{\mu}_d \sim N(0.5, 1)$ and $c = 2$. The Poisson approximation is the mean of $K + 1$ where $K \sim \text{Poisson}(\eta^2 c/c_d)$. The true value was obtained through simulations.

simulation results in Figure 2. The approximation seems to be good. Furthermore, we see that diversification happens very slowly in the L_1-constrained situation.

For $\Sigma = \eta^2((1 - \rho)I + \rho 11^\top)$ with $\rho > 0$ the following theorem gives bounds on the number of stocks one invests in as $d \to \infty$.

Theorem 5.5. *Suppose $\Sigma = \eta^2((1 - \rho)I + \rho 11^\top)$ with $\rho > 0$ and suppose we use the L_1-constrained optimal strategy with plug-in estimators $\hat{\mu}_i$ for μ_i, where $(\hat{\mu}_1, \ldots, \hat{\mu}_d)^\top \sim N(\xi 1, \Sigma/t_{est})$ for some $\xi \in \mathbb{R}$. If we choose the threshold $c = \alpha(2\log(d))^{-1/2}$ then for even k,*

$$\lim_{d \to \infty} \mathbb{P}(\#\{i : \pi_i^* \neq 0\} > k) \leq 2\mathbb{P}(K \geq k/2 + 1),$$

where K is a Poisson distribution with expected value $\alpha\eta\sqrt{1 - \rho}\sqrt{t_{est}}$.

Again, using the resulting Poisson approximation, one can see that the diversification happens only slowly.

6   Other Restricted Strategies

6.1 L_0-restricted Strategies

Consider funds that are restricted to invest in certain sectors and/or certain countries. This means that these funds use strategies that only invest in a fixed finite number of stocks. After renumbering the stocks this means that $\hat{\pi}_j = 0$ for $j > k$, where $k \in \mathbb{N}$ is some fixed constant. Then, if one uses the resulting Merton-strategy with plug-in estimators based on only these k stocks one does not suffer from the deterioration of performance as $d \to \infty$.

6.2 L_2-restricted Strategies

Imposing an L_2 restriction on its own does not guarantee that the expected utility does not degenerate to $-\infty$ as $d \to \infty$. Indeed, consider $\Sigma = (1 - \rho)I + \rho 11^\top$ and $\mu - r1 = \xi 1$ for some $\rho \neq 0$ and some $\xi \neq 0$. The strategy $\pi = \frac{c}{\sqrt{d}}1$ satisfies the L_2-restriction $\|\pi\|_2 \leq c$, but has an expected utility that degenerates to $-\infty$ for $d \to \infty$. Indeed, $\pi^\top\Sigma\pi = (1 - \rho)c^2 + \rho c^2 d$ and $\pi^\top(\mu - r1) = c\xi\sqrt{d}$. Thus the corresponding expected utility converges to $-\infty$ as $d \to \infty$.

6.3   The $1/d$-Strategy

Consider the strategy that invests the same amount into all stocks, i.e. $\pi^{\alpha_*/d} = \frac{\alpha_*}{d}1$ for some $\alpha_* > 0$. Assuming that $\mu_1, \ldots, \mu_d$ are i.i.d., as $d \to \infty$,

$$(6.1) \qquad V(\pi^{\alpha_*/d}) = \mathbb{E}(V(\pi^{\alpha_*/d}|\mu)) \to \log(X_0) + rT + T\alpha_*\mathbb{E}[\mu_1 - r] - \frac{T}{2}\eta^2\alpha_*{}^2\rho.$$

No degeneration occurs.

6.4   Equal Weighting of the most Extreme stocks (EWE)

In this section we suggest a strategy that improves upon the $1/d$-strategy by investing in the stocks that have the most extreme variance-adjusted returns. As it turns out, we will get potentially a better mean log-return, without a higher variance of the return. In this section we will assume that Σ is known.

Consider the strategy

$$\pi_{k_i}^{\text{EWE}} = \frac{\alpha_*}{\beta d} \operatorname{sgn}(\hat{a}_{k_i}) \mathbb{I}(i \leq \beta d), \quad i = 1, \ldots, d,$$

where $\hat{a}_i = \frac{\hat{\mu}_i - r}{\Sigma_{ii}}$, $\alpha_* > 0$ and $\beta \in (0, 1)$ are constants, and k_i are such that $|\hat{a}_{k_1}| > |\hat{a}_{k_2}| > \cdots > |\hat{a}_{k_d}|$. As this strategy uses an Equal Weighting of the most Extreme stocks, we call it the *EWE-strategy*.

Remark 6.1. This strategy can be seen as satisfying an L_p restriction for all $0 \leq p \leq \infty$, namely $\|\pi^{\text{EWE}}\|_p \leq \alpha_*(\beta d)^{1/p-1}$ for $0 < p \leq \infty$ and $\|\pi^{\text{EWE}}\|_0 \leq \beta d$.

Theorem 6.2. *Suppose* $\Sigma = \eta^2 \left[(1 - \rho)I + \rho 11^\top\right]$. *Then*

$$(6.2) \quad V(\pi^{EWE}|\mu) \geq \log(X_0) + rT + T\mathbb{E}[(\mu - r1)^\top \pi^{EWE}] - \frac{T\eta^2\alpha_*^2}{2\beta d}\left[1 + \rho(\beta d - 1)\right].$$

Furthermore, suppose that $\mu_1, \ldots, \mu_d$ *are i.i.d. and that, conditionally on* $\mu_1, \ldots, \mu_d$, *the estimator satisfies* $\hat{\mu} \sim N(\mu, \frac{1}{t_{est}}\Sigma)$. *This implies that* $\hat{\mu}_i \sim N(\mu_A + \mu_i, \frac{1}{t_{est}}(1 - \rho)\eta^2), i = 1, \ldots, d$ *are i.i.d. conditionally on* $\mu^A \sim N(\mu_A^*, \frac{1}{t_{est}}\rho\eta^2)$ *and* $\mu_1, \ldots, \mu_d$. *Then*

$$(6.3)$$

$$\lim_{d \to \infty} V(\pi^{EWE}) = \lim_{d \to \infty} \mathbb{E}(V(\pi^{EWE}|\mu)) \geq \log(X_0) + rT - \frac{T}{2}\eta^2{\alpha_*}^2\rho$$

$$+ T\frac{\alpha_*}{\beta}\mathbb{E}[\mathbb{P}(\hat{a}_1 > c|\mu_1, \mu_A)(\mu_1 - r) - \mathbb{P}(\hat{a}_1 < -c|\mu_1, \mu_A)(\mu_1 - r)],$$

where c *is such that* $\mathbb{P}(|\hat{a}_1| > c|\mu_A) = \beta$.

A proof is given in Appendix F.

Comparing (6.1) and (6.3), one sees that the variance penalty of strategies, $\frac{T}{2}\alpha_*^2\rho$ for the $1/d$-strategy is not larger than the variance penalty for the EWE-strategy. Furthermore, the EWE-strategy is superior to the $1/d$-strategy in the limit if

$$g(\beta) = \beta^{-1}\mathbb{E}[\mathbb{P}(\hat{a}_1 > c|\mu_1, \mu_A)(\mu_i - r) - \mathbb{P}(\hat{a}_1 < -c|\mu_1, \mu_A)(\mu_i - r)] - \mathbb{E}[\mu_i - r] \geq 0,$$

which essentially is the difference between the mean log-returns of the two strategies. Some simulation based evaluations of the previous expression can be seen in Figure 3. We used the following simple setup: $\mu_i \sim N(\mu_0, .05^2)$, $\eta^2/t_{test} = 0.1^2$. The EWE strategy is superior in many situations. Especially, if β, the proportion of stocks one is investing in, is small.

7 Some Examples

In this section, we investigate the performance of the various trading strategies by looking at their theoretical performance in a specific setup (Section 7.1) and by looking at their real life performance in an out-of-sample test (Section 7.2).

Both will be based on a data set which consists of daily returns of 373 stocks that where part of the S&P 500 index on the 1st of January 2006 and had daily returns for all trading days between 2001 and 2008. This choice might lead to some selection bias as we are choosing companies which existed for several years and survived the financial crisis in 2008. The out-of-sample results in Section 7.2 might therefore be slightly optimistic.
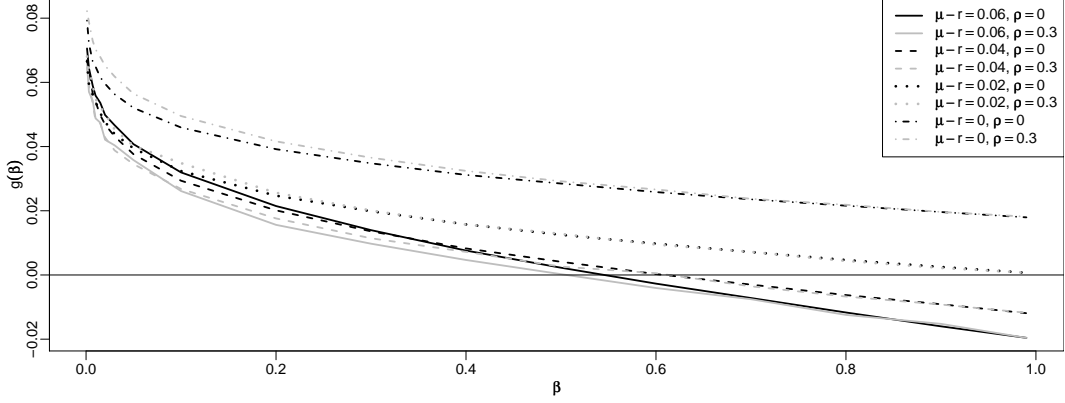
Figure 3: Comparison of the EWE-strategy with the $1/d$-strategy. Plot of $g(\beta)$ against β. If $g(\beta) > 0$ then the EWE-strategy is superior.

Throughout this section we assume $X_0 = 1$. There were $n = 2011$ trading days in these 8 years. We assume a yearly interest rate of $r = 0.02$.

We will investigate how the strategies perform as the number d of stocks varies. For this we use a specific random ordering of the stocks and will allow the strategies to invest in the first d stocks of this ordering.

Based on the observed stock prices at the time points $0, \Delta, 2\Delta, \ldots, (n-1)\Delta$, we use the following unbiased estimators of μ and Σ:

$$
(7.1) \qquad
\begin{aligned}
\hat{\mu}^{\text{data}} &= \frac{1}{\Delta}\hat{\xi} + \frac{1}{2}\operatorname{diag}(\widehat{\Sigma}^{\text{data}}), \\
\widehat{\Sigma}^{\text{data}}_{\mu,\nu} &= \frac{1}{\Delta(n-2)}\sum_{i=0}^{n-2}\left[R_\mu(i) - \hat{\xi}_\mu\right]\left[R_\nu(i) - \hat{\xi}_\nu\right]
\end{aligned}
$$

for $\mu, \nu = 1, \ldots, d$, where $R_\mu(i) = \log\left(\frac{S_\mu((i+1)\Delta)}{S_\mu(i\Delta)}\right)$, $\hat{\xi}_\mu = \frac{1}{n-1}\sum_{i=0}^{n-2} R_\mu(i)$.

7.1   Theoretical Utility - Simulation Study

In this subsection, we consider the theoretical performance of the strategies mentioned in the previous sections. For strategies for which we do not have explicit formulas for the expected utilities (e.g. for unknown Σ), we will use simulation to compute their expected utility.

We use the data described above to define the model we will simulate from. We set the true return vector to $\mu = \hat{\mu}^{\text{data}}$ and we let the covariance matrix Σ be given by $\Sigma = \hat{\sigma}^2(\frac{1}{5}11^\top + \frac{4}{5}I)$, where $\hat{\sigma}^2 = \frac{1}{d}\sum_{i=1}^{d}\widehat{\Sigma}^{\text{data}}_{ii}$. A similar analysis, not reported here, for $\Sigma = \hat{\sigma}^2 I$ and $\Sigma = \hat{\Sigma}^{\text{data}}$ lead to similar results.

All of the following is conditional on these choices of μ and Σ. In particular, the S&P 500 data set will not be used any further in this subsection and, in contrast to some of the examples in previous sections, μ will not be random.

We consider up to three degrees of information about the model parameters that the strategy π can use:

- μ, Σ known,

- Σ known, μ has to be estimated and

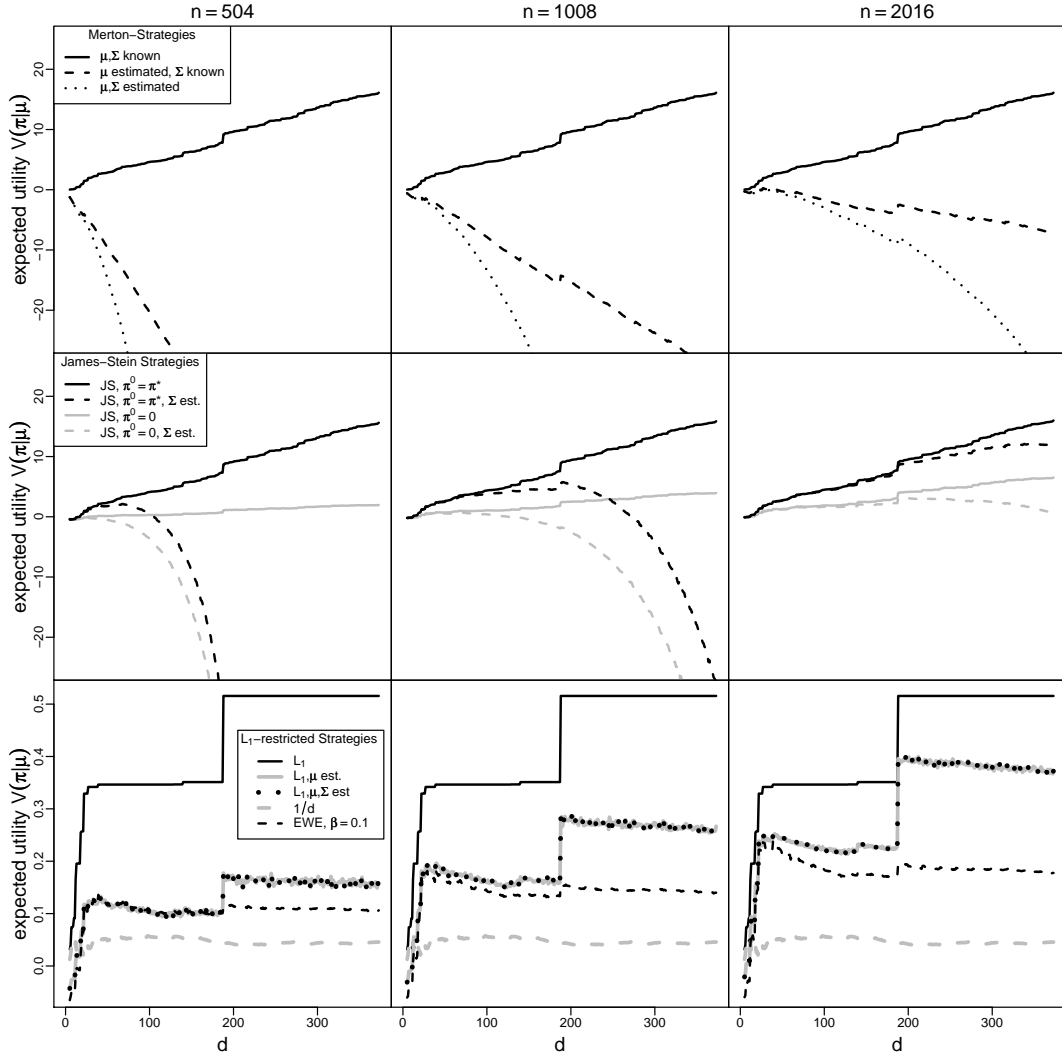- both μ, Σ have to be estimated.

15

Figure 4: Comparison of the expected utility of several strategies. The first row contains Merton strategies (Section 3), the second row James-Stein type strategies (Section 4), the third row L_1-restricted strategies (Sections 5, 6).

Figure 4 shows the expected utility $V(\pi|\mu)$ when using roughly $T = 1$ year as investment period. In the first two information scenarios, we have used our explicit results for the expected utility $V(\pi|\mu)$ for the Merton strategy with known parameters (2.3), the plug-in Merton (3.2) and the James-Stein strategy (4.1). For the 1/d-strategy, an explicit formula can be obtained by using (2.1). For other strategies π, we simulate n=504, 1008, or 2016 past observations spaced one trading day apart (corresponding to 2, 4 or 8 years) and estimate the parameters μ and Σ analogously to the estimation in (7.1). We repeat this 1000 times and based on the estimated parameters obtain estimated strategies $\pi^{(1)}, \ldots, \pi^{(1000)}$. We then approximate $V(\pi|\mu)$ in (2.1) by the Monte-Carlo estimate

$$\log(X_0) + rT + T\frac{1}{1000}\sum_{i=1}^{1000}\left[(\pi^{(i)})^\top(\mu - r1) - \frac{1}{2}(\pi^{(i)})^\top\Sigma\pi^{(i)}\right].$$

The first row of Figure 4 compares the expected utility for the Merton strategy with and without plug-in estimators. If both μ and Σ are known then the expected

utility is increasing roughly linearly in the number of stocks. This illustrates the theoretical results from Section 2. When μ has to be estimated then, consistently with the theoretical results of Section 3, the expected utility is decreasing linearly in the number of assets. Having to estimate Σ in addition to μ makes the situation even worse. The loss in utility is now non-linear. Furthermore, we see that having an increased sample size for estimation improves the estimated strategies. Consistently with (3.2), the loss in expected utility through the estimation of μ alone is halved when the observation period is doubled.

The second row of Figure 4 shows the expected utility for different James-Stein-type strategies. The best overall performance is obtained by using a James-Stein estimator that shrinks towards the optimal strategy, i.e. $\pi^0 = \pi^\star$. Compared to the optimal utility of the Merton strategy with known parameters in the first row, only very little utility is lost. Obviously, in practice this strategy cannot be used, since the true parameter is unknown. The expected utility of a James-Stein estimator which shrinks towards $\pi^0 = 0 \in \mathbb{R}^d$ performs less well than if one shrinks towards the optimal strategy. A James-Stein estimator does improve the performance of the portfolio significantly when compared to the Merton plug-in strategies. When we have to estimate the covariance matrix Σ for the shrinkage, then the expected utility is decreasing very rapidly in the number of stocks.

Finally, the last row of Figure 4 compares different strategies that all respect the L_1-constraint $\|\pi\|_1 \leq 1$: the optimal L_1-restricted strategy (with combinations of known/estimated parameters), the equally weighted portfolio, the EWE-strategy described in Section 6.4 (with $\beta = 0.1$ and $\alpha_* = 1$) and the equally weighted portfolio ($1/d$-strategy). The best performance can be achieved by using an L_1-constraint when the parameters are known. In this situation the optimal portfolio only contains a very small number of stocks. The expected utility therefore only changes if a new asset becomes available which has a much better expected return and volatility than those which were previously available. That is why we observe a step-function type of behaviour. If the parameters μ and Σ have to be estimated in the L_1-constraint case the expected utility is obviously worse than in the case with known parameters. The expected utility stays positive in the majority of situations which was not the case in the plug-in-Merton case or the JS-case when both parameters had to be estimated. The difference between having to estimate Σ or not is marginal in the L_1-constrained case. The $1/d$-strategy has a very stable performance. It outperforms the L_1-constrained portfolio and the EWE-strategy when parameters have to be estimated and the number of available stocks is very small. Otherwise, the $1/d$-strategy is strictly dominated by the other L_1-restricted strategies. The EWE-strategy performs well and guarantees a stable performance.

As the sample size n increases, the performance of the estimated optimal L_1-constrained strategy approaches the performance of the optimal L_1-constrained strategy with known parameters. Furthermore, the expected utility of the EWE-strategy improves. The $1/d$-strategy and the L_1-strategy with known parameters do not use any estimated parameters and thus their performance is not affected by n.

7.2   Out-of-Sample Performance

Based on the 8 years of S&P 500 data, we now consider the out-of-sample performance of the strategies. We estimated the covariance matrix and the mean returns on data from 4 years and run the investment strategy for the following year.
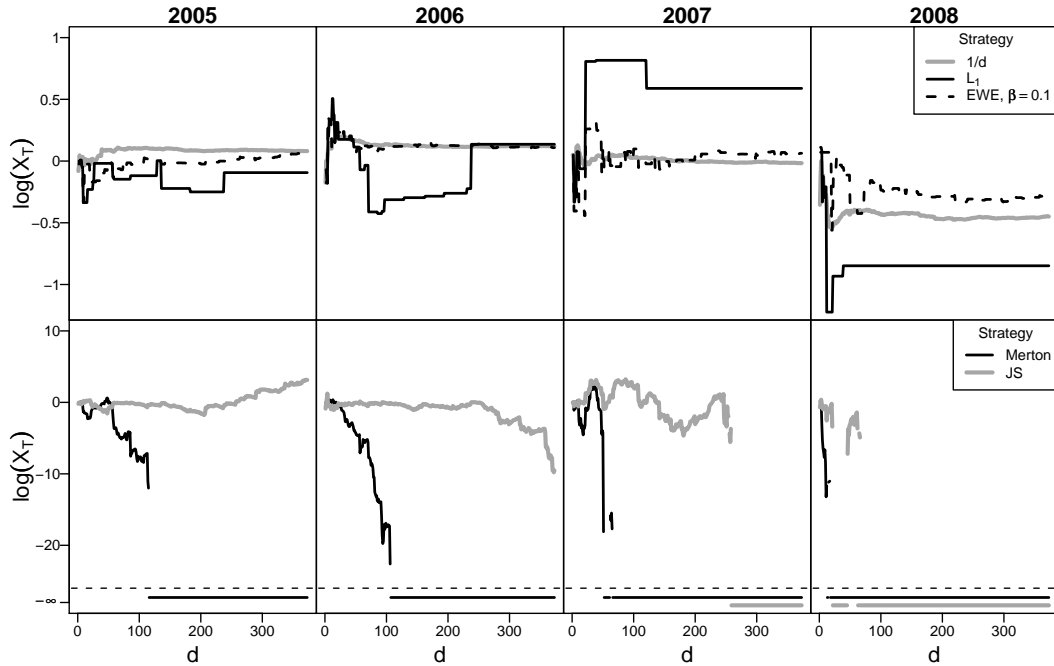
Figure 5: Out-of-sample performance for several strategies. Note the different scale for the utility for the L_1-constrained strategies in the first row and the Merton and the James-Stein strategy in the second row. In the latter case the observed utility can be $-\infty$.

As the Merton model is a continuous time model we need to approximate it by trading in discrete time. We do this by trading once per trading day. The discrete trading may lead to negative wealth at a time point t, in which case we set the utility $\log(X_T)$ to $-\infty$. As discussed in Rogers (2001), to avoid going bankrupt, the investor trading at discrete time steps has to restrict the investment proportions π to be nonnegative and to satisfy $\sum_{i=1}^{d} \pi_i \leq 1$. Then the loss due to discrete time trading is small.

Results can be seen in Figure 5, where the number d of available stocks is plotted against the utility $\log(S_T)$ with T being one year.

The plug-in Merton strategy performs badly in all years, being equal to $-\infty$ very quickly as d increases. The James-Stein strategy (which uses $\pi^0 = 0$) shows results with a very high volatility, and shows a utility of $-\infty$ in 2007 and 2008 for large d.

The L_1-optimal strategy with plug-in-estimators has a large volatility. A good performance in 2007 is destroyed by the extremely poor performance in 2008. This volatile behaviour is due to the slow diversification of the L_1-strategy.

The EWE-strategy and the $1/d$-strategy are not very much affected by the number of available stocks. Overall, the EWE-strategy performs best, narrowly beating the $1/d$-strategy. This is particularly evident in 2008, the year of a financial crisis.

This is no contradiction to the results for the theoretical expected utility in Section 7.1, where the optimal L_1-restricted strategy with estimated parameters was doing best. In this subsection, we consider a real-data situation. Several assumptions of the theoretical model will be violated. In particular, the assumption that μ and Σ are constant over time probably not be true and will have a heavy impact when, as in the last column of Figure 5, pre-crisis data (2004-2007) are used to trade in the crisis year 2008. Not surprisingly, the strategies that do best (1/d-strategy,

EWE-strategy) are the ones that are least dependent on the estimates.

8   Discussion

The L_1-constrained portfolio selection problem is an example of an optimal investment problem with convex constraints on the investment strategies. Such problems were analysed by Cvitanic and Karatzas (1992) using duality methods. In principle, one could use their approach as well to tackle the L_1-constraint situation and consider general existence results. Since we find that the primal formulation already allows the derivation of analytic formulae for the optimal strategies we prefer the direct approach here.

In the unconstrained Merton portfolio optimisation problem with logarithmic utility the optimal investment strategy is given by $\Sigma^{-1}(\mu - r1)$ which is, up to a scaling factor, the optimal solution to the one-period mean-variance optimisation problem with one riskless and d risky assets (usually referred to as Tobin model or generalised Markowitz model). In what sense do our results carry over to the Markowitz/Tobin situation? In a classical one-period mean-variance model an objective function could have the form $f(\pi) = \mathbb{E}[\pi^\top R] - \frac{1}{2}\mathbb{V}ar(\pi^\top R)$ where R models the random return vector and π as before models the investment weights. We see immediately, that by writing $\mathbb{V}ar(\pi^\top R) = \pi^\top \Sigma \pi$ we derive the optimal strategy in the Merton and in the static mean-variance framework by solving the same quadratic optimisation problem as long as π is deterministic. Therefore, our results for the optimal strategies (also for the L_1-constrained strategies) carry over to the static mean-variance optimisation problem. However, the expressions for the expected utility are different in the Merton and the mean-variance situation. Consider e.g. the estimator $\hat{\pi}$ which has to be used when model parameters are unknown. In the Merton context we consider the expression

$$V(\hat{\pi}|\mu) = \log(X_0) + rT + T\mathbb{E}\left[\hat{\pi}^\top(\mu - r1) - \frac{1}{2}\hat{\pi}^\top \Sigma \hat{\pi}\right],$$

whereas in the mean-variance context we consider

$$f(\hat{\pi}) = \mathbb{E}[\hat{\pi}^\top R] - \frac{1}{2}\mathbb{V}ar(\hat{\pi}^\top R) \neq \mathbb{E}[\hat{\pi}^\top R] - \frac{1}{2}\mathbb{E}(\hat{\pi}^\top \Sigma \hat{\pi}).$$

These two expressions do not coincide and therefore the loss due to having to estimate the model parameters in the Merton context and the static mean-variance context is not directly comparable.

Our model does not include any transaction costs. Analysing the effect of estimation in a model that includes transaction costs would be of interest. However, no explicit results for the strategy and no analytic results for the expected utility are known in these models. For proportional transaction costs, Davis and Norman (1990) have shown that the optimal strategy can be expressed in terms of a *no-trading-interval*: While the proportion of wealth in the share stays within this interval it is optimal not to trade. Otherwise just enough trading is done to push this proportion into this interval. This interval usually contains the Merton proportion. From a mathematical point of view one has to solve a free boundary problem and even in one dimension (meaning one risky asset) there are no analytic results for the boundaries of this interval. There are numerical algorithms available which could be applied for higher dimensions. But those algorithms suffer the curse of dimensionality and it will be effectively impossible to solve this problem even for known parameters for high-dimensional problems.

9   Conclusions

Our theoretical results for known Σ about the behaviour of the utility when the number of stocks d goes to ∞ can be summarised as follows. The plug-in Merton strategy degenerates to $-\infty$ in realistic cases. The James-Stein type strategy does better than plug-in Merton. In particular, the James-Stein strategy does not degenerate in many realistic scenarios. The utility of L_1-constrained strategies can never degenerate to $-\infty$. The L_1-constrained optimal strategy with plug-in estimators diversifies only slowly, i.e. it invests in few stocks only.

The specific situation considered in Section 7.1 covers, via simulations, the case when Σ needs to be estimated as well. In this case we observed the following. The expected utility of the plug-in Merton and the James-Stein strategy degenerates to $-\infty$. For the optimal L_1-strategy with plug-in estimators, the difference between having to estimate Σ or not is negligible. The EWE-strategy (Equal Weighting of Extreme stocks) performs better than the $1/d$-strategy. The optimal L_1-strategy with plug-in estimators does better than the EWE-strategy for large d.

However, in the out-of-sample tests, the EWE-strategy and the $1/d$-strategy do better than the optimal L_1-strategy. This could be explained by the L_1-strategy investing only in few stocks, whereas the EWE-strategy and the $1/d$-strategy diversify much better, investing in more stocks and weighting them all equally. This seems to be important when there are structural changes in the market, such as in the financial crisis of 2008. An explanation why the EWE-strategy seems to outperform the $1/d$-strategy is that the former tries to pick stocks with a higher return, while still diversifying reasonably.

A   Inverse of a Specific Covariance Matrix

The following lemma can be derived by solving $I = \Sigma(aI + b11^{\top})$ for a and b.

Lemma A.1. *Suppose* $\Sigma = \eta^2 \left[(1-\rho)I + \rho 11^{\top}\right]$, *where* $0 \leq \rho \leq 1$ *and* $1 = (1,\ldots,1)^{\top} \in \mathbb{R}^d$. *Then*

$$\Sigma^{-1} = \frac{1}{\eta^2(1-\rho)}\left[I - \frac{\rho}{1-\rho+\rho d}11^{\top}\right] \text{ for } \rho < 1.$$

B   Expected Utility of the Merton Strategy

This section contains the proof of (2.4) in Example 2.1. Using Lemma A.1, the quadratic term in (2.3) can be written as

$$(\mu - r1)^{\top}\frac{1}{\eta^2(1-\rho)}\left[I - \frac{\rho}{1-\rho+\rho d}11^{\top}\right](\mu - r1)$$

$$= \frac{1}{\eta^2(1-\rho)}\left(\|\mu - r1\|_2^2 - \frac{\rho}{1-\rho+\rho d}\left[(\mu - r1)^{\top}1\right]^2\right)$$

$$= \frac{1}{\eta^2(1-\rho)}\left(\sum_{i=1}^{d}(\mu_i - r)^2 - \frac{\rho}{1-\rho+\rho d}\left[\sum_{i\neq j}(\mu_i - r)(\mu_j - r) + \sum_{i=1}^{d}(\mu_i - r)^2\right]\right).$$

The expected value of this is

$$\frac{1}{\eta^2(1-\rho)}\left[d\mathbb{E}[(\mu_1-r)^2] - \frac{\rho}{1-\rho+\rho d}\left[d(d-1)(\mathbb{E}(\mu_1-r))^2 + d\mathbb{E}[(\mu_1-r)^2]]\right]\right]$$

$$= \frac{d}{\eta^2(1-\rho)(1-\rho+\rho d)}\left[(1-\rho+\rho d-\rho)\mathbb{E}[(\mu_1-r)^2] - (\rho d-\rho)(\mathbb{E}(\mu_1-r))^2\right]$$

$$= \frac{d\left[(1+\rho(d-2))\mathbb{V}ar[\mu_1] + (1-\rho)(\mathbb{E}(\mu_1-r))^2\right]}{\eta^2(1-\rho)(1+\rho(d-1))}.$$

C   Proof - James-Stein Strategy

Proof of Theorem 4.1. Using (2.1) yields

$$V(\hat{\pi}^{JS,\pi^0}|\mu) = \log(X_0) + rT - \frac{1}{2}T\mathbb{E}[-2(\hat{\pi}^{JS,\pi^0})^\top\Sigma\pi^* + (\hat{\pi}^{JS,\pi^0})^\top\Sigma\hat{\pi}^{JS,\pi^0}],$$

where $\pi^* = \Sigma^{-1}(\mu - r1)$ is the optimal strategy. Noting that $\hat{\pi}^{JS,\pi^0} = \hat{\pi} - \frac{a\hat{\pi}^0}{(\hat{\pi}^0)^\top\Sigma\hat{\pi}^0}$ with $\hat{\pi}^0 = \hat{\pi} - \pi^0$ and expanding the term in the expectation gives

$$-2\hat{\pi}^\top\Sigma\pi^* + 2a\frac{(\hat{\pi}^0)^\top\Sigma\pi^*}{(\hat{\pi}^0)^\top\Sigma\hat{\pi}^0} + \hat{\pi}^\top\Sigma\hat{\pi} - 2a\frac{(\hat{\pi}^0)^\top\Sigma\hat{\pi}}{(\hat{\pi}^0)^\top\Sigma\hat{\pi}^0} + \frac{a^2}{(\hat{\pi}^0)^\top\Sigma\hat{\pi}^0}.$$

Collecting terms gives

$$V(\hat{\pi}^{JS,\pi^0}|\mu) = V(\hat{\pi}|\mu) + \frac{T}{2}\left\{2a\mathbb{E}\left[(\hat{\pi}-\pi^*)^\top\Sigma\frac{\hat{\pi}^0}{(\hat{\pi}^0)^\top\Sigma\hat{\pi}^0}\right] - a^2\mathbb{E}\left[\frac{1}{(\hat{\pi}^0)^\top\Sigma\hat{\pi}^0}\right]\right\}$$

$$= V(\hat{\pi}|\mu) + \frac{T}{2}a\left\{2\mathbb{E}\left[(\hat{\zeta}-\zeta^*)^\top\frac{\hat{\zeta}}{\hat{\zeta}^\top\hat{\zeta}}\right] - a\mathbb{E}\left[\frac{1}{\hat{\zeta}^\top\hat{\zeta}}\right]\right\},$$

where $\hat{\zeta} = \sigma^\top\hat{\pi}^0 = \sigma^{-1}(\hat{\mu}-r1) - \sigma^\top\pi^0 \sim N(\zeta^*, I/t_{est})$ and $\zeta^* = \sigma^\top(\pi^* - \pi^0)$.

Using Stein's lemma (Stein, 1981) as in (Young and Smith, 2005, p.34) yields

$$\mathbb{E}\left[(\hat{\zeta}-\zeta^*)^\top\frac{\hat{\zeta}}{\hat{\zeta}^\top\hat{\zeta}}\right] = \mathbb{E}\left[\frac{(\hat{\zeta}/t_{est} - \zeta^*/t_{est})^\top\hat{\zeta}/t_{est}}{\hat{\zeta}/t_{est}^\top\hat{\zeta}/t_{est}}\right] = \mathbb{E}\left[\frac{d-2}{\hat{\zeta}^\top\hat{\zeta}}\right]/t_{est}.$$

$t_{est}\hat{\zeta}^\top\hat{\zeta}$ follows a $\chi_d^2(2\lambda)$ distribution with non-centrality parameter $2\lambda = t_{est}\zeta^{*\top}\zeta^* = t_{est}(\pi^* - \pi^0)^\top\Sigma(\pi^* - \pi^0)$. Thus $t_{est}\hat{\zeta}^\top\hat{\zeta}$ can be written as a mixture of central χ_{d+2K}^2 random variables where the mixture variable K follows a Poisson distribution with parameter λ. Thus,

$$\mathbb{E}\left[\frac{1}{t_{est}\hat{\zeta}^\top\hat{\zeta}}\right] = \mathbb{E}\left[\frac{1}{d-2+2K}\right].$$

$\square$

Lemma C.1. *Let K be a Poisson distributed random variable with mean λ and let $q \geq 1$. Then,*

$$\mathbb{E}\left(\frac{1}{q+K}\right) = \frac{1}{q+\lambda} + O\left(\frac{1}{q^2}\right) \qquad as\ q+\lambda \to \infty,$$

where λ may depend on q and $O\left(\frac{1}{q^2}\right)$ does not depend on λ.

Proof. A Taylor expansion of $K \mapsto \frac{1}{q+K}$ around $K = \lambda$ gives

$$(C.1) \qquad \frac{1}{q+K} = \frac{1}{q+\lambda} - \frac{1}{(q+\lambda)^2}(K-\lambda) + \frac{1}{(q+\xi_K)^3}(K-\lambda)^2,$$

for some ξ_K between K and λ. Using that the last term is non-negative and taking expectations, we get

$$\mathbb{E}\left(\frac{1}{q+K}\right) \geq \frac{1}{q+\lambda}.$$

Furthermore,

$$\frac{1}{(q+\xi_K)^3} \leq \frac{1}{(q+\lambda/2)^3} + \mathbb{I}(K < \lambda/2)\frac{1}{q^3}.$$

Using this, taking expectations in (C.1) and using (Houdré, 2002, (1.6)) with $f(x) = -x$, which gives $\mathbb{P}(K < \lambda/2) \leq \exp(\lambda/2 - 3/2\lambda \log(3/2))$, we get

$$\mathbb{E}\left(\frac{1}{q+K}\right) \leq \frac{1}{q+\lambda} + \frac{\lambda}{(q+\lambda/2)^3} + \frac{\lambda}{q^3}e^{\frac{\lambda}{2}-\frac{3}{2}\lambda \log(\frac{3}{2})}.$$

The second term can be bounded above by $8/q^2$ and $\lambda \exp(\lambda/2 - 3/2\lambda \log(3/2))$ is bounded. $\qquad \square$

D   Proof for the Structure of L_1-constrained Strategies

Proof of Theorem 5.2. The solution for the unconstrained case follows directly from (2.2) and Lemma A.1.

Let $\pi_i^+ = \max(\pi_i, 0)$, $\pi_i^- = \max(-\pi_i, 0)$. Hence, $\pi = \pi^+ - \pi^-$. We consider the $(2d)$-dimensional vector $\tilde{\pi}^\top = (\pi^{+\top}, \pi^{-\top})$. Let $1_k = (1, \ldots, 1)^\top \in \mathbb{R}^k$ for $k \in \mathbb{N}$.

The optimisation problem can be written as follows:

$$\begin{cases} \tilde{\pi}^\top \begin{pmatrix} \Sigma & -\Sigma \\ -\Sigma & \Sigma \end{pmatrix} \tilde{\pi} - 2\left((\mu - r1_d)^\top, (r1_d - \mu)^\top\right)\tilde{\pi} \to \min \\ \text{subject to} \\ \begin{pmatrix} I \\ -1_{2d}^\top \end{pmatrix} \tilde{\pi} \geq \begin{pmatrix} 0 \\ -c \end{pmatrix}. \end{cases}$$

The following Kuhn-Tucker conditions characterise uniquely the solution to this problem (Lawson and Hanson, 1995, p.159,160).

$$2\begin{pmatrix} \Sigma & -\Sigma \\ -\Sigma & \Sigma \end{pmatrix}\tilde{\pi} - 2\begin{pmatrix} \mu - r1_d \\ r1_d - \mu \end{pmatrix} - u + 1_{2d}v = 0,$$
$$\tilde{\pi} \geq 0, 1_{2d}^\top\tilde{\pi} \leq c, v \geq 0, u \geq 0,$$
$$u_i = 0 \quad \forall i \in \{1, \ldots, 2d\} : \tilde{\pi}_i > 0,$$
$$v = 0 \text{ or } 1_{2d}^\top\tilde{\pi} = c,$$

where $v \in \mathbb{R}$ and $u \in \mathbb{R}^{2d}$. The first equation is equivalent to

$$(D.1) \qquad 2\eta^2(1-\rho)\begin{pmatrix} \pi \\ \pi \end{pmatrix} + 2\eta^2\rho 1_d 1_d^\top \pi - 2\begin{pmatrix} \mu - r1_d \\ \mu - r1_d \end{pmatrix} + \begin{pmatrix} -u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} 1_d \\ -1_d \end{pmatrix}v = 0,$$

where $u = (u_1^\top, u_2^\top)^\top$, $u_1, u_2 \in \mathbb{R}^d$. It is straightforward to verify that π^* satisfies the Kuhn-Tucker conditions with $v = a^+ - a^-$,

$$u_{1i} = \begin{cases} 0, & \text{if } \pi_i^* > 0, \\ -2(\mu_i - r - a^+) & \text{if } \pi_i^* = 0, \\ 2(a^+ - a^-) & \text{if } \pi_i^* < 0, \end{cases} \quad u_{2i} = \begin{cases} 2(a^+ - a^-) & \text{if } \pi_i^* > 0, \\ 2(\mu_i - r - a^-) & \text{if } \pi_i^* = 0, \\ 0 & \text{if } \pi_i^* < 0. \end{cases}$$

Thus π^* is a solution to the optimisation problem.

Next, we show that the equations for a^+ and a^- always have a unique solution. From quadratic optimisation theory it is known that the optimisation problem always has a unique solution. This solution has to satisfy the Kuhn-Tucker equations. Thus it suffices to show that any solution to the Kuhn-Tucker equation can be written in the form of π^*, which we will do next.

For i with $\pi_i > 0$, (D.1) implies $\eta^2(1-\rho)\pi_i + \eta^2\rho 1_d^\top \pi - (\mu_i - r) + v/2 = 0$. Hence,

$$\pi_i = \frac{1}{\eta^2(1-\rho)} \left(\mu_i - r - a^+ \right),$$

where $a^+ = v/2 + \eta^2\rho 1_d^\top \pi$. Similarly, for i with $\pi_i < 0$,

$$\pi_i = \frac{1}{\eta^2(1-\rho)} \left(\mu_i - r - a^- \right),$$

where $a^- = -v/2 + \eta^2\rho 1_d^\top \pi$. Thus $a^+ + a^- = 2\eta^2\rho 1_d^\top \pi$. For i with $\pi_i = 0$,

$$2\eta^2\rho 1^\top \pi - 2(\mu_i - r) + v = u_{1i} \geq 0,$$
$$2\eta^2\rho 1^\top \pi - 2(\mu_i - r) - v = -u_{2i} \leq 0.$$

Hence, $a^+ \geq \mu_i - r$ and $a^- \leq \mu_i - r$. Thus if $\|\pi\|_1 = c$, then $\pi = \pi^*$. $\qquad \square$

E Some Results from Extreme Value Theory

Let $X \sim N(\mu, \sigma^2)$. We will say that $|X|$ has a *folded normal distribution* and denote its distribution by $FN(\mu, \sigma^2)$.

Theorem E.1. *The folded normal distribution is in the domain of attraction of the Gumbel distribution, i.e. let $X_1, \ldots, X_d \sim FN(\mu, \sigma^2)$ be independent, then there exist norming constants $c_d > 0, b_d \in \mathbb{R}$ such that*

$$\frac{\max(X_1, \ldots, X_d) - b_d}{c_d} \xrightarrow{\mathcal{L}} H \quad (d \to \infty),$$

where H denotes the Gumbel distribution. The norming constants c_d, b_d are (implicitly) given by

$$1 - \Phi\left(\frac{b_d - \mu}{\sigma}\right) + 1 - \Phi\left(\frac{b_d + \mu}{\sigma}\right) = \frac{1}{d}, \quad c_d = a(b_d),$$

(E.1)

$$a(x) = \frac{\Phi\left(\frac{x-\mu}{\sigma}\right)(x-\mu) + \sigma\phi\left(\frac{x-\mu}{\sigma}\right) + \Phi\left(\frac{x+\mu}{\sigma}\right)(x+\mu) + \sigma\phi\left(\frac{x+\mu}{\sigma}\right) - 2x}{2 - \Phi(\frac{x-\mu}{\sigma}) - \Phi(\frac{x+\mu}{\sigma})},$$

where Φ (resp. ϕ) is the cdf (pdf) of a standard normal distribution.

The following two lemmas are needed for the proof of Theorem E.1.

Lemma E.2. *The cumulative distribution function of $X \sim FN(\mu, \sigma^2)$ is $F(x) = -1 + \Phi(\frac{x-\mu}{\sigma}) + \Phi(\frac{x+\mu}{\sigma})$ for $x \geq 0$.*

Proof. $F(x) = \mathbb{P}(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi\left(\frac{-x-\mu}{\sigma}\right).$ $\qquad\qquad\qquad\square$

Lemma E.3. *Let F be the cdf of $X \sim FN(\mu, \sigma^2)$ and let $\bar{F}(x) = 1 - F(x)$. Then*

$$\lim_{x \to \infty} \frac{\bar{F}(x)F''(x)}{(F'(x))^2} = -1,$$

where $'$ denotes differentiation.

Proof. First, we consider the case $\sigma = 1$. By l'Hôpital's rule,

$$\lim_{x \to \infty} \frac{\bar{F}(x)}{\frac{1}{x-\mu}\phi(x-\mu) + \frac{1}{x+\mu}\phi(x+\mu)} =$$

$$= \lim_{x \to \infty} \frac{-\phi(x-\mu) - \phi(x+\mu)}{\frac{-1}{(x-\mu)^2}\phi(x-\mu) - \phi(x-\mu) + \frac{-1}{(x+\mu)^2}\phi(x+\mu) - \phi(x+\mu)}$$

$$= \lim_{x \to \infty} \frac{-1}{-1 + \frac{\frac{-1}{(x-\mu)^2}\phi(x-\mu) + \frac{-1}{(x+\mu)^2}\phi(x+\mu)}{\phi(x-\mu) + \phi(x+\mu)}} = 1.$$

The last equality holds because

$$0 \leftarrow \frac{-1}{(x-\mu)^2} \leq \frac{\frac{-1}{(x-\mu)^2}\phi(x-\mu) + \frac{-1}{(x+\mu)^2}\phi(x+\mu)}{\phi(x-\mu) + \phi(x+\mu)} \leq \frac{-1}{(x+\mu)^2} \to 0.$$

Hence,

$$\lim_{x \to \infty} \frac{\bar{F}(x)F''(x)}{(F'(x))^2} = \lim_{x \to \infty} \frac{\left[\frac{\phi(x-\mu)}{x-\mu} + \frac{\phi(x+\mu)}{x+\mu}\right]\left[-(x-\mu)\phi(x-\mu) - (x+\mu)\phi(x+\mu)\right]}{\left[\phi(x-\mu) + \phi(x+\mu)\right]^2}.$$

Expanding the negative of the numerator gives

$$\phi(x-\mu)^2 + \phi(x+\mu)^2 + \phi(x-\mu)\phi(x+\mu)\left[\frac{x+\mu}{x-\mu} + \frac{x-\mu}{x+\mu}\right].$$

Thus

$$\lim_{x \to \infty} \frac{\bar{F}(x)F''(x)}{(F'(x))^2} = -1 - \lim_{x \to \infty} \frac{\phi(x-\mu)\phi(x+\mu)\left[\frac{x+\mu}{x-\mu} + \frac{x-\mu}{x+\mu} - 2\right]}{\left[\phi(x-\mu) + \phi(x+\mu)\right]^2}.$$

The limit on the right hand side is 0 because the term in the bracket in the numerator simplifies as follows

$$\frac{x^2 + 2\mu x + \mu^2 + x^2 - 2\mu x + \mu^2}{x^2 - \mu^2} - 2 = \frac{4\mu^2}{x^2 - \mu^2} \to 0 \quad (x \to \infty),$$

and because

$$0 \leq \frac{\phi(x-\mu)\phi(x+\mu)}{[\phi(x-\mu) + \phi(x+\mu)]^2} = \frac{\phi(x-\mu)\phi(x+\mu)}{\phi(x-\mu)^2 + \phi(x+\mu)^2 + 2\phi(x-\mu)\phi(x+\mu)} \leq \frac{1}{2}.$$

In the general case $\sigma \neq 1$, we have $F(x) = G(x/\sigma)$ where G is the cdf of $|X|$ where $X \sim N(\mu/\sigma, 1)$. Hence,

$$\lim_{x \to \infty} \frac{\bar{F}(x)F''(x)}{(F'(x))^2} = \lim_{x/\sigma \to \infty} \frac{\bar{G}(x/\sigma)G''(x/\sigma)/\sigma^2}{(G'(x/\sigma))^2/\sigma^2} = -1.$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Theorem E.1. Lemma E.3 together with (Embrechts et al., 1997, Example 3.3.23, Proposition 3.3.25) imply that F is in the domain of attraction of the Gumbel distribution. Furthermore, (Embrechts et al., 1997, Theorem 3.3.26) shows that the norming constants can be defined by

$$b_d = F^{-1}(1 - \frac{1}{d}), \quad c_d = a(b_d), \quad a(x) = \int_x^\infty \frac{\overline{F}(t)}{\overline{F}(x)} dt.$$

Evaluating the integral gives (E.1). □

Proof of Theorem 5.4. Corollary 5.3 shows that the optimal strategy with plug-in estimators invests in $k < d$ stocks if and only if $\frac{1}{\eta^2} \sum_{i=1}^{k-1}(Z_{i,d} - Z_{k,d}) < c \leq \frac{1}{\eta^2} \sum_{i=1}^{k}(Z_{i,d} - Z_{k+1,d})$, where $Z_{1,d} \geq Z_{2,d} \geq \ldots \geq Z_{d,d}$ are the upper order statistics of $Z_i = |\hat{\mu}_i - r|$, $i = 1, \ldots, d$. The random variables $Z_1, \ldots, Z_d$ are i.i.d. folded normally distributed.

Theorem E.1 shows that the folded normal distribution is in the domain of attraction of the Gumbel distribution and therefore (Embrechts et al., 1997, Corollary 4.2.11, p. 202) implies that for $k \geq 1$,

$$\frac{\sum_{i=1}^{k} Z_{i,d} - k Z_{k+1,d}}{c_d} \xrightarrow{\mathcal{L}} \sum_{i=1}^{k} E_i \quad (d \to \infty),$$

where $E_1, \ldots, E_k$ are i.i.d. exponentially distributed (with parameter 1) random variables and $c_d > 0$ are norming constants specified in Theorem E.1. Noting that $\mathbb{P}(\sum_{i=1}^{k} E_i \leq c) = \mathbb{P}(K + 1 = k)$ finishes the proof. □

Proof of Theorem 5.5. For even k, using notation from Theorem 5.2,

$$\begin{aligned} \mathbb{P}(\#\{i : \pi_i^* \neq 0\} > k) &\leq \mathbb{P}(\#\{i : \pi_i^* > 0\} > k/2 \text{ or } \#\{i : \pi_i^* < 0\} > k/2) \\ &\leq \mathbb{P}(\#\{i : \pi_i^* > 0\} > k/2) + \mathbb{P}(\#\{i : \pi_i^* < 0\} > k/2) \\ &\leq \mathbb{P}(\hat{\mu}_{k/2+1,d} - r - a^+ > 0) + \mathbb{P}(\hat{\mu}_{d-k/2,d} - r - a^- < 0), \end{aligned}$$

where $\hat{\mu}_{1,d} \geq \cdots \geq \hat{\mu}_{d,d}$. On the event $\{\hat{\mu}_{k/2+1,d} - r - a^+ > 0\}$, we have

$$\begin{aligned} c &\geq \sum_{i=1}^{d} |\pi_i^*| \geq \sum_{i=1}^{d} \max(\pi_i^*, 0) \geq \frac{1}{\eta^2(1-\rho)} \sum_{i=1}^{k/2+1} (\hat{\mu}_{i,d} - r - a^+) \\ &\geq \frac{1}{\eta^2(1-\rho)} \sum_{i=1}^{k/2+1} ((\hat{\mu}_{i,d} - r) - (\hat{\mu}_{k/2+1,d} - r)) = \frac{1}{\eta^2(1-\rho)} \sum_{i=1}^{k/2+1} (\hat{\mu}_{i,d} - \hat{\mu}_{k/2+1,d}), \end{aligned}$$

and, similarly, on the event $\{\hat{\mu}_{d-k/2,d} - r - a^- < 0\}$, we have

$$c \geq \sum_{i=1}^{d} |\pi_i^*| \geq \frac{1}{\eta^2(1-\rho)} \sum_{d-k/2}^{d} (a^- - (\hat{\mu}_{i,d} - r)) \geq \frac{1}{\eta^2(1-\rho)} \sum_{d-k/2}^{d} (\hat{\mu}_{d-k/2,d} - \hat{\mu}_{i,d}).$$

Thus,

$$\begin{aligned} \mathbb{P}(\#\{i : \pi_i^* \neq 0\} > k) \leq &\mathbb{P}\left(\sum_{i=1}^{k/2+1} [\hat{\mu}_{i,d} - \hat{\mu}_{k/2+1,d}] \leq c\eta^2(1-\rho) \right) \\ &+ \mathbb{P}\left(\sum_{i=d-k/2}^{d} [\hat{\mu}_{d-k/2,d} - \hat{\mu}_{i,d}] \leq c\eta^2(1-\rho) \right). \end{aligned}$$

Since $\mathbb{C}ov(\hat{\mu}) = \frac{1}{t_{est}}\eta^2((1-\rho)I + \rho 11^\top)$, there are independent and normally distributed $\mu_A, \mu_1^0, \ldots, \mu_d^0$ with $\mathbb{V}ar(\mu_i^0) = \eta^2(1-\rho)/t_{est}$ such that $\hat{\mu}_i = \mu_i^0 + \mu_A$. As a consequence we can replace $\hat{\mu}_{i,d} - \hat{\mu}_{k/2+1,d}$ by $\mu_{i,d}^0 - \mu_{k/2+1,d}^0$ in the previous expression.

The normal distribution is in the domain of attraction of a Gumbel distribution. Therefore, by (Embrechts et al., 1997, Example 3.3.29, Corollary 4.2.11),

$$\frac{\sqrt{t_{est}}}{\eta\sqrt{1-\rho}}(2\log(d))^{1/2}\left(\sum_{i=1}^{k/2+1} \mu_{i,d}^0 - (k/2+1)\mu_{k/2+1,d}^0\right) \overset{\mathcal{L}}{\to} \sum_{i=1}^{k/2+1} E_i \quad (d \to \infty),$$

where $\mu_{1,d}^0 \geq \cdots \geq \mu_{d,d}^0$ and $E_1, \ldots, E_{k/2+1}$ are i.i.d. standard exponential. A similar argument can be made for the lower tail.

With the chosen threshold $c = \alpha(2\log(d))^{-1/2}$ we get

$$\lim_{d\to\infty}\mathbb{P}(\#\{i : \pi_i^* \neq 0\} > k) \leq 2\mathbb{P}\left(\sum_{i=1}^{k/2+1} E_i \leq \lambda\right) = 2\mathbb{P}(K \geq k/2 + 1),$$

where $\lambda = \alpha\eta\sqrt{1-\rho}\sqrt{t_{est}}$ and K is a Poisson distribution with parameter λ. $\square$

F   Proofs for the EWE-Strategy

Proof of Theorem 6.2. Plugging π^{EWE} into (2.1), the last term in the expectation is

$$(\pi^{\text{EWE}})^\top \Sigma \pi^{\text{EWE}} \leq \eta^2\alpha^2\left[(1-\rho)\beta d + \rho\beta^2 d^2\right] = \eta^2\alpha^2\beta d\left[1 + \rho(\beta d - 1)\right],$$

and thus,

$$V(\pi^{\text{EWE}}|\mu) \geq \log(X_0) + rT + T\mathbb{E}[(\pi^{\text{EWE}})^\top(\mu - r1)] - \frac{T}{2}\eta^2\frac{\alpha_*^2}{\beta d}\left[1 + \rho(\beta d - 1)\right].$$

Next, we prove the second part of the statement. For all $i = 1, \ldots, d$,

$$\mathbb{E}[\pi_i^{\text{EWE}}(\mu_i - r)] = \alpha\mathbb{E}[\mathbb{P}(\pi_i^{EWE} > 0|\mu_A, \mu_i)(\mu_i - r) - \mathbb{P}(\pi_i^{EWE} < 0|\mu_A, \mu_i)(\mu_i - r)]$$
$$\overset{(*)}{\to} \alpha\mathbb{E}[\mathbb{P}(\hat{a}_1 > c|\mu_1, \mu_A)(\mu_i - r)] - \mathbb{P}(\hat{a}_1 < -c|\mu_1, \mu_A)(\mu_i - r)].$$

To see $(*)$:

$$\mathbb{P}(\pi_i^{EWE} > 0|\mu_A, \mu_i) = \mathbb{P}\left(\frac{1}{d}\sum_{j=1}^{d}\mathbb{I}\left(|\hat{a}_j| < |\hat{a}_i|\right) \geq 1 - \beta, \hat{a}_i > 0\Big|\mu_A, \mu_i\right).$$

As $d \to \infty$, for $i \neq 1$,

$$\frac{1}{d}\sum_{j=1}^{d}\mathbb{I}\left(|\hat{a}_j| < |\hat{a}_i|\right) \to \mathbb{P}\left(|\hat{a}_1| < |\hat{a}_i||\mu_A, \mu_1, \ldots, \mu_d\right).$$

Thus by the bounded convergence theorem, $(*)$ holds.

Letting $d \to \infty$ in the final term of (6.2) completes the proof. $\square$

References

Bai, Z., Liu, H., and Wong, W. (2009). Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory. *Math. Finance*, 19(4):639–667.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., and Loris, I. (2009). Sparse and stable Markowitz portfolios. *P. Natl. Acad. Sci.*, 106(30):12267–12272.

Cvitanic, J. and Karatzas, I. (1992). Convex duality in constrained portfolio optimization. *Ann. Appl. Probab.*, 2(4):767–818.

Davis, M. H. A. and Norman, A. R. (1990). Portfolio selection with transaction costs. *Mathematics of Operations Research*, 15(4):676–713.

DeMiguel, V., Garlappi, L., and Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Rev. Financ. Stud.*, 22(5):1915–1953.

Detemple, J. (1986). Asset pricing in a production economy with incomplete information. *Journal of Finance*, 383–391.

Detemple, J. (1991). Further results on asset pricing with incomplete information. *Journal of Economic dynamics and Control*, 15(3):425–453.

Detemple, J., Garcia, R., and Rindisbacher, M. (2003). A Monte Carlo method for optimal portfolios. *Journal of Finance*, 58(1):401–446.

Dothan, M. and Feldman, D. (1986). Equilibrium interest rates and multiperiod bonds in a partially observable economy. *Journal of Finance*, 41(2):369–382.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.*, 32(2):407–451.

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, 70(350):311–319.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer-Verlag.

Fan, J., Zhang, J., and Yu, K. (2009). Asset allocation and risk assessment with gross exposure constraints on vast portfolios. Preprint, Princeton University.

Garlappi, L., Uppal, R., and Wang, T. (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Rev. Financ. Stud.*, 20(1):41.

Gennotte, G. (1986). Optimal portfolio choice under incomplete information. *Journal of Finance*, 41(3):733–746.

Golosnoy, V. and Okhrin, Y. (2009). Flexible shrinkage in portfolio selection. *J. Econ. Dyn. Control.*, 33(2):317–328.

Gruber, M. (1998). *Improving Efficiency by Shrinkage: The James-Stein and Ridge Regression Estimators*. New York: Marcel Dekker.

Houdré, C. (2002). Remarks on deviation inequalities for functions of infinitely divisible random vectors. *Ann. Probab.*, 30(3):1223–1237.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *J. Finance*, 58(4):1651–1684.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob., Vol. 1 (Univ. of Calif. Press)*, 361–379.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *J. Financ. Quant. Anal.*, 21(3):279–292.

Karatzas, I. and Zhao, X. (2001). Bayesian adaptive portfolio optimization. In *Handbook of mathematical finance*. Cambridge: Cambridge University Press.

Lawson, C. L. and Hanson, R. J. (1995). *Solving Least Squares Problems*. SIAM, 3rd edition.

Ledoit, O. and Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *J. Portfolio Manage.*, 30(4):110–119.

Markowitz, H. M. (1952). Portfolio selection. *J. Finance*, 7(1):77–91.

Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investment*. John Wiley & Sons.

Merton, R. (1971). Optimum consumption and portfolio rules in a continuous-time model. *J. Econ. Theory*, 3(4):373–413.

Merton, R. (1980). On Estimating the Expected Return on the Market: An Exploratory Investigation. *J. Financ. Econ.*, 8:323–361.

Ocone, D. and Karatzas, I. (1991). A generalized Clark representation formula, with application to optimal portfolios. *Stochastics: An International Journal of Probability and Stochastic Processes*, 34(3):187–220.

Pesaran, M. and Zaffaroni, P. (2008). Optimal asset allocation with factor models for large portfolios. IEPR working paper 08.07.

Rieder, U. and Bäuerle, N. (2005). Portfolio optimization with unobservable Markov-modulated drift process. *J. Appl. Probab.*, 42:362–378.

Rogers, L. C. G. (2001). The relaxed investor and parameter uncertainty. *Finance Stoch.*, 5(2):131–154.

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, 9(6):1135–1151.

Stein, J. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proc. Third Berkeley Symp. on Math. Statist. and Prob.*, 197–206.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. B*, 58(1):267–288.

Young, G. and Smith, R. (2005). *Essentials of Statistical Inference*. Cambridge: Cambridge University Press.