Mathieu S. Dubois and Luitgard A. M. Veraart

Optimal diversification in the presence of parameter uncertainty for a risk averse investor

Article (Published version) (Refereed)

Original citation:
Dubois, Mathieu S. and Veraart, Luitgard A. M. (2015) *Optimal diversification in the presence of parameter uncertainty for a risk averse investor.* SIAM Journal on Financial Mathematics, 6 (1). pp. 201-241. ISSN 1945-497X

DOI: 10.1137/130942826

SIAM J. FINANCIAL MATH.
Vol. 6, pp. 201–241

Optimal Diversification in the Presence of Parameter Uncertainty for a Risk Averse Investor*

Mathieu S. Dubois† and Luitgard A. M. Veraart†

Abstract. We consider an investor who faces parameter uncertainty in a continuous-time financial market. We model the investor's preference by a power utility function leading to constant relative risk aversion. We show that the loss in expected utility is large when using a simple plug-in strategy for unknown parameters. We also provide theoretical results that show the trade-off between holding a well-diversified portfolio and a portfolio that is robust against estimation errors. To reduce the effect of estimation, we constrain the weights of the risky assets with an L_1-norm leading to a sparse portfolio. We provide analytical results that show how the sparsity of the constrained portfolio depends on the coefficient of relative risk aversion. Based on a simulation study, we demonstrate the existence and the uniqueness of an optimal bound on the L_1-norm for each level of relative risk aversion.

Key words. optimal investment, power utility, parameter uncertainty, risk aversion, L_1-norm constraint, sparse portfolio

AMS subject classifications. 91G10, 91G70, 93E20

DOI. 10.1137/130942826

1. Introduction.

We consider a financial market consisting of one risk-free asset and a large number of risky assets following a multidimensional geometric Brownian motion. In this market we assume that there is an investor with a power utility function seeking to maximize the expected utility of her final wealth. If all parameters are known, [23] shows that the optimal fraction of wealth allocated in each risky asset is characterized by the drift and the volatility matrix of assets returns.[1]

These two quantities are not directly observable, and they are typically replaced by estimates computed from historical data.[2] For instance, we can plug simple estimators, such as the sample or the maximum likelihood estimators, into the analytical expression of the optimal portfolio weights. However, the resulting plug-in strategy is likely to differ considerably from the true optimal strategy. The main source of error comes from the estimation of the drift. Indeed, the accuracy of the estimator of the drift does not depend on the frequency of observations but on the length of the observation interval. To obtain a reasonable precision

*Received by the editors October 24, 2013; accepted for publication (in revised form) December 15, 2014; published electronically March 26, 2015.
http://www.siam.org/journals/sifin/6/94282.html

†London School of Economics and Political Science, Department of Mathematics, Houghton Street, London WC2A 2AE, UK (m.dubois@lse.ac.uk, l.veraart@lse.ac.uk). The first author gratefully acknowledges financial support by the Jones/Markham scholarship.

[1]When all parameters are known, [23] shows that the utility maximization problem is reduced to a mean-variance problem.

[2]It is also possible to use estimates of the covariance matrix or, equivalently, the volatility matrix, based on forward-looking information; see the recent work of [19] for a successful application in portfolio selection.

for estimators of the drift, one needs to use a very long period of observation; see [24].

Moreover, when the number of risky assets is large, the problem becomes even more prominent as estimation errors accumulate across the positions in the risky assets. Based on [23] and restricted to a logarithmic utility function, [12] shows that the expected utility associated with the plug-in strategy can degenerate to $-\infty$ as the number of assets increases.[3]

In this paper we extend the approach of [12] to general power utility functions. By taking into account the coefficient of relative risk aversion (RRA) explicitly, we are able to pin down the influence of both risk aversion and estimation risk on the expected utility.

Next, we impose an L_1-constraint on the portfolio weights. This constraint induces sparsity in the portfolio, i.e., most of the weights are set to zero, and it naturally reduces the accumulation of estimation error. Our objective is to select the bound of the L_1-constraint which gives the optimal degree of sparsity in the portfolio.

Holding a sparse portfolio is known to be an efficient way to reduce exposure to estimation risk.[4] In a minimum-variance framework with a large number of risky assets (500 stocks), [14] shows that the no-short sale constraint is likely to improve the performance of the plug-in strategy based on the sample covariance matrix. In this case, the estimation error is large, and constraining the amount of short positions can help because the associated plug-in strategy is sparse. However, the authors of [14] also demonstrate that the portfolio has too many weights set to zero. As a result, the poor performance of only a few assets can dramatically influence the performance of the portfolio.

To overcome this problem [6], [3], and [11] generalize the no-short sale constraint to the less restrictive L_1-constraint. The L_1-constraint is more flexible because it imposes an upper bound on the proportion of short positions. Thus, the set of admissible portfolios is augmented by relaxing the constraint while keeping a reasonable sparsity. With a suitable bound the constrained plug-in strategy has a smaller out-of-sample variance than benchmark portfolios such as the no-short sale minimum-variance portfolio and the equally weighted portfolio; see [6], [3], and [11]. Moreover, it also outperforms strategies based on James–Stein estimators in the static framework of [6] and in the dynamic framework of [12].

The identification of a good bound for the L_1-constraint is decisive for the performance of the constrained plug-in strategy. The empirical results of [11] demonstrate that the out-of-sample variance of the constrained minimum-variance plug-in strategy is convex in the bound of the L_1-constraint. In particular, the variance can be reduced by half by moving from the no-short sale constraint to the optimal L_1-constraint. Relaxing the constraint further increases the variance up to 25 percent. Therefore, a relatively small interval has to be identified for the bound. Alternatively, [6] suggests selecting the bound, which minimizes the variance, using the cross-validation method. None of these papers characterize the structure of the constrained strategy or justify explicitly the existence of an optimal bound. We address these problems as outlined below.

[3]In the static framework it is well known that plug-in mean-variance efficient portfolios perform poorly out-of-sample. In particular, no standard plug-in strategy outperforms consistently the equally weighted portfolio benchmark in terms of Sharpe ratio, certainty equivalence, and turnover; see [7] and references therein.

[4]Sparsity could also be induced by considering an optimal subset of available assets, i.e., by solving an L_0-norm problem. Even with a quadratic objective, this is an NP-hard problem; see [26] and [1]. In the portfolio selection literature, simple convex constraints such as L_1-constraints inducing sparsity are favored because of their computational tractability.

In section 2, we introduce the general setup and recall the optimal investment strategies when parameters are assumed to be observable. We then drop the observability assumption of the drift in section 3. We estimate the drift vector with the maximum likelihood estimator (MLE) and assume that the volatility matrix is known. This assumption is justified by assuming that prices are observed continuously, and hence the volatility is directly observable from the quadratic variation of the logarithmic asset price, but the drift is not.

In section 3, Theorem 3.2, we show that the expected utility of the plug-in strategy is equal to the theoretical expected utility of the optimal strategy with known parameters times a loss factor. For a fixed investment horizon, this loss factor is increasing in the number of risky assets. In particular, the expected utility can degenerate as the number of risky assets increases. When the drift is estimated, the diversification of the plug-in strategy clearly hurts.

In section 4, we introduce the L_1-norm as a way to constrain investment weights. We demonstrate that the sparsity of the optimal L_1-constrained strategy depends, to a large extent, on the coefficient of RRA. To understand the relation between the structure of the constrained strategy and the coefficient of RRA, we provide the analytical solution of the optimal L_1-constrained portfolio for independent risky assets in Theorem 4.4. In this case only the assets with the largest absolute excess returns are selected. The L_1-constrained portfolio rule consists of shrinking the excess returns towards zero by an intensity which is the same for all assets. If the absolute excess return of an asset is smaller than this constant, we do not invest in it. The number of assets invested in and the shrinkage intensity depend on both the bound of the L_1-constraint and the level of risk aversion. The L_1-constrained strategy becomes less sparse, as the coefficient of RRA increases, for both the true and the estimated drift. In terms of diversification, increasing the coefficient of RRA is similar to relaxing the constraint.

When facing parameter uncertainty, we show, in Proposition 4.7, that imposing an L_1-constraint rules out degeneracy of the expected utility of the plug-in strategy. Indeed, even if the number of assets goes to infinity, the L_1-constrained portfolio remains sparse, which prevents accumulation of estimation error.

With a fixed number of assets, the key point is to analyze the trade-off between the loss due to the lack of diversification, introduced by the L_1-constraint, and the loss due to estimation error. As we relax the constraint, the loss due to under-diversification decreases, while the loss due to estimation increases. These two losses move in opposite directions. Depending on the structure of the drift, the volatility matrix, and the method of estimation, there can be an L_1-bound which minimizes the total loss of the constrained plug-in strategy for each level of risk aversion.

For a general volatility matrix, we do not have closed form results for optimal L_1-constrained strategies. Therefore we use in section 5 a simulation study to investigate the properties and the performance of the L_1-constrained portfolio, when assets are correlated. Similarly to the independent case, the L_1-constrained strategy becomes less sparse as the coefficient of RRA increases. We present numerical examples which show that the L_1-bound can be chosen in an optimal way to minimize the loss due to estimation. This optimal choice depends crucially on the level of risk aversion.

Finally, in section 6 we consider a universe of stocks based on the S&P 500 from 2006 to 2011, and we investigate the out-of-sample performance of the unconstrained and the constrained plug-in strategies. We trade daily over one-month intervals to test the strategies. We calibrate the optimal bound of the constrained strategy based on two different numerical methods. Our results confirm that the unconstrained strategy has very unstable returns. We also demonstrate that imposing the appropriate L_1-constraint improves greatly the performance of the plug-in strategy. While, on average, the constrained strategy has a higher variance than the equally weighted portfolio, it delivers a utility of terminal wealth which is in the same range. Hence, the L_1-constrained plug-in strategy has a comparable performance to the equally weighted portfolio, even when the drift is estimated.

2. Model setup.

We consider a financial market where trading takes place continuously over a finite time interval $[0,T]$ for $0 < T < \infty$. The market consists of one risk-free asset with time-t price $S_0(t)$ and d risky assets with time-t price $S_i(t)$ for $i = 1, \dots, d$. Their dynamics are given by

$$
\begin{aligned}
dS_0(t) &= S_0(t)\, r dt, & S_0(0) &= 1,\\
dS_i(t) &= S_i(t)\left(\mu_i dt + \sum_{j=1}^{d} \sigma_{ij} dW_j(t)\right), & S_i(0) &> 0 \quad \text{for } i = 1, \dots, d,
\end{aligned}
$$

where $r \geq 0$ is the constant interest rate, $\mu = (\mu_1, \dots, \mu_d)^\top$ 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. Furthermore, $W = (W_1, \dots, W_d)^\top$ is a standard d-dimensional Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

We denote by $X^\pi(t)$ the investor's wealth at time t when using strategy π, which is given by

$$
dX^\pi(t) = \sum_{i=0}^{d} \pi_i(t)\, X^\pi(t)\, \frac{dS_i(t)}{S_i(t)},
$$

for a constant initial wealth $X^\pi(0) = X(0) > 0$. Here, $\pi_i(t)$ denotes the fraction of wealth invested in the ith asset at time t. Hence, $\sum_{i=0}^{d} \pi_i(t) = 1$ for all t. Using $\pi_0(t) = 1 - \sum_{i=1}^{d} \pi_i(t)$ and setting $\pi(t) = (\pi_1(t), \dots, \pi_d(t))^\top$, we obtain

$$
\frac{dX^\pi(t)}{X^\pi(t)} = \left(r + \pi(t)^\top (\mu - r1)\right) dt + \pi^\top(t)\, \sigma dW(t)\,, \tag{2.1}
$$

where $1 = (1, \dots, 1)^\top \in \mathbb{R}^d$. For strategies $(\pi(t))_{t\geq 0}$ that are adapted to the filtration $(\mathcal{F}(t))_{t\geq 0}$ with $\mathcal{F}(t) = \sigma(W(s), s \leq t)$ and that are sufficiently integrable, the solution of (2.1) is

$$
X^\pi(T) = X(0) \exp\left(\int_0^T \left(r + \pi(t)^\top (\mu - r1) - \frac{1}{2}\pi(t)^\top \Sigma \pi(t)\right) dt + \int_0^T \pi(t)^\top \sigma dW(t)\right),
$$

where $\Sigma = \sigma\sigma^\top$. Note that Σ is positive definite.

2.1. The investor's objective and the classical solution. We consider an investor with a constant relative risk aversion (CRRA) utility function of the type

$$U_\gamma(x) = \begin{cases} \frac{x^{1-\gamma}}{1-\gamma} & \text{for } \gamma > 1, \\ \log(x) & \text{for } \gamma = 1 \end{cases}$$

for $x > 0$, where γ is the coefficient of RRA.

The investor's goal is to maximize

$$V_\gamma(\pi|\mu,\sigma) := \mathbb{E}[U_\gamma(X^\pi(T))] \tag{2.2}$$

over strategies π which are sufficiently integrable so that $V_\gamma(\pi|\mu,\sigma)$ is well defined. We call such strategies admissible. We use the notation $V_\gamma(\pi|\mu,\sigma)$ to emphasize that the objective function depends on γ, μ, and σ.

The paper [23] shows that the optimal strategy, denoted by π^* and called the Merton ratio, consists of holding a constant proportion in each asset:

$$\pi^*(t) = \frac{1}{\gamma}\Sigma^{-1}(\mu - r1) \quad \forall t \in [0,T].$$

Therefore, the Merton ratio also maximizes the mean-variance term

$$M_\gamma(\pi) = \pi^\top(\mu - r1) - \frac{\gamma}{2}\pi^\top\Sigma\pi.$$

Finally, the corresponding expected utility is given by

$$V_\gamma(\pi^*|\mu,\sigma) = \begin{cases} K_\gamma \exp\left(\frac{(1-\gamma)T}{2\gamma}(\mu - r1)^\top \Sigma^{-1}(\mu - r1)\right), & \gamma > 1, \\ K_\gamma + \frac{T}{2}(\mu - r1)^\top \Sigma^{-1}(\mu - r), & \gamma = 1, \end{cases} \tag{2.3}$$

with

$$K_\gamma = \begin{cases} \frac{X(0)^{1-\gamma}}{1-\gamma}\exp((1-\gamma)rT), & \gamma > 1, \\ \log(X(0)) + rT, & \gamma = 1. \end{cases} \tag{2.4}$$

Regardless of the magnitude of the initial wealth $X(0)$, K_γ is always strictly negative for $\gamma > 1$.

2.2. The effect of diversification for known parameters. When the true parameters are known, the investor optimally diversifies her portfolio by investing the corresponding Merton ratio in each stock. As more stocks become available, one simply maximizes over a larger set of strategies, and the expected utility increases at a rate which, by (2.3), depends on γ and the growth of the quadratic form $(\mu - r1)^\top \Sigma^{-1}(\mu - r1)$. Moreover, if the spectrum of the matrix Σ^{-1} is bounded from above and away from zero, analyzing the convergence of the Euclidean norm of excess returns $||\mu - r1||_2$ is sufficient to characterize the asymptotic behavior of the expected utility.

When studying the expected utility as a function of the number of risky assets d, we are in effect considering a sequence of markets. This sequence is built with a market containing the first $1,\ldots,d$ risky assets in the same order, and then a new risky asset is included and considered as the $(d+1)$st asset. In this setting, the drift, the volatility matrix, the covariance matrix, and the Brownian motion are denoted, for the market with d risky assets, by $\mu^{(d)}$, $\sigma^{(d)}$, $\Sigma^{(d)}$, and $W^{(d)}$, respectively. A portfolio strategy in the market with d risky assets is denoted by $\pi^{(d)}$.

Proposition 2.1. *Let $\left(\Sigma^{(d)}\right)_{d\geq 1} \subset \mathbb{R}^{d\times d}$ be such that its eigenvalues $\lambda_i^{(d)}$, $i=1,\ldots,d$, satisfy*

$$\underline{m}_\lambda = \lim_{d\to\infty} \min_{i=1,\ldots,d} 1/\lambda_i^{(d)}, \qquad \overline{m}_\lambda = \lim_{d\to\infty} \max_{i=1,\ldots,d} 1/\lambda_i^{(d)}. \tag{2.5}$$

Suppose that $\underline{m}_\lambda > 0$ and $\overline{m}_\lambda < \infty$. Then, for $\gamma > 1$ and for all d we have that

$$\begin{aligned} &K_\gamma \exp\left(\frac{(1-\gamma)\,T}{2\gamma}\underline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2\right) \\ &\leq V_\gamma\left(\left(\pi^*\right)^{(d)} |\mu^{(d)}, \sigma^{(d)}\right) \\ &\leq K_\gamma \exp\left(\frac{(1-\gamma)\,T}{2\gamma}\overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2\right), \end{aligned}$$

and for $\gamma = 1$,

$$K_1 + \frac{T}{2}\underline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2 \leq V_1\left(\left(\pi^*\right)^{(d)} |\mu^{(d)}, \sigma^{(d)}\right) \leq K_1 + \frac{T}{2}\overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2.$$

A proof is given in Appendix A. As $||\mu^{(d)} - r1^{(d)}||_2$ is an increasing positive sequence in d, it always admits a limit. If this limit is finite, the expected utility is bounded away from zero. If the limit is infinite, the expected utility reaches zero and the positive effect of diversification is fully exploited. The case $\gamma = 1$ is similar.

3. Performance of plug-in strategies.

When asset prices are continuously observed, we can obtain the true volatility matrix σ since the quadratic variation of the log-stock price is observable. However, this is not the case for the drift μ. Indeed, the accuracy of the estimation of the drift depends on the length of the estimation period and not on the frequency of observations. We use the MLE of the drift over the observation period $[-t_{obs}, 0]$ for a constant $t_{obs} > 0$. The MLE for μ_i, $i = 1,\ldots,d$, is given by

$$\hat{\mu}_i = \frac{\log\left(S_i\left(0\right)\right) - \log\left(S_i\left(-t_{obs}\right)\right)}{t_{obs}} + \frac{1}{2}\sum_{j=1}^{d} \sigma_{ij}^2. \tag{3.1}$$

Based on the estimator $\hat{\mu}$, one can implement the time-constant plug-in strategy

$$\hat{\pi} = \frac{1}{\gamma}\Sigma^{-1}\left(\hat{\mu} - r1\right). \tag{3.2}$$

$\hat{\pi}$ is an unbiased Gaussian estimator of π^*, in particular,

$$\hat{\pi} \sim \mathcal{N}\left(\pi^*, V_0^2\right) \qquad \text{with } V_0^2 = \frac{1}{\gamma^2 t_{obs}}\Sigma^{-1}. \tag{3.3}$$

Furthermore, the expected utility of the plug-in strategy is given by the moment generating function of the mean-variance term $M_\gamma(\hat{\pi})$, as the following lemma shows.

Lemma 3.1. *Let $\gamma > 1$; then it holds that*

$$V_\gamma(\hat{\pi}|\mu,\sigma) = K_\gamma \mathbb{E}\left[\exp\left((1-\gamma)TM_\gamma(\hat{\pi})\right)\right]. \tag{3.4}$$

The lemma is proved in Appendix B. We now characterize the loss in expected utility due to estimation when implementing the plug-in strategy $\hat{\pi}$, based on the MLE of the drift.

Theorem 3.2. *Let $\gamma > 1$ and $t_{obs} > T$. Then the expected utility of the plug-in strategy $\hat{\pi}$ is given by*

$$V_\gamma(\hat{\pi}|\mu,\sigma) = L_\gamma(\hat{\pi},\pi^*)\, V_\gamma(\pi^*|\mu,\sigma) \tag{3.5}$$

with

$$L_\gamma(\hat{\pi},\pi^*) = \left(1 + \frac{(1-\gamma)T}{\gamma t_{obs}}\right)^{-\frac{d}{2}}. \tag{3.6}$$

A proof is given in Appendix B. For the case $\gamma = 1$, [12] has shown that the loss is linear in d:

$$V_1(\hat{\pi}|\mu,\sigma) = V_1(\pi^*|\mu,\sigma) - L_1(\hat{\pi},\pi^*) \qquad \text{with } L_1(\hat{\pi},\pi^*) = \frac{T}{2t_{obs}}d > 0. \tag{3.7}$$

While the loss factor does not depend on the value of the true parameters μ and Σ, it is an increasing function of γ and d.

Since $\hat{\pi}$ is a consistent estimator of π^*, the expected utility of the plug-in strategy converges to the expected utility of the optimal strategy as the length of the observation period $t_{obs} \to \infty$.

By (3.3), the accuracy of the plug-in strategy depends on the length of the observation period, and the rate of convergence in (3.5) is very slow. For instance, with $T = 1$, $\gamma = 2$, and $d = 200$ risky assets, we need three centuries of observations, $t_{obs} = 300$, to get a loss factor close to one, $L_\gamma(\hat{\pi},\pi^*) \approx 1.18$. We will see from (3.9) that this corresponds to a 15.3% loss in certainty equivalent. Therefore, the loss can be reduced significantly only by taking a very long estimation period, and, for a feasible estimation period, using a plug-in strategy $\hat{\pi}$ results in a poor expected utility.

The following corollary gives a sufficient condition for the expected utility to degenerate.

Corollary 3.3. *Let $\gamma \geq 1$, $t_{obs} > T$ and suppose that the sequence of eigenvalues of $\Sigma^{(d)}$ verifies (2.5). If $||\mu^{(d)} - r1^{(d)}||_2^2$ is in $o(d)$, then*

$$V_\gamma\left(\hat{\pi}^{(d)}|\mu^{(d)},\sigma^{(d)}\right) \to -\infty \qquad \text{as } d \to \infty.$$

A proof is given in Appendix B. For instance, with $\mu_i^{(d)} - r = 1/i$, $i = 1,\dots,d$, the sequence $||\mu^{(d)} - r1^{(d)}||_2^2$ has a finite limit and the corollary applies.

It is already well known that strategies based on the MLE of the drift perform poorly. It is common to obtain extreme positions due to estimation error, and, for high-dimensional problems, the accumulation of error leads to a large loss in expected utility. What is new here is a full description of the loss due to estimation as a function of the coefficient of RRA and the number of risky assets. Additionally, we provide in Corollary 3.3 a sufficient condition for the degeneracy of the expected utility as $d \to \infty$.

3.1. Measures of economic loss. Theorem 3.2 establishes an analytic relationship between the expected utility obtained from using the optimal strategy with known drift and the expected utility from using a plug-in strategy. In general it is hard to interpret different levels of expected utility, as utility functions describe a preference ordering which is invariant to linear transformations. Therefore we provide some discussion on how one can measure economic loss that is due to using a plug-in strategy rather than an optimal strategy.

3.1.1. Mean-variance loss function. For known parameters, problem (2.2) is equivalent to maximizing the (instantaneous) mean-variance term. When deviating from the optimal strategy, a standard choice to measure economic loss is the mean-variance loss function:[5]

$$L_\gamma^M(\hat{\pi}, \pi^*) = M_\gamma(\pi^*) - \mathbb{E}\left[M_\gamma(\hat{\pi})\right].$$

The mean-variance loss is then given by

$$L_\gamma^M(\hat{\pi}, \pi^*) = \frac{d}{2\gamma t_{obs}}. \tag{3.8}$$

$L_\gamma^M(\hat{\pi}, \pi^*)$ captures the fact that a smaller fraction of wealth is invested in the risky assets as γ increases.[6]

However, $L_\gamma^M(\hat{\pi}, \pi^*)$ does not measure estimation risk consistently if one considers an investor with CRRA power utility function for $\gamma > 1$. Indeed, by Lemma 3.1, the expected utility of final wealth is proportional to the moment generating function of the mean-variance term. In general there is a nonmonotonic relation between the moment generating function and the expectation of the mean-variance term. Hence, the equivalence between our setting and the mean-variance approach does not hold when the implemented strategy is random.

3.1.2. Certainty equivalents and efficiency. To account for both sources of risk consistently, namely the risk due to the driving Brownian motions and the risk due to parameter uncertainty, the loss due to estimation has be to be quantified in terms of expected utility. A strategy $\hat{\pi}$ is suboptimal if it generates a loss in expected utility, $V_\gamma(\hat{\pi}|\mu, \sigma) \le V_\gamma(\pi^*|\mu, \sigma)$.

We now look at the loss in certainty equivalents and show its relation to the relative loss in expected utility.

Definition 3.4. *For the optimal strategy and the plug-in strategy, the* certainty equivalents *are the scalar quantities $CE_\gamma^{\hat{\pi}}$ and $CE_\gamma^{\pi^*}$, respectively, such that*

$$U_\gamma\left(CE_\gamma^{\hat{\pi}}\right) = V_\gamma(\hat{\pi}|\mu, \sigma) \quad \text{and} \quad U_\gamma\left(CE_\gamma^{\pi^*}\right) = V_\gamma(\pi^*|\mu, \sigma).$$

The certainty equivalents are the cash amounts delivering the same utility as the corresponding strategies. From the definition of CRRA utility functions one obtains immediately that, for the certainty equivalents $CE_\gamma^{\hat{\pi}}$ and $CE_\gamma^{\pi^*}$, the following relationship holds true:

$$\frac{CE_\gamma^{\hat{\pi}}}{CE_\gamma^{\pi^*}} = \begin{cases} L_\gamma(\hat{\pi}, \pi^*)^{\frac{1}{1-\gamma}} & \text{for } \gamma > 1, \\ \exp\left(-L_\gamma(\hat{\pi}, \pi^*)\right) & \text{for } \gamma = 1. \end{cases} \tag{3.9}$$

[5] See, e.g., [16] or [31] and the references therein.

[6] Paper [15] uses the relative mean-variance loss function $\frac{L_\gamma^M(\hat{\pi},\pi^*)}{|M_\gamma(\pi^*)|}$. In this case, the loss function does not depend on the coefficient of RRA.

The ratio of certainty equivalents can also be interpreted in terms of the efficiency measure that has been introduced in the literature to compare different expected utilities; see [29]. Along the lines of [29, Def. 1] we define the efficiency in our context as follows.

Definition 3.5. *The* efficiency $\Theta_\gamma(\pi)$ *of an investor with relative risk aversion* γ *using strategy* π *relative to the Merton investor (who uses the optimal strategy* π^**) is the amount of wealth at time* 0 *which the Merton investor would need to obtain the same maximized expected utility at time* T *as the investor with strategy* π *who started at time* 0 *with unit wealth.*

Using the results of Theorem 3.2 we obtain that, for CRRA utility functions, the ratio of the certainty equivalents (3.9) is exactly the efficiency.

Theorem 3.6. *The efficiency of the investor who uses the simple plug-in strategy* (3.2) *is given by*

$$\Theta_\gamma(\hat{\pi}) = \begin{cases} L_\gamma(\hat{\pi}, \pi^*)^{\frac{1}{1-\gamma}} & \text{for } \gamma > 1, \\ \exp\left(-L_1(\hat{\pi}, \pi^*)\right) & \text{for } \gamma = 1. \end{cases}$$

A proof can be found in Appendix B.

We see that for $\gamma > 1$ the relative loss factor $L_\gamma(\hat{\pi}, \pi^*)$ is the efficiency raised to power $1/(1-\gamma)$, and, for $\gamma = 1$, the absolute loss $L_1(\hat{\pi}, \pi^*)$ is minus the logarithm of the corresponding efficiency. Hence, there is a one-to-one monotonic relation between the relative loss in expected utility and efficiency. Furthermore, for $\gamma > 1$, the loss factor $L_\gamma(\hat{\pi}, \pi^*)$ is always greater than one and the efficiency is always smaller than one. When there is no estimation risk, both quantities are equal to one.

As the loss factor is increasing in the number of assets and the power $1/(1-\gamma)$ is negative, the efficiency is sharply decreasing with the number of assets. Namely, the more assets available, the lower the initial wealth of the Merton investor to obtain the same expected utility as the plug-in investor. This is illustrated in Figure 1.

While the loss factor $L_\gamma(\hat{\pi}, \pi^*)$ measures loss in expected utility consistently for a fixed level of risk aversion, its magnitude should not be compared across different levels of risk aversion. The expected utility of the plug-in investor is characterized as the product of the loss factor and the expected utility of the Merton investor, but both quantities depend on the investor's risk aversion γ. Although the loss factor itself is an increasing function in γ, this fact is not sufficient to draw conclusions on how expected utilities of plug-in investors with different risk-aversion parameters relate to each other.

We therefore look at the efficiency of the plug-in investor as a function of γ. For a fixed number of risky assets, the efficiency is an increasing function of γ. Hence, the plug-in strategy becomes more efficient as γ increases. If we consider two plug-in investors with different parameters of relative risk aversion γ_1 and γ_2, with $\gamma_1 > \gamma_2$, the more risk averse investor, i.e., the one with risk aversion γ_1, will be more efficient relative to the Merton investor than the plug-in investor, who is less risk averse with risk aversion γ_2. The reason for this behavior is that the more risk averse investor invests a smaller fraction of his wealth in the risky assets. This is in line with the behavior of the mean-variance loss function in (3.8), in which the effect of estimation is also reduced as the coefficient of RRA increases.

3.2. Drift versus covariance estimation. So far we have only considered the estimation problem of the drift and assumed that the matrix Σ is observable. We have justified at the

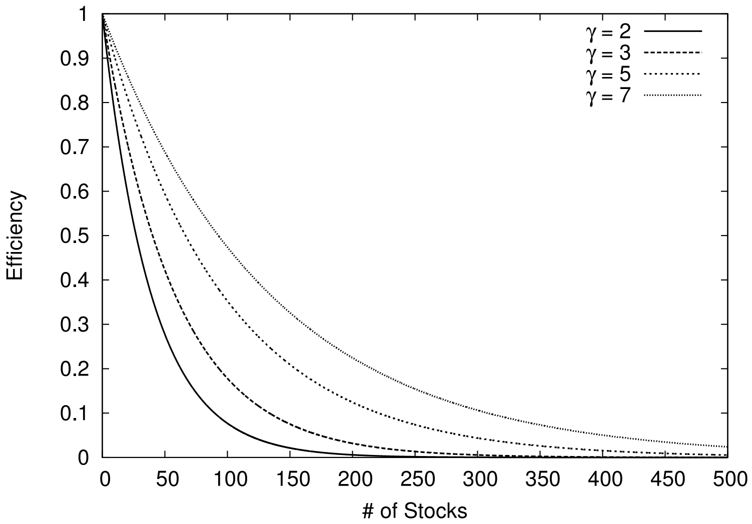

Figure 1. *Plot of the efficiency of the plug-in investor relative to the Merton investor as a function of the number of risky assets d for different levels of risk aversion γ. For a given γ, the efficiency depends only on the number of risky assets, the investment horizon $T = 1$, and the observation period $t_{obs} = 10$.*

beginning of section 3 that as long as we are in continuous time, the quadratic variation of the stock price is observable, and hence Σ is known.

As soon as we move to a discrete-time setting the situation changes. If we assume that observing the asset prices continuously is no longer possible, the covariance matrix Σ also needs to be estimated. Hence any discussion on estimating the covariance matrix is linked to the discussion on discrete- versus continuous-time settings.

The effects of discrete trading have already been studied by [29], and a detailed analysis on discrete trading and observations in the context of parameter uncertainty is available in [2]. Note that one cannot just suitably discretize a strategy that is optimal in continuous time to obtain a strategy that is optimal in discrete time. A strategy that is optimal in discrete time has different characteristics, e.g., short-selling is forbidden. Furthermore, [2] shows that, with parameter uncertainty on the drift and the covariance, the expected utility of the "discrete trader" does not converge to the expected utility of the "continuous trader," as the time step goes to zero.

Note that this is in contrast to the static Markowitz mean-variance approach, where there is no rebalancing. In a static mean-variance context, the structure and the performance of plug-in strategies using estimators for both the mean and the covariance matrix have been studied in depth; see, e.g., [10]. Since these results are already available and we are studying a continuous-time setting, we will not analyze the theoretical problem of the estimation of the covariance matrix any further.

4. L_1-restricted portfolio.

To avoid the degeneracy of the expected utility due to parameter uncertainty, we reduce the dimension of the portfolio by imposing an L_1-constraint on the investment strategies. For $c \geq 0$, the L_1-constrained problem is

$$\max_{\pi \in \mathcal{A}_c} V_\gamma\left(\pi|\mu,\sigma\right), \tag{4.1}$$

where $\mathcal{A}_c$ is the set of admissible constrained strategies π such that

$$||\pi\left(t,w\right)||_1 = \sum_{i=1}^{d} |\pi_i\left(t,w\right)| \leq c \qquad \text{for } m \otimes \mathbb{P}\text{-a.e. } (t,w)$$

and m is the Lebesgue measure on $[0,T]$.

Proposition 4.1. *Problem* (4.1) *reduces to the static problem*

$$\begin{cases} \max\limits_{\pi \in \mathbb{R}^d} V_\gamma\left(\pi|\mu,\sigma\right) \\ \text{subject to } ||\pi||_1 \leq c. \end{cases} \tag{4.2}$$

In particular, the optimal strategy π_c^* is deterministic and constant.

The proposition is proved in Appendix C.[7]

Remark 4.2. Note that the summation starts from $i=1$, i.e., we only restrict the portfolio weights in the risky assets. The bound c of the L_1-constraint controls the level of sparsity in the portfolio. Let $\tilde{\pi} = \left(\pi_0, \pi_1, \ldots, \pi_d\right)^\top$, with $\pi_0 = 1 - \sum_{i=1}^d \pi_i$. With $\pi_i^+ = \max\{\pi_i, 0\}$ and $\pi_i^- = -\min\{\pi_i, 0\}$ we denote by $\tilde{\pi}^{(l)}$ and $\tilde{\pi}^{(s)}$ the total percentages of long and short positions, respectively, and then

$$\tilde{\pi}^{(l)} = \sum_{i=0}^{d} \pi_i^+ = \sum_{i=0}^{d} \left(|\pi_i| + \pi_i\right)/2 = \frac{1}{2}\left(\|\tilde{\pi}\|_1 + 1\right),$$
$$\tilde{\pi}^{(s)} = \sum_{i=0}^{d} \pi_i^- = \sum_{i=0}^{d} \left(|\pi_i| - \pi_i\right)/2 = \frac{1}{2}\left(\|\tilde{\pi}\|_1 - 1\right).$$

Then $\tilde{\pi}^{(l)} - \tilde{\pi}^{(s)} = 1$ and $\tilde{\pi}^{(l)} + \tilde{\pi}^{(s)} = \|\tilde{\pi}\|_1 = \|\pi\|_1 + |\pi_0|$. Now with $\sum_{i=1}^d |\pi_i| \leq c$ we obtain that

$$\begin{aligned} \tilde{\pi}^{(s)} &= \sum_{i=0}^{d} \pi_i^- = \sum_{i=0}^{d} \left(|\pi_i| - \pi_i\right)/2 = \frac{1}{2}\left(\|\tilde{\pi}\|_1 - 1\right) = \frac{1}{2}\left(\|\pi\|_1 + \left|1 - \sum_{i=1}^{d} \pi_i\right| - 1\right) \\ &\leq \frac{1}{2}\left(c + \left|1 - \sum_{i=1}^{d} \pi_i\right| - 1\right) \leq \frac{1}{2}\left(c - 1 + 1 + c\right) = c. \end{aligned}$$

Hence, c is an upper bound on the total percentage of short positions held in the portfolio.

[7]See also the dual approach of [4] for power utility functions with $\gamma < 1$, and see [17] for $\gamma > 1$ and cone constraints.

4.1. Structure of the optimal strategy. We study the effect of the L_1-constraint on the optimal strategy for given parameters μ and σ. This allows us to understand how assets are selected and to characterize the sparsity of the strategy as a function of γ. As the optimal strategy of the initial constrained problem is constant and deterministic, the constrained optimization reduces to the mean-variance problem with the same L_1-constraint:

$$\begin{cases} \max\limits_{\pi\in\mathbb{R}^d} M_\gamma(\pi) \\ \text{subject to } ||\pi||_1 \le c. \end{cases} \tag{4.3}$$

The mean-variance term can be rewritten as

$$M_\gamma(\pi) = \pi^\top(\mu - r1) - \frac{\gamma}{2}\pi^\top\Sigma\pi = -\frac{1}{2}\left\|\sqrt{\gamma}\sigma^\top\pi - \frac{1}{\sqrt{\gamma}}\sigma^{-1}(\mu - r1)\right\|_2^2 + K, \tag{4.4}$$

where $||\cdot||_2$ is the Euclidean norm and K does not depend on π. The L_1-constraint holds only on the weights of the risky assets. Therefore, we have a standard L_1-constrained ordinary least-square (OLS) problem; see [30].

Lemma 4.3. *Problem (4.2) is equivalent to the constrained optimization of the quadratic form*

$$\begin{cases} \min\limits_{\pi\in\mathbb{R}^d} \left\|\sqrt{\gamma}\sigma^\top\pi - \frac{1}{\sqrt{\gamma}}\sigma^{-1}(\mu - r1)\right\|_2^2 \\ \textit{subject to } ||\pi||_1 \le c. \end{cases} \tag{4.5}$$

Note that (4.5) can be reduced to the case $\gamma = 1$ by considering the covariance matrix $\tilde{\Sigma} = \gamma\Sigma$. In that sense, we can interpret the optimization problem for general RRA parameter $\gamma \ge 1$ as an optimization problem in which a higher level of RRA is treated equivalently to larger entries in the covariance matrix for an investor with $\gamma = 1$.

To highlight the fundamental role of the coefficient of RRA γ on the sparsity of the constrained strategy, we provide an analytical solution of (4.5) for diagonal volatility matrices.

Theorem 4.4. *Suppose* $\Sigma = \operatorname{diag}\left(\sigma_1^2, \dots, \sigma_d^2\right)$ *and* $\mu \in \mathbb{R}^d$ *with*

$$|\mu_1 - r| > |\mu_2 - r| > \cdots > |\mu_d - r|.$$

Then $\pi_i^* = \frac{1}{\gamma\sigma_i^2}(\mu_i - r1)$, $i = 1, \dots, d$, *is a solution to (4.3) if* $||\pi^*||_1 \le c$. *Otherwise, the unique solution is*

$$\pi_c^* = \left(\frac{\operatorname{sgn}(\mu_1 - r)}{\gamma\sigma_1^2}(|\mu_1 - r| - a), \dots, \frac{\operatorname{sgn}(\mu_k - r)}{\gamma\sigma_k^2}(|\mu_k - r| - a), 0, \dots, 0\right)^\top, \tag{4.6}$$

where sgn *is the sign function and*

$$a = \frac{1}{\sum_{i=1}^k \frac{1}{\sigma_i^2}}\left(\sum_{i=1}^k \frac{|\mu_i - r|}{\sigma_i^2} - \gamma c\right), \tag{4.7}$$

$$k = \min\left\{j = 1, \dots, d : c \le \sum_{i=1}^j \frac{1}{\gamma\sigma_i^2}(|\mu_i - r| - |\mu_{j+1} - r|)\right\}, \tag{4.8}$$

and $\mu_{d+1} = r$.

The theorem is proved in Appendix C. The argument to find the structure of the constrained weights is as follows. As the optimal unconstrained strategy π^* does not satisfy the L_1-constraint in general, the weights have to be shrunk. For a diagonal volatility matrix, the constrained solution is of the form

$$(\pi_c^*)_i = \frac{\operatorname{sgn}(\mu_i - r)}{\gamma\sigma_i^2}\left(|\mu_i - r| - a\right)^+ \qquad \text{for each } i = 1, \dots, d, \tag{4.9}$$

where $a > 0$ and $(\cdot)^+$ denotes the positive part (see the proof of Theorem 4.4 in Appendix C).[8] Because of the structure of (4.9), we invest only in assets with the highest absolute excess returns, and we classify them by decreasing order of absolute excess return. The excess returns are adjusted by the shrinking constant a. The shrinkage intensity has to be decreasing in c to reflect the fact that the strategy is less constrained for a large c.

To satisfy the constraint, we have to deviate from π^*, but the shrinkage intensity should not change the sign of the position, long or short.[9] Therefore, the weight of an asset is set to zero if its absolute excess return is smaller than a. The index k is then defined as the last asset with an absolute excess return larger than a,

$$k = \min_{j=1,\dots,d}\left\{|\mu_{j+1} - r| \le a\right\}. \tag{4.10}$$

Since the L_1-constraint is binding, we obtain the expression (4.7) of the positive shrinkage intensity a as a function of k.

We can also write k in terms of c and γ. By (4.7) and (4.10), k is the smallest index such that the following inequality holds:

$$\sum_{i=1}^{k} \frac{|\mu_{k+1} - r|}{\gamma\sigma_i^2} \le \sum_{i=1}^{k} \frac{|\mu_i - r|}{\gamma\sigma_i^2} - c. \tag{4.11}$$

The left-hand side of inequality (4.11) corresponds to the L_1-norm of the weights of k risky independent assets with a common drift μ_{k+1} and volatility matrix σ. Therefore, the strategy includes assets with largest excess returns until the difference between the L_1-norm of the corresponding unconstrained weights and the bound c exceeds the L_1-norm of the weights of these k fictitious assets. Rewriting inequality (4.11), we get the full characterization of k in (4.8) and the constrained strategy in (4.6). The index k increases with γ, and the larger γ is, the less sparse the constrained portfolio.

Finally, we establish some regularity properties of the optimal constrained strategy and the corresponding expected utility as functions of the bound c.

Proposition 4.5. *Let π_c^* be the optimal solution of problem (4.2).*

1. *The solution map $c \mapsto \pi_c^*$ is continuous on $\mathbb{R}^+$.*
2. *The expected utility map $c \mapsto V_\gamma(\pi_c^* | \mu, \sigma)$ is continuous and concave on $\mathbb{R}^+$.*

[8] This structure was first characterized in [30] for orthogonal matrices. See also the corresponding result in [12] for a multiple of the identity matrix.

[9] For a diagonal matrix, the constrained and the unconstrained solutions will have the same sign componentwise. However, when there is enough correlation, the signs can be different; see [30, sect. 2.3].

A proof is provided in Appendix C. This result holds for any volatility matrix, and it shows that we can adjust continuously the sparsity of our strategy while keeping the continuity and the concavity of the expected utility as a function of c. In particular, as we relax the constraint, the expected utility of the constrained problem converges to the expected utility of the Merton ratio; see Corollary C.4.

4.2. The constrained plug-in strategy: Sparsity and estimation. We first characterize the expected utility for any time-constant L_1-constrained strategy π_c independent of $W(T)$. Similarly to Lemma 3.1, the expected utility is given by the moment generating function of the mean-variance term $M_\gamma(\pi_c)$.

Lemma 4.6. *Assume that π_c is a time-constant and possibly random strategy independent of $W(T)$ such that $||\pi_c||_1 \le c$. Then*

$$V_\gamma(\pi_c|\mu,\sigma) = K_\gamma \mathbb{E}\left[\exp\left((1-\gamma)\,TM_\gamma(\pi_c)\right)\right]. \tag{4.12}$$

For the unconstrained estimated strategy, the expected utility can degenerate to $-\infty$ for $\gamma \ge 1$ as the number of available risky assets increases. The first advantage of using an L_1-norm is ruling out the degeneracy of the expected utility as the number of assets grows to infinity.

Proposition 4.7. *Assume that $\pi_c^{(d)}$ is a time-constant and possibly random strategy independent of $W^{(d)}(T)$ such that $||\pi_c^{(d)}||_1 \le c$ for each d. Suppose*

$$\lim_{d\to\infty}\max_{i=1,\dots,d}|\mu_i^{(d)} - r| < \infty \qquad \textit{and} \qquad \lim_{d\to\infty}\max_{i,j=1,\dots,d}\Sigma_{ij}^{(d)} < \infty;$$

then $V_\gamma(\pi_c^{(d)}|\mu^{(d)},\sigma^{(d)})$ is bounded from below for $\gamma \ge 1$ as $d \to \infty$.

The statement is proved in Appendix C.

Note that this statement is true for any type of constant L_1-constrained strategy. Hence, this does include plug-in strategies in which drift and possibly even covariances are replaced by estimators.

We see that the L_1-constraint is especially helpful when we face parameter uncertainty for a large number of assets. Indeed, the loss due to estimation accumulates and is so large that we can potentially gain from holding a sparse portfolio. The key to the performance of the L_1-constraint is the trade-off between the gain due to diversification and the loss due to estimation error.

To understand this in terms of loss functions, let π^* be the Merton strategy and π_c^* the solution of the constrained problem with the true drift μ and bound c. When μ is known, the sparsity of the constrained strategy, π_c^*, implies a loss in expected utility. We define the loss factor $L_\gamma(\pi_c^*,\pi^*)$ by the relationship

$$V_\gamma(\pi_c^*|\mu,\sigma) = L_\gamma(\pi_c^*,\pi^*)\,V_\gamma(\pi^*|\mu,\sigma).$$

At the beginning of the investment period, the value of the MLE of the drift $\hat{\mu}$ is fixed, and we obtain the constrained plug-in strategy $\hat{\pi}_c$ by solving (4.5) with $\hat{\mu}$ and the bound c. The resulting strategy $\hat{\pi}_c$ is not normally distributed and is a biased estimator of π_c^*. Thus, there is loss due to estimation on the constrained strategy. The corresponding loss factor $L_\gamma(\hat{\pi}_c, \pi_c^*)$ is defined by

$$V_\gamma(\hat{\pi}_c|\mu, \sigma) = L_\gamma(\hat{\pi}_c, \pi_c^*)\, V_\gamma(\pi_c^*|\mu, \sigma).$$

The total loss consists of both the loss due to insufficient diversification and the loss due to estimation error; i.e., the total loss factor $L_\gamma(\hat{\pi}_c, \pi^*)$ is defined by

$$L_\gamma(\hat{\pi}_c, \pi^*) = L_\gamma(\pi_c^*, \pi^*)\, L_\gamma(\hat{\pi}_c, \pi_c^*),$$

and then we obtain

$$V_\gamma(\hat{\pi}_c|\mu, \sigma) = L_\gamma(\hat{\pi}_c, \pi^*)\, V_\gamma(\pi^*|\mu, \sigma).$$

By Proposition 4.5, the loss factor due to under-diversification, $L_\gamma(\pi_c^*, \pi^*)$, is continuous in c. The loss factor due to estimation, $L_\gamma(\hat{\pi}_c, \pi_c^*)$, and the total loss factor, $L_\gamma(\hat{\pi}_c, \pi^*)$, are also continuous functions in the bound c as the following proposition shows.

Proposition 4.8. *Let t_{obs} be large enough such that $t_{obs}\gamma\Sigma + (1-\gamma)\, Td||\Sigma||_\infty I_d$ is positive definite. Then the maps*

$$c \mapsto L_\gamma(\hat{\pi}_c, \pi_c^*) \qquad \textit{and} \qquad c \mapsto L_\gamma(\hat{\pi}_c, \pi^*)$$

are continuous on $\mathbb{R}^+$.

The statement is proved in Appendix C. The proof consists of showing that the map $c \mapsto V_\gamma(\hat{\pi}_c|\mu, \sigma)$ is continuous. Because the term $(1-\gamma)\, Td||\Sigma||_\infty$ is negative and grows linearly with d, Proposition 4.8 can be applied only if t_{obs} is roughly larger than d. If the L_1-constraint effectively shrinks all the weights towards zero, i.e., $|(\hat{\pi}_c)_i| \leq |\hat{\pi}_i|$ for $i = 1, \ldots, d$ a.s., we can get rid of the dependency on d. In this case, the continuity holds if $t_{obs}\gamma\Sigma + (1-\gamma)\, T||\Sigma||_2 I_d$ is positive definite. This property is, in particular, true for a diagonal covariance matrix.

As we relax the constraint, the strategy π_c^* is more diversified and $L_\gamma(\pi_c^*, \pi^*)$ converges to one, i.e.,

$$L_\gamma(\pi_c^*, \pi^*) \to 1 \quad \text{as } c \to \infty. \tag{4.13}$$

The loss factor due to estimation error, $L_\gamma(\hat{\pi}_c, \pi_c^*)$, behaves in an opposite way. As c increases, more stocks are included in the strategy and the estimation error forces $L_\gamma(\pi_c^*, \hat{\pi}_c)$ to move away from one. By Proposition 4.8, the loss factor $L_\gamma(\hat{\pi}_c, \pi_c^*)$ and the total loss factor $L_\gamma(\hat{\pi}_c, \pi^*)$ both converge to the loss factor of the unconstrained plug-in strategy,[10] i.e.,

$$L_\gamma(\hat{\pi}_c, \pi_c^*) \to L_\gamma(\hat{\pi}, \pi^*) \quad \text{and} \quad L_\gamma(\hat{\pi}_c, \pi^*) \to L_\gamma(\hat{\pi}, \pi^*) \quad \text{as } c \to \infty. \tag{4.14}$$

For t_{obs} and c finite, the loss factors are bigger than one and the aim is to find a bound such that the total loss factor is closest to one. The existence of an optimal bound depends on the structure of the true parameters and the accuracy of the estimator $\hat{\mu}$. We study the behavior of the loss factors in more detail in the next section.

Remark 4.9. As in the unconstrained case, the measure of efficiency, introduced in Definition 3.5, is a function of the loss factor of the constrained plug-in strategy. Since the expected utility is given by (4.12), the following relationship holds true for $\gamma > 1$:

$$\Theta_\gamma(\hat{\pi}_c) = L_\gamma(\hat{\pi}_c, \pi^*)^{\frac{1}{1-\gamma}}.$$

[10]See Corollary C.4 in Appendix C.

5. Simulation study. In this section, we investigate the structure and the performance of the L_1-constrained portfolio when risky assets are correlated. We consider a volatility matrix that is nondiagonal. For this situation we do not have an analytic form for the constrained strategy, and therefore we use simulations to compute its expected utility. Our data set consists of a random sample of $d = 250$ stocks that have been listed at least once on the S&P 500 and have had daily returns for all trading days between January 2001 and December 2011. There is no problem of survivorship bias because our main goal is to show the existence and the uniqueness of an optimal bound c for a given universe of stocks. Throughout this section, we assume an initial normalized endowment of $X(0) = 1$, an annual risk-free rate of $r = 0.02$, and an investment horizon of $T = 1$ year.

5.1. Methodology. Based on the daily log-returns of the stocks, we compute the following unbiased estimators $\tilde{\mu}$, $\tilde{\Sigma}$ of μ and Σ, respectively:

$$\tilde{\mu} = \frac{1}{\Delta}\tilde{\xi} + \frac{1}{2}\operatorname{diag}\left(\tilde{\Sigma}\right), \tag{5.1}$$

$$\tilde{\Sigma}_{ij} = \frac{1}{\Delta(N-1)} \sum_{k=0}^{N-1} \left[R_i(k) - \tilde{\xi}_i\right]\left[R_j(k) - \tilde{\xi}_j\right] \qquad \text{for } i, j = 1, \dots, d, \tag{5.2}$$

with

$$\tilde{\xi}_i = \frac{1}{N}\sum_{k=0}^{N-1} R_i(k) \quad \text{and} \quad R_i(k) = \log\left(\frac{S_i((k+1)\Delta)}{S_i(k\Delta)}\right).$$

The time step is defined by $\Delta = t_{obs}/N$, with $t_{obs} = 11$ years and $N = 2767$ days.

We then assume a standard Merton market as introduced in section 2, where we set $\mu = \tilde{\mu}$ for the drift and $\Sigma = \tilde{\Sigma}$ for the covariance matrix. Furthermore, we evaluate the estimator of the drift $\hat{\mu}$ by sampling from its law; i.e., we use the fact that $\hat{\mu} \sim \mathcal{N}_d(\mu, \Sigma/t_{obs})$. In this setting, we compute the investment weights and the associated expected utility for

- π^*, the unconstrained strategy using the true μ;
- $\hat{\pi}$, the unconstrained plug-in strategy using the estimator $\hat{\mu}$;
- π_c^*, the L_1-constrained strategy using the true μ; and
- $\hat{\pi}_c^*$, the L_1-constrained plug-in strategy using the estimator $\hat{\mu}$.

For a nondiagonal volatility matrix, we do not have analytical results for the L_1-constrained strategy. Therefore, we compute the optimal weights numerically. To do so, we plug the true or the estimated drift in (4.3) and solve the quadratic optimization problem with the least angle regression (LARS) algorithm; see [9]. This algorithm generates the optimal portfolio weights for all binding bounds.

5.2. Computation of the loss function. We want to compute the expected utility of the unconstrained and the constrained strategies both when the drift is known and when it is estimated. For the unconstrained case, we know the explicit form of the expected utility associated with the Merton ratio (2.3) and with the plug-in strategy (3.2). For the constrained case with known drift, the strategy is deterministic and the associated expected utility can be computed directly by using (4.12).

When the drift is estimated, the constrained strategy $\hat{\pi}_c$ is random, and we approximate its expected utility using a Monte-Carlo method. More specifically, we sample M independent and identically distributed (i.i.d.) realizations of the MLE of the drift. Then, for each realization, we solve for the associated constrained strategy and obtain i.i.d. realizations $Y^1_{\gamma,c}, \ldots, Y^M_{\gamma,c}$ of $Y_{\gamma,c}$ given by

$$Y_{\gamma,c} = \exp\left((1-\gamma)\, T\left(\hat{\pi}_c^\top (\mu - r1) - \frac{\gamma}{2}\hat{\pi}_c^\top \Sigma \hat{\pi}_c\right)\right).$$

For K_γ given in (2.4), we define the Monte-Carlo estimator of the expected utility by

$$\bar{V}_{\gamma,M}\left(\hat{\pi}_c|\mu,\sigma\right) = K_\gamma \frac{1}{M}\sum_{i=1}^{M} Y^i_{\gamma,c},$$

which approximates $V_\gamma\left(\hat{\pi}_c|\mu,\sigma\right) = K_\gamma \mathbb{E}\left(Y_{\gamma,c}\right)$; see (4.12). The accuracy of the method is measured by the standard deviation of the Monte-Carlo estimator given by

$$\sqrt{\mathbf{Var}\left(\bar{V}_{\gamma,M}\left(\hat{\pi}_c|\mu,\sigma\right)\right)}.$$

For a fixed c, the standard deviation of the random variable $\bar{V}_{\gamma,M}\left(\hat{\pi}_c|\mu,\sigma\right)$ is a function of γ. Therefore, the number of realizations necessary to attain a given accuracy varies with γ. In this section, the number of realizations, M, is fixed to 5000 for all $\gamma \in [1,7]$.[11]

Furthermore, the loss in expected utility due to estimation $L_\gamma\left(\hat{\pi}_c, \pi^*_c\right)$ is approximated by the ratio of the Monte-Carlo estimator $\bar{V}_{\gamma,M}\left(\hat{\pi}_c|\mu,\sigma\right)$ and the expected utility of the constrained strategy with known parameters $V_\gamma\left(\pi^*_c|\mu,\sigma\right)$. To remove the influence of the multiplicative factor $V_\gamma\left(\pi^*_c|\mu,\sigma\right)$ on the accuracy of the estimation of the loss factors, we apply the logarithmic transformations

$$\ell_\gamma\left(\pi^*_c, \pi^*\right) = \log\left(L_\gamma\left(\pi^*_c, \pi^*\right)\right), \quad \ell_\gamma\left(\hat{\pi}_c, \pi^*_c\right) = \log\left(L_\gamma\left(\hat{\pi}_c, \pi^*_c\right)\right).$$

As mentioned in subsection 4.2, the total loss in expected utility is the product of the loss due to under-diversification measured in terms of the factor $L_\gamma\left(\pi^*_c, \pi^*\right)$, and the loss due to estimation, measured in terms of the factor $L_\gamma\left(\hat{\pi}_c, \pi^*_c\right)$. The logarithm of the total loss factor is then given by

$$\ell_\gamma\left(\hat{\pi}_c, \pi^*\right) = \ell_\gamma\left(\pi^*_c, \pi^*\right) + \ell_\gamma\left(\hat{\pi}_c, \pi^*_c\right).$$

Computing the logarithm enables us to have a natural interpretation of the loss factors. Indeed, in this additive setting, the log-loss equals to zero when there is no loss in expected utility. By (4.13) and (4.14), the log-loss factors converge to zero and to the log-loss factor of the unconstrained plug-in strategy, respectively:

$$\ell_\gamma\left(\pi^*_c, \pi^*\right) \to 0,\ \ell_\gamma\left(\hat{\pi}_c, \pi^*_c\right) \to \ell_\gamma\left(\hat{\pi}, \pi^*\right) \quad \text{and} \quad \ell_\gamma\left(\hat{\pi}_c, \pi^*\right) \to \ell_\gamma\left(\hat{\pi}, \pi^*\right) \quad \text{as } c \to \infty.$$

Finally, as the log-convexity implies the convexity of the loss function itself, it is sufficient to study the convexity of the log-loss.

[11]See discussion in Appendix D.

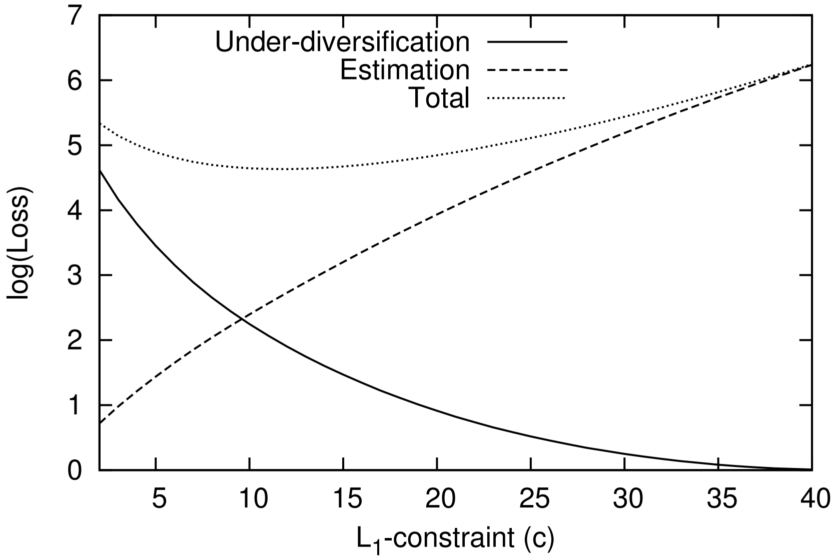

Figure 2. *Plot of the logarithm of the loss factor due to under-diversification $\ell_\gamma(\pi_c^*, \pi^*)$, due to estimation $\ell_\gamma(\hat{\pi}_c, \pi_c^*)$, and of the total loss factor $\ell_\gamma(\hat{\pi}_c, \pi^*)$ as a function of the bound of the L_1-constraint c with $\gamma = 5$.*

Table 1

Comparison between the log-loss in expected utility and the efficiency in the constrained and unconstrained case for different values of risk aversion γ. The log-loss factor is equal to zero when there is no loss.

RRA	Optimal	L_1-constrained		Unconstrained	
γ	c	$\ell_\gamma(\hat{\pi}_c, \pi^*)$	$\Theta_\gamma(\hat{\pi}_c)$	$\ell_\gamma(\hat{\pi}, \pi^*)$	$\Theta_\gamma(\hat{\pi})$
2	29	2.88	0.06	5.82	0.003
3	20	3.86	0.15	7.82	0.02
5	12	4.63	0.31	9.44	0.10
7	8	4.97	0.44	10.14	0.19

5.3. Existence of an optimal bound. Figure 2 depicts the profile of the log-loss factor as a function of c for $\gamma = 5$. We see that the log-loss factors are continuous in c. Furthermore, the total log-loss, $\ell_\gamma(\hat{\pi}_c, \pi^*)$, is convex in c and minimized at $c^* = 29$. Hence, the total loss factor $L_\gamma(\hat{\pi}_c, \pi^*)$ is also convex and minimized at c^*. Equivalently, the expected utility of the constrained plug-in strategy is maximized at this optimal bound c^*.

Table 1 shows that the optimal bound decreases sharply with γ. Again, as γ increases, we need to use a more restrictive bound to optimally control the loss factor. For each γ, the loss due to under-diversification represents between 41% and 44% of the total loss factor, while loss due to estimation represents between 56% and 59%. Moreover, the L_1-constraint greatly helps the log-loss to be closer to zero. The logarithm of the loss is reduced by roughly 50% at the optimal bound, and the loss factor itself is reduced by at least 95.7%.

In terms of efficiency, the L_1-constraint also helps significantly. For small levels of risk aversion, e.g., $\gamma =$ 2, 3, the efficiency is improved dramatically by imposing an L_1-constraint. For high levels of risk aversion, e.g., $\gamma = 7$, the effect of estimation is less important, as a smaller fraction is invested in the risky assets. In this case, holding an L_1-sparse portfolio doubles the efficiency of the investor.

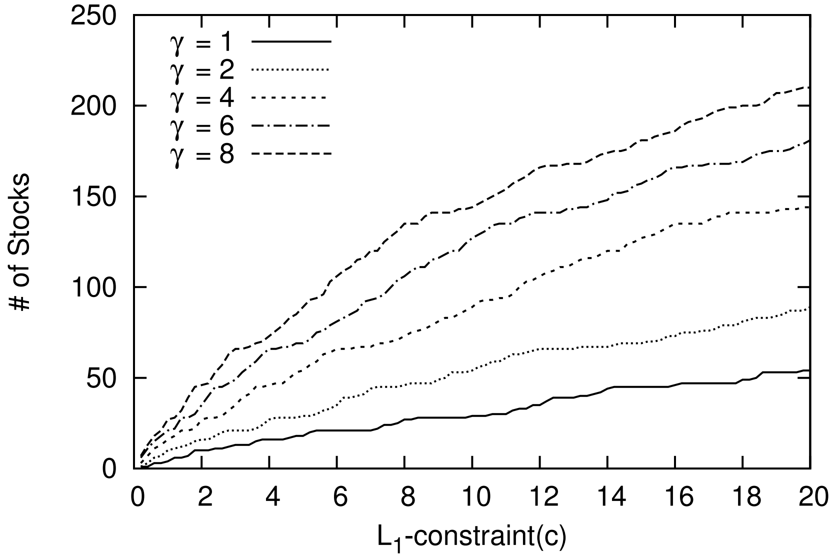

Figure 3. *Number of stocks invested in as a function of the bound c of the L_1-constraint for different RRA coefficients γ. We have chosen $r = 0.02$ for the annual risk-free rate.*

5.4. Structure and stability of the L_1-constrained strategy. Regarding the sparsity of the constrained strategy, the results of subsection 4.1 extend to our general covariance matrix. Figure 3 shows that the number of stocks invested in increases with γ. For instance, at $c = 7$, we invest in 44 stocks with $\gamma = 2$ and in 93 stocks with $\gamma = 6$.

On the one hand, if γ is large, we are less constrained, as a smaller fraction is invested in each selected stock, and this enables us to hold a more diversified portfolio. On the other hand, if γ is small, we still take some relatively strong positions at the cost of having a very sparse portfolio.

As we relax the constraint, we notice that the number of stocks invested in increases stepwise. The steps are especially long for small γ as we take large positions. In effect, c needs to be increased significantly until new stocks are added to the portfolio.

For independent stocks, we should invest in stocks with the highest absolute excess returns up to a certain index k, which is increasing in γ (Theorem 4.4). The weights of the selected stocks are shrunk towards zero, while the other weights are set to zero. If stocks with highest absolute returns have positive excess returns, the portfolio consists exclusively of long positions. and the L_1-constraint acts only as a restriction on the number of stocks.

Table 2 reports the composition on the L_1-constrained portfolio for correlated stocks when the drift μ is known. Because of the structure of the covariance matrix, stocks with negative returns are selected, although their absolute excess return is close to zero. Indeed, the L_1-constraint selects the stocks with the highest returns and then jumps to stocks with negative excess returns. In contrast to the no-short sale constraint, we still keep a limited proportion of short positions. The short positions represent 47% of the amount of the L_1-norm of the unconstrained portfolio, while they are reduced to 30% for the constrained case when we use L_1-bound $c = 3$. Hence, the constraint controls both the sparsity and the proportion of short positions; see also Remark 4.2.

Table 2

Characteristics of the 21 stocks selected for the constrained strategy π_c^* with known parameters. The first column corresponds to the general index of stock i. The second column corresponds to the global rank in term of excess returns. The third column shows the associated excess return. The fourth and fifth columns report the weights. Parameters: Bound of the L_1-constraint $c = 3$, coefficient of RRA $\gamma = 2$, annual risk-free rate $r = 0.02$.

Stock index i	Excess return Rank	$\mu_i - r$	Unconstrained weight π_i^*	Constrained weight $(\pi_c^*)_i$
238	1	0.60	0.57	0.39
216	2	0.43	0.18	0.13
199	3	0.41	0.57	0.33
150	4	0.39	0.52	0.18
234	5	0.37	0.66	0.21
224	6	0.36	0.57	0.14
112	7	0.34	0.40	0.18
211	8	0.33	0.52	0.02
168	9	0.32	1.00	0.28
9	12	0.29	0.48	0.04
183	15	0.27	0.58	0.17
175	17	0.25	1.44	0.06
219	222	0.01	−0.27	−0.03
64	236	−0.02	−1.19	−0.17
177	238	−0.02	−0.55	−0.08
245	240	−0.03	−1.06	−0.25
93	241	−0.03	−0.18	−0.03
120	242	−0.03	−0.10	−0.02
249	245	−0.03	−0.09	−0.02
198	249	−0.08	−0.17	−0.11
222	250	−0.10	−0.38	−0.17

Table 3

This table reports the expected value (and standard deviation) of the number of stocks invested in, the number of shorts positions, the fraction in L_1-norm of short positions, and the average of the standard deviation of the weights. The quantities presented are the sample mean and standard deviation over $M = 5000$ realizations. The optimal bound c of the L_1-constrained strategy is chosen as in Table 1. The unconstrained Merton plug-in strategy corresponds to $c = \infty$. Parameters: Coefficient of RRA $\gamma = 2$, annual risk-free rate $r = 0.02$.

	Expected value of			Weights
$\gamma = 2$	$\#\{i : (\hat{\pi}_c)_i \neq 0\}$	$\#\{i : (\hat{\pi}_c)_i < 0\}$	$\lVert\hat{\pi}_c^-\rVert_1/\lVert\hat{\pi}_c\rVert_1$ (%)	Std. dev.
Merton	250 (0.00)	120.92 (4.81)	48.36 (0.47)	0.66
L_1, $c = 29$	86.39 (5.78)	40.70 (4.27)	45.75 (1.71)	0.21

In Table 3, we compare the structure of the plug-in strategies when the drift is estimated using (5.1) and $\gamma = 2$. For the L_1-constrained strategy, the bound c is chosen to be the optimal bound minimizing the loss in expected utility, as in Table 1. We report the expected value of the number of stocks invested in, the number of shorts positions, and the fraction in L_1-norm of short positions. At the optimal bound, the number of stocks and short positions is reduced from 250 to approximately 86 stocks and from approximately 121 to 41 positions, respectively.

For the unconstrained strategy, the number and the fraction in L_1-norm of short positions both represent 48.36% of the positions and the L_1-norm, respectively. We demonstrated in Table 2 that, by taking a restrictive bound, the proportion of short positions is reduced substantially. At the optimal bound, this is not the case. Indeed, the number of short positions and the fraction of short positions represents 47.11% and 45.75% of the L_1-norm.[12] Therefore, at the optimal level of diversification, the magnitude of the short positions between the unconstrained and the constrained strategies is essentially the same. This shows the superiority of the L_1-norm over the no-short sale constraint. We are able to control parameter uncertainty while keeping short positions, representing almost half of the portfolio in L_1-norm.

To measure the stability of the previous quantities, we also report their standard deviation. The unconstrained strategy invests in all available, i.e., here 250, stocks. For the L_1-constrained strategy with a fixed bound, the number of stocks invested in is random as it depends on the values of $\hat{\mu}$. The standard deviation of stocks invested in is 5.78 or, equivalently, it represents 6.7% of the expected number of stocks held in the portfolio. Compared to the unconstrained strategy, the variability of the number and the fraction in L_1-norm of short positions is proportionally higher for the L_1-constrained strategy. While the structure of the portfolio is more sensitive to the estimation of the drift, the overall stability of the portfolio is improved. Indeed, the average standard deviation of the weights is 68.2% smaller for the constrained strategy. Since the L_1-norm of the weights is bounded, extreme positions are forbidden, and in turn variability is reduced.

6. Out-of-sample study.

Our goal is to investigate the out-of-sample structure and performance of the unconstrained and the optimally constrained plug-in strategies using empirical data. Therefore we will no longer make any parametric assumptions about the evolution of the asset prices.

From the theoretical considerations in the previous sections we have learned the following: We should rebalance our portfolio as frequently as possible (i.e., continuously in the best case). We will therefore trade daily on daily data. To reduce the effect of parameter uncertainty, we should choose a suitable bound c for the L_1-constraint. This is what we do by selecting the bound maximizing the utility, as we will discuss in detail in subsection 6.2. Finally, we measure the performance of the strategies in terms of (expected) utility of terminal wealth.

When looking at the out-of-sample results we need to keep in mind that the performance of the strategies is now also affected by the effects of discrete trading and discrete observations, which was not the case in the previous sections. This is a problem that arises with all continuous-time models.

6.1. The data and general methodology.

Our data set consists of the stocks that have been listed at least once in the S&P 500 and have had daily returns for all trading days between January 2001 and December 2011.

We test our method between 2006 and 2011. At the beginning of each year we select randomly a sample of 250 stocks and fix it as the universe of stocks in which to invest. Based on the five previous years of daily returns, we estimate the drift and volatility matrix of these stocks. We calibrate the optimal bound of the L_1-constraint as outlined in subsection 6.2.

Trading now takes place over intervals of a length of one month. At the beginning of

[12]We find similar results for $\gamma = 3, 5, 7$.

Table 4

This table reports the summary statistics of the optimal bounds c for the leave-one-block-out (LOB) and the cross-validation (CV) methods for the three time periods that we consider. Parameters: Coefficient of RRA $\gamma = 2$, investment horizon $T = 1/12$, and annual risk-free rate $r = 0.02$.

	Optimal L_1-bound		
	1st quartile	Median (mean)	3rd quartile
2006–2007			
LOB	0.78	21.65 (65.00)	68.42
CV	10.85	12.00 (13.39)	15.02
2008–2009			
LOB	0.00	1.80 (29.78)	23.20
CV	3.93	10.90 (12.82)	20.32
2010–2011			
LOB	0.65	5.10 (44.63)	91.15
CV	2.20	3.25 (3.27)	4.60

each month, we assume a normalized initial endowment of one unit of cash and an annual risk-free rate of $r = 0.02$. At the end of the month we record the terminal utility. We average the terminal utilities that we obtain from trading 24 times consecutively over a one-month interval, i.e., a two-year period. We consider three consecutive two-year time periods and compare the performance of the plug-in strategies to the equally weighted portfolio, which is a hard benchmark to beat; see [7].

6.2. Choosing the L_1-bound. To find a suitable (and ideally optimal) bound for the L_1-constraint c, we present two alternative methods, namely a method we call the *leave-one-block-out* (LOB) method and another called a *cross-validation* (CV) method.

For both methods we start as follows. On the first trading day of each month, we divide the multivariate time series of daily returns of the five previous years into 60 blocks of one month. Based on the 59 first blocks, we estimate the drift and the volatility matrix using (5.1) and (5.2). Next, we compute the constrained plug-in strategy in the interval $[0, ||\hat{\pi}||_1]$ for each value of the bound c that is on a grid with grid size $\Delta c = 0.1$. Finally, we invest (rebalancing daily) on the remaining block with respect to the constrained plug-in strategy (see also (6.1)).

In the LOB method, we select the bound c maximizing the utility of final wealth. In the CV method, we repeat this procedure 1000 times by taking a random sample with replacement of the 60 blocks, and we select the bound c maximizing the average utility of final wealth.

As the method of estimation is the same for all months, estimation risk is constant through time, and the variation of the optimal bound depends mainly on the market conditions of the calibration period.

Table 4 contains the numerical results for the optimal L_1-bounds using the two different methods. For the LOB method, the optimal bound varies widely. While the first quartile stays small over all periods, the median (mean) and the third quartile are of different magnitude within and across each period. Moreover, there are five months in 2008, and three months in 2009, with an optimal bound equal to zero. During these months, the strategy consists of holding only the risk-free asset.

In 2006–2007 and 2010–2011, the evolution of the optimal bound calibrated with the CV method is stable as the difference between the first and third quartiles is small. In 2008–2009, there is a large variation of the optimal bound because it is a period of transition. Indeed, the mean over time of the optimal bound is 20.62 in 2008 and 5.03 in 2009. Furthermore, we observe a substantial drop in the median (mean) of the optimal bound, from 12.00 (13.39) in 2006–2007 to 3.25 (3.27) in 2010–2011. In 2010–2011, the CV method incorporates the fact that preceding years were very volatile, and it delivers a more conservative and restrictive bound.

6.3. Performance of the plug-in strategies. At the beginning of each month we pick the optimal L_1-bound which was identified with one of our two methods. We estimate the drift and the volatility matrix using (5.1) and (5.2) based on the past five years of daily returns. Then we compute the unconstrained and constrained plug-in strategies. The strategies are then held constant over a period of one month, i.e., $T = 1/12$. This means we rebalance the positions daily to keep the weights constant over the investment period. For each month M and strategy π^M, which is constant in time over the month, the dynamics of the wealth denoted by X^M is given by

$$X^M(t+1) = X^M(t)\left(1 + r + \sum_{i=1}^{d} \pi_i^M \left(r_i(t+1) - r\right)\right), \tag{6.1}$$

where $r_i(t+1)$ is the simple net return of asset i between t and $t+1$. Since there are 21 trading days in one month, we have $t = 0, \dots, 20$. At the end of each month M, we obtain a utility of final wealth $U_\gamma\left(X^M(21)\right)$.

In Table 5, we report the summary statistics of the utility of final wealth for three blocks of two years, i.e., $M = 1, \dots, 24$ in each block. Over all periods, the unconstrained plug-in strategy performs poorly because of its extreme variance. For instance, the unconstrained strategy gives its highest return in August 2007 and directly leads to bankruptcy a month later. Furthermore, the standard deviation of monthly returns of the wealth is at its peak in 2006–2007 with a magnitude of 1219.95%. However, on a smaller scale, its remains between 89.61% and 120.02% over the subsequent periods. As a result, from January 2008 to May 2009, the wealth reaches zero in almost every month. As we measure the performance with a power utility function with $\gamma > 1$, hitting zero for the wealth translates into infinitely negative utility.

The constrained plug-in strategy calibrated with a bound with the LOB method is halfway between the unconstrained strategy and the constrained strategy calibrated with the CV method. It is the most successful strategy for two periods, with a utility of -0.74 in 2006–2007 and of -0.81 in 2010–2011. This performance comes at the cost of a very large variance of monthly returns. In terms of standard deviation of utility of final wealth and monthly returns, the strategy is similar to the unconstrained portfolio. Therefore, amidst the financial crisis of 2008–2009, the wealth also hits zero.

The constrained plug-in strategy calibrated with the CV method performs better than the unconstrained strategy in all periods, and it is more stable over time than both the unconstrained and the L_1-constrained with LOB strategies. Although, on average, its returns have a larger standard deviation than the equally weighted portfolio, the mean utility of final

Table 5

This table reports the summary statistics of the utility and the monthly returns of final wealth out-of-sample. The quantities presented are computed over each block of 24 months. The optimal bounds c of the L_1-constrained strategies are calibrated using the leave-one-block-out (LOB) and cross-validation (CV) methods. Parameters: Coefficient of RRA $\gamma = 2$, initial wealth $X^M(0) = 1$, investment horizon $T = 1/12$, and annual risk-free rate $r = 0.02$.

	Utility of final wealth			Return of portfolio (%)		
	Min.	Mean (std. dev.)	Max	Min.	Mean (std. dev.)	Max.
			2006–2007			
Merton	$-\infty$	$-\infty$	-0.02	-100	290.84 (1219.95)	5765.79
L_1, c with LOB	-1.00	-0.74 (0.32)	-0.02	0.17	355.86 (1201.51)	5765.79
L_1, c with CV	-1.38	-0.97 (0.26)	-0.43	-27.36	12.88 (41.37)	134.02
EWE	-1.08	-0.99 (0.03)	-0.95	-7.44	1.01 (3.14)	5.09
			2008–2009			
Merton	$-\infty$	$-\infty$	-0.29	-100	-60.07 (89.61)	245.44
L_1, c with LOB	$-\infty$	$-\infty$	-0.12	-100	66.57 (170.09)	744.64
L_1, c with CV	-4.95	-1.46 (0.99)	-0.29	-79.81	0.55 (72.24)	242.19
EWE	-1.27	-1.01 (0.13)	-0.74	-21.35	0.59 (12.95)	34.50
			2010–2011			
Merton	-10.11	-2.43 (2.55)	-0.20	-90.11	6.39 (120.02)	398.11
L_1, c with LOB	-1.00	-0.81 (0.24)	-0.20	0.17	48.25 (99.01)	398.11
L_1, c with CV	-1.17	-0.99 (0.08)	-0.86	-14.28	1.48 (8.22)	15.72
EWE	-1.13	-0.99 (0.06)	-0.87	-11.56	1.07 (6.19)	15.21

wealth is comparable in all periods. The utility of the constrained strategy is higher in 2006–2007, measuring -0.97 versus -0.99. In 2008–2009, the large proportion of estimated short positions hurts the performance of the constrained strategy and has a smaller mean utility of -1.46 versus -1.01. Finally, the mean utility of final wealth is equal for both strategies in 2010–2011 with a value of -0.99.

Given the nature of the out-of-sample study, trading takes place in discrete time. It is well known that strategies that are optimal in continuous time cannot just be discretized to obtain strategies that are optimal in discrete time; see [29] and [2] for extensive discussions. In particular, short-selling is never optimal when trading takes place in discrete time. Since in our study the L_1-constraint does not rule out short-selling, the superiority of the L_1-constrained portfolio over the unconstrained portfolio is not just a consequence of no short-selling.

6.4. Structure and stability of the strategies. In Table 6, we report the mean and standard deviation of stocks invested in, of the short positions, of the fraction of short positions in L_1-norm, and the mean of the monthly portfolio turnover for four different investment strategies: the unconstrained Merton strategy, referred to as *Merton* in Table 6, the L_1-constrained strategy where the bound was computed using the LOB method, the L_1-constrained strategy where the bound was computed using the CV method, and the equally weighted portfolio, referred to as *EW*.

Table 6

This table reports the mean value (standard deviation) of the number of stocks invested in, of the number of shorts positions, of the fraction in L_1-norm of short positions, and the mean monthly turnover for four different investment strategies. The mean is computed over each block of 24 months. The optimal bound c of the L_1-constrained strategy is calibrated with the LOB and CV methods. Parameters: Coefficient of RRA $\gamma = 2$, investment horizon $T = 1/12$, and annual risk-free rate $r = 0.02$.

	Mean value of			
	$\#\{i : \pi_i \neq 0\}$	$\#\{i : \pi_i < 0\}$	$\|\|\pi^-\|\|_1/\|\|\pi\|\|_1$ (%)	Turnover
		2006–2007		
Merton	250 (0.00)	123.13 (5.01)	47.92 (0.71)	∞
L_1, LOB	90.67 (95.38)	42.58 (47.22)	32.25 (19.38)	12.31
L_1, CV	43.58 (9.82)	17.96 (4.84)	35.48 (4.00)	0.78
EW	250 (0.00)	0.00 (0.00)	0.00 (0.00)	0.01
		2008–2009		
Merton	250 (0.00)	120.08 (5.73)	49.39 (0.36)	∞
L_1, LOB	48.75 (75.00)	22.58 (36.21)	34.82 (21.39)	∞
L_1, CV	32.46 (18.69)	14.67 (8.46)	44.74 (5.39)	1.66
EW	250 (0.00)	0.00 (0.00)	0.00 (0.00)	0.02
		2010–2011		
Merton	250 (0.00)	121.71 (5.75)	49.14 (0.24)	28.65
L_1, LOB	83.79 (99.88)	41.38 (50.08)	31.58 (19.66)	4.95
L_1, CV	18.92 (8.73)	9.67 (4.48)	33.47 (10.95)	0.08
EW	250 (0.00)	0.00 (0.00)	0.00 (0.00)	0.01

The monthly portfolio turnover is defined as

$$\text{monthly turnover} \;=\frac{1}{20}\sum_{t=2}^{21}\sum_{i=1}^{d}\left(\left|\pi_i-\pi_i\left(t^-\right)\right|\right),$$

where π_i is the constant target weight for asset i, and $\pi_i\left(t^-\right)$ is the fraction of wealth (i.e., the weight) invested in asset i just before rebalancing at time t. All portfolios are rebalanced daily, and 21 corresponds to the number of trading days in a month. If the wealth hits zero during the month, we set the turnover to ∞ and stop the rebalancing.

For the unconstrained strategy, the structure remains unchanged throughout the whole test period. The mean fraction of short positions of the L_1-norm stays between 120 and 124, the mean ratio $||\pi^-||_1/||\pi||_1$ stays between 47% and 50%, and the portfolio turnover is infinite in both 2006–2007 and 2008–2009. It is finite in 2010–2011 but still large.

For the constrained plug-in strategy with the LOB method, the mean number of risky assets invested in and that of short positions are the largest in 2006–2007 and 2010–2011. In 2008–2009, these quantities are significantly smaller because there are actually several months where there is no investment in risky assets; see also Table 4. Nevertheless, on average, the mean ratio $||\pi^-||_1/||\pi||_1$ stays stable over each period, with a larger standard deviation in 2008–2009. In terms of turnover, the strategy performs poorly. The turnover is very high in 2006–2007 and is infinite in 2008–2009. In 2010–2011, it is reduced but still much larger than the turnover of the constrained plug-in strategy calibrated with the CV method and the equally weighted portfolio. This turnover is high because of the variability of the optimal bound to different market conditions.

For the constrained plug-in strategy calibrated with the CV method, the mean number of risky assets invested in and that of short positions decrease in each period. The mean fraction of short positions of the L_1-norm does not. The mean ratio $||\pi^-||_1/||\pi||_1$ is actually the largest in 2008–2009, with a value of 44.74%, and it is similar in 2006–2007 and 2010–2011, with 35.48% and 33.47%, respectively. For 2008–2009, the mean of estimated excess returns decreases sharply during the year 2008 and implies a significantly larger fraction of short positions in the constrained portfolio. The constrained plug-in strategy calibrated with the CV method is more stable than the constrained plug-in strategy calibrated with the LOB method. The standard deviation of each measure is reduced by at least 40%. Finally, although its turnover is much larger than the equally weighted portfolio turnover in 2006–2007 and 2008–2009, the turnovers are comparable during the period 2010–2011.

6.5. Transaction costs. For a complete analysis of the out-of-sample performance it is interesting to consider transaction costs as well. It is well known that if transaction costs are included, holding the Merton ratio, and hence having to continuously rebalance the portfolio, is no longer optimal. In particular, the investor can find herself in a position where it is optimal to not trade at all. This would happen if her positions would be within the so-called no-trading region, which is used to characterize the optimal trading strategies; see, e.g., [5]. Nevertheless, we can still investigate what transaction costs one would have to pay when using the trading strategies that we have derived in our setting without transaction costs.

Paper [18] distinguishes between implicit costs such as bid-ask spreads, which are proportional to the cash amount of shares traded, and explicit trading costs such as brokerage commissions, which are fixed costs per share traded. This paper measures both types of trading costs as a percentage of the face value of investment. However, [13] argues that brokerage commissions should not be measured as a percentage of face value. Indeed, brokers ignore most available prices and charge commissions in exact cents per share. Moreover, large trades tends to have higher commission fees per share and may also have larger proportional transaction costs because of their potential price impact.

Assuming that proportional trading costs are constant and equal for long and short positions, the total trading costs for one month are defined as

$$\text{monthly trading costs} = \sum_{t=2}^{21}\sum_{i=1}^{d} \kappa S_i\left(t\right)\left|N_i\left(t\right) - N_i\left(t^-\right)\right| + C\cdot\#\{i : N_i\left(t\right) \neq 0\},$$

where κ is the coefficient of charged proportion, $N_i\left(t\right)$ is the number of stock i held at time t, $N_i\left(t^-\right)$ is number of stocks i held just before rebalancing at time t, and C is the commission price per share.

On average, κ represents 0.3% of the face value invested in each stock, and fixed cost commissions C vary between 2 and 5 cents per share. For simplicity, we take $\kappa = 0.3\%$ and $C = 0.025$. As a benchmark, we consider a small investor, with a coefficient of RRA $\gamma = 2$ and an initial wealth of $X_0^M = 25000$, so that trades do not have any price impact. Note that total trading costs are not proportional to the initial wealth invested. Their dependence on the initial wealth is especially important for the plug-in strategies as the number of shares varies widely.

During the financial crisis of 2008–2009, the volatility of the markets leads to a large variability of the positions held in the portfolio even for the constrained plug-in strategies; see Table 6. As a result, the proportional transaction costs are much higher for these strategies compared to the equally weighted portfolio. During 2006–2007 and 2010–2011, the constrained strategies perform, on average, at least as well as the equally weighted portfolio.[13] While having a significantly higher performance during these years, the constrained strategy calibrated with the LOB method faces large transaction costs. For instance, the average total trading costs are 6625.08 and 9963.01 units of cash in 2006 and 2010, respectively. In the same years, the total transaction costs of the constrained strategy calibrated with CV and the equally weighted portfolio are 664.34 and 140.35 in 2006, and 95.42 and 140.68 in 2010. Hence, in 2010, both strategies deliver the same average utility and the transaction costs are, on average, smaller for the constrained strategy.

In summary, outside the financial crisis, L_1-constrained strategies deliver, on average, a higher utility than the equally weighted portfolio. Regarding transaction costs, the constrained strategies have large proportional costs and low fixed commissions. Furthermore, because of fixed commissions, the equally weighted portfolio and the unconstrained portfolio suffer from complete diversification. The trade-off between the two types of costs depends on the size of the positions and the sparsity of the strategies. For the constrained strategy with LOB, total trading costs remain large, while the constrained strategy with CV can have lower trading costs than the equally weighted portfolio. Note that the effect of proportional trading costs will be reduced for a higher level of risk aversion, as the weights in the risky assets, and in turn the number of shares, decrease with γ.

6.6. A note on the estimation of the covariance matrix. In our out-of-sample study we need to estimate the covariance matrix as well. A numerical problem emerges here, since the sample covariance matrix tends to be ill-conditioned, and therefore the computation of Σ^{-1} can be numerically unstable. Successful regularization techniques such as the shrinkage approach of [20] have been proposed to control the conditioning of the matrix.

In our out-of-sample study, we have therefore also investigated the performance of the plug-in strategy based on the covariance estimator proposed by [20]. While it performs better than the simple unconstrained plug-in strategy in 2010–2011, the wealth also hits zero in 2006–2007 and 2008–2009, as it does for the simple unconstrained plug-in strategy.

In certain months, the method of estimation of [20] can have a positive first order effect. For example, in the period 2010–2011, the minimum utility is reached in both cases in January 2011, with a utility of -10.11 without shrinkage and -6.52 with shrinkage. However, with the measure of average utility, there is no difference during 2006–2007 and 2008–2009 because the wealth hits zero at least once. As our measure of performance is nonsymmetric, it smooths the difference between good realizations and assigns very low values to bad realizations.

We also tested the performance of the L_1-constrained strategies with the covariance estimator of [20]. The values of the optimal bound are very close to the results reported earlier, and the performance of the strategy is essentially the same.

For both types of plug-in strategies (i.e., unconstrained and L_1-constrained), we start with

[13] In Table 5, the utilities are given for an investor with a normalized initial wealth $X^M(0) = 1$. For $X^M(0) = 25000$, the results differ by the multiplicative factor $(X(0))^{1-\gamma}$.

Table 7

This table reports the minimum, mean, and maximum of utility of terminal wealth averaged over 24 months for three time periods. Four different strategies are considered: the unconstrained plug-in strategy using (5.1) and (5.2), denoted by Unconst; the unconstrained plug-in strategy using (5.1) and the method proposed by [20] for the covariance matrix, denoted by Unconst, L&W; and the two corresponding strategies where we imposed an L_1-constraint, denoted by L_1 and L_1, L&W, respectively.

	Minimum/mean/maximum of utility over 24 months		
$\gamma = 2$	2006–2007	2008–2009	2010–2011
Unconst	$-\infty/-\infty/-0.02$	$-\infty/-\infty/-0.29$	$-10.11/-2.43/-0.20$
Unconst, L&W	$-\infty/-\infty/-0.03$	- $\infty/-\infty/-0.11$	$-6.52/-1.80/-0.20$
L_1	$-1.38/-0.97/-0.43$	$-4.95/-1.46/-0.29$	$-1.17/-0.99/-0.86$
L_1, L&W	$-1.36/-0.97/-0.43$	$-4.80/-1.48/-0.26$	$-1.16/-0.99/-0.85$

the sample covariance matrix of log-returns, and it is shrunk towards the covariance matrix based on the constant correlation model as discussed in [20]. We provide our numerical results in Table 7.

7. Conclusion.

For a coefficient of RRA bigger than one, we have shown that the loss in expected utility due to parameter uncertainty depends on the coefficient of RRA and the number of risky assets in a highly nonlinear fashion. In particular, as the number of risky assets increases, the loss can become infinite. Therefore, the challenge is to reduce the number of risky assets to limit the exposure to estimation risk when implementing the plug-in strategy. Putting an L_1-constraint on the weights of the plug-in strategy induces sparsity in the portfolio and is an efficient method for reducing the negative effect of parameter uncertainty on its performance.

By characterizing the structure of the L_1-constrained strategy for independent stocks, we show that the level of sparsity is determined by the coefficient of RRA. For a general covariance matrix structure, we demonstrate, based on a simulation study, that there exists an optimal bound minimizing the loss due to estimation for each level of risk aversion. Hence, estimation risk can be efficiently controlled with the L_1-constraint by taking into account the level of risk aversion of the investor.

Based on a CRRA utility maximization framework, we provide an economical justification for using the L_1-constraint as a way to reduce estimation risk. Indeed, for each level of risk aversion, we choose the appropriate level of sparsity and attain the optimal trade-off between the gain of diversification and the loss due to estimation risk. Finally, we show that L_1-constrained strategies can be applied successfully on empirical data.

Appendix A. Proof for section 2.

Proof of Proposition 2.1. The matrix $(\Sigma^{-1})^{(d)}$ is symmetric positive definite with spectrum $(1/\lambda_i^{(d)})_{i=1,\dots,d}$ and can be diagonalized. Since the change of basis of this matrix is orthogonal, the following inequalities hold for each $x \in \mathbb{R}^d$:

$$\underline{m}_\lambda ||x||_2^2 \leq \min_{i=1,\dots,d} \frac{1}{\lambda_i^{(d)}} ||x||_2^2 \leq x^\top \left(\Sigma^{-1}\right)^{(d)} x \leq \max_{i=1,\dots,d} \frac{1}{\lambda_i^{(d)}} ||x||_2^2 \leq \overline{m}_\lambda ||x||_2^2.$$

Thus,

$$\underline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2 \leq \left(\mu^{(d)} - r1^{(d)}\right)^\top \left(\Sigma^{-1}\right)^{(d)} \left(\mu^{(d)} - r1^{(d)}\right) \leq \overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2,$$

and, for $\gamma > 1$,

$$\begin{aligned} &K_\gamma \exp\left(\frac{(1-\gamma)\,T}{2\gamma}\underline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2\right) \\ &\leq V_\gamma\left(\left(\pi^*\right)^{(d)}|\mu^{(d)}, \sigma^{(d)}\right) \\ &\leq K_\gamma \exp\left(\frac{(1-\gamma)\,T}{2\gamma}\overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2\right). \end{aligned}$$

The argument is similar for $\gamma = 1$. ■

Appendix B. Proofs for section 3.

Lemma B.1. *Let π be constant in time, sufficiently integrable, and independent of $W(T)$; then*

$$V_\gamma\left(\pi|\mu,\sigma\right) = K_\gamma \mathbb{E}\left[\exp\left(\left(1-\gamma\right)TM_\gamma\left(\pi\right)\right)\right].$$

Proof of Lemma B.1. For $\gamma > 1$, we rewrite the expected utility

$$V_\gamma\left(\pi|\mu,\sigma\right) = \frac{X\left(0\right)^{1-\gamma}}{1-\gamma}\mathbb{E}\left[\exp\left(\left(1-\gamma\right)\left(r+\pi^\top\left(\mu-r1\right) - \frac{1}{2}\pi^\top\Sigma\pi\right)T + \left(1-\gamma\right)\pi^\top\sigma W\left(T\right)\right)\right]$$

as

$$V_\gamma\left(\pi|\mu,\sigma\right) = K_\gamma \mathbb{E}\left[\exp\left(\left(1-\gamma\right)T\left(\pi^\top\left(\mu-r1\right) - \frac{\gamma}{2}\pi^\top\Sigma\pi\right)\right)Z\left(T\right)\right]$$

with

$$Z\left(T\right) = \exp\left(\left(1-\gamma\right)\pi^\top\sigma W(T) - \frac{\left(1-\gamma\right)^2}{2}\pi^\top\Sigma\pi T\right) \qquad \text{and } \mathbb{E}\left[Z\left(T\right)|\,\pi\right] = 1.$$

Hence,

$$V_\gamma\left(\pi|\mu,\sigma\right) = K_\gamma \mathbb{E}\left[\exp\left(\left(1-\gamma\right)T\left(\pi^\top\left(\mu-r1\right) - \frac{\gamma}{2}\pi^\top\Sigma\pi\right)\right)\right]. \qquad ■$$

Lemma B.2. *Let π be constant in time, sufficiently integrable, and independent of $W(T)$; then*

$$V_\gamma\left(\pi|\mu,\sigma\right) = \mathbb{E}\left[\exp\left(-\frac{\left(1-\gamma\right)\gamma T}{2}\left(\pi-\pi^*\right)^\top\Sigma\left(\pi-\pi^*\right)\right)\right]V_\gamma\left(\pi^*|\mu,\sigma\right).$$

Proof of Lemma B.2. From Lemma B.1 we obtain

$$V_\gamma\left(\pi|\mu,\sigma\right) = K_\gamma \mathbb{E}[\exp\left(\left(1-\gamma\right)TM_\gamma\left(\pi\right)\right)]. \tag{B.1}$$

We write the mean-variance term as

$$\begin{aligned}
M_\gamma(\pi) &= \pi^\top(\mu - r1) - \frac{\gamma}{2}\pi^\top\Sigma\pi \\
&= \pi^\top(\mu - r1) - \frac{\gamma}{2}(\pi - \pi^*)^\top\Sigma(\pi - \pi^*) + \frac{\gamma}{2}\pi^{*\top}\Sigma\pi^* - \gamma\pi^\top\Sigma\pi^* \\
&= \pi^\top(\mu - r1) - \frac{\gamma}{2}(\pi - \pi^*)^\top\Sigma(\pi - \pi^*) + \frac{\gamma}{2}\pi^{*\top}\Sigma\pi^* - \pi^\top(\mu - r1) \\
&= -\frac{\gamma}{2}(\pi - \pi^*)^\top\Sigma(\pi - \pi^*) + \frac{\gamma}{2}\pi^{*\top}\Sigma\pi^* \\
&= -\frac{\gamma}{2}(\pi - \pi^*)^\top\Sigma(\pi - \pi^*) + \frac{1}{2\gamma}(\mu - r1)^\top\Sigma^{-1}(\mu - r1).
\end{aligned}$$

Therefore, by (2.3) and (B.1),

$$V_\gamma(\pi|\mu,\sigma) = L_\gamma(\pi,\pi^*)\,V_\gamma(\pi^*|\mu,\sigma),$$

with

$$L_\gamma(\pi,\pi^*) = \mathbb{E}\left(\exp\left(-\frac{(1-\gamma)\gamma T}{2}(\pi - \pi^*)^\top\Sigma(\pi - \pi^*)\right)\right). \qquad ∎$$

Proof of Lemma 3.1. By (3.1) and (3.2), $\hat{\mu}$ and $\hat{\pi}$ are independent of $W(T)$. Hence, we can apply Lemma B.1, and the result follows. ∎

Lemma B.3. *Let $Y \sim \mathcal{N}_d(0,\Lambda)$, where Λ is symmetric and positive definite. Let $B \in \mathbb{R}^d$, and let $C \in \mathbb{R}^{d\times d}$ be a symmetric positive definite matrix such that Λ and C commute. Let $t \in \mathbb{R}$ and suppose that $\Lambda^{-1} - tC$ is symmetric and positive definite. Then*

$$\mathbb{E}\left[\exp\left(t\left(B^\top Y + \frac{1}{2}Y^\top CY\right)\right)\right] = \exp\left(\frac{1}{2}t^2B^\top\left(\Lambda^{-1} - tC\right)^{-1}B\right)\frac{1}{\sqrt{|I - t\Lambda C|}}.$$

Proof of Lemma B.3.

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(t\left(B^\top Y + \frac{1}{2}Y^\top CY\right)\right)\right] \\
&= \frac{1}{\sqrt{|\Lambda|}}\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\exp\left(t\left(B^\top y + \frac{1}{2}y^\top Cy\right)\right)\exp\left(-\frac{1}{2}y^\top\Lambda^{-1}y\right)dy \\
&= \frac{1}{\sqrt{|\Lambda|}}\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\exp\left(tB^\top y\right)\exp\left(-\frac{1}{2}y^\top\left(\Lambda^{-1} - tC\right)y\right)dy \\
&= \frac{\sqrt{|\left(\Lambda^{-1} - tC\right)^{-1}|}}{\sqrt{|\Lambda|}}\mathbb{E}\left[\exp\left(tB^\top Z\right)\right] \qquad \text{with } Z \sim \mathcal{N}_d\left(0,\left(\Lambda^{-1} - tC\right)^{-1}\right) \\
&= \frac{1}{\sqrt{|I - t\Lambda C|}}\exp\left(\frac{1}{2}t^2B^\top\left(\Lambda^{-1} - tC\right)^{-1}B\right). \qquad ∎
\end{aligned}$$

Proof of Theorem 3.2. By (3.1) and (3.2), $\hat{\pi}$ is independent of $W(T)$. By Lemma B.2, we need to compute the loss factor

$$\mathbb{E}\left(\exp\left(-\frac{(1-\gamma)\gamma T}{2}(\hat{\pi} - \pi^*)^\top\Sigma(\hat{\pi} - \pi^*)\right)\right).$$

Since $\hat{\pi} - \pi^* \sim N\left(0, V_0^2\right)$, we can apply Lemma B.3 with $\Lambda = V_0^2 = \frac{1}{\gamma^2 t_{obs}} \Sigma^{-1}$, $B = 0$, $C = \Sigma$, and $t = -(1-\gamma)\gamma T$, and we obtain

$$V_\gamma(\hat{\pi}|\mu,\sigma) = \frac{1}{\sqrt{|I + (1-\gamma)\gamma T V_0^2 \Sigma|}} V_\gamma(\pi^*|\mu,\sigma) = \frac{1}{\sqrt{\left(1 + \frac{(1-\gamma)}{\gamma t_{obs}} T\right)^d}} V_\gamma(\pi^*|\mu,\sigma).$$

Note that Λ and C do indeed commute here. For $\gamma > 1$, the term $1 + \frac{(1-\gamma)}{\gamma t_{obs}} T$ is strictly positive because of the assumption $t_{obs} > T$. ∎

Proof of Corollary 3.3. Assume $t_{obs} > T$ so that the expected utility of the plug-in strategy is well defined. For $\gamma > 1$ we see from (3.6) that

$$\lim_{d\to\infty} L_\gamma\left(\hat{\pi}^{(d)}, (\pi^*)^{(d)}\right) = \infty.$$

By Proposition 2.1 and because the loss factor is positive, the expected utility of the plug-in strategy is bounded above as follows:

$$V_\gamma\left(\hat{\pi}^{(d)}|\mu^{(d)}, \sigma^{(d)}\right) \le L_\gamma\left(\hat{\pi}^{(d)}, (\pi^*)^{(d)}\right) K_\gamma \exp\left(\frac{(1-\gamma)T}{2\gamma} \overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2\right).$$

If the right-hand side of the inequality goes to $-\infty$, the expected utility $V_\gamma(\hat{\pi}|\mu,\sigma)$ degenerates. We have the following equivalences:

$$\begin{aligned}
&\lim_{d\to\infty} L_\gamma\left(\hat{\pi}^{(d)}, (\pi^*)^{(d)}\right) K_\gamma \exp\left(\frac{(1-\gamma)T}{2\gamma} \overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2\right) = -\infty \\
\Leftrightarrow &\lim_{d\to\infty} L_\gamma\left(\hat{\pi}^{(d)}, (\pi^*)^{(d)}\right) \exp\left(\frac{(1-\gamma)T}{2\gamma} \overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2\right) = \infty \\
\Leftrightarrow &\lim_{d\to\infty} \log\left(L_\gamma\left(\hat{\pi}^{(d)}, (\pi^*)^{(d)}\right) \exp\left(\frac{(1-\gamma)T}{2\gamma} \overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2\right)\right) = \infty \\
\Leftrightarrow &\lim_{d\to\infty} -\frac{d}{2} \log\left(1 + \frac{(1-\gamma)T}{\gamma t_{obs}}\right) + \frac{(1-\gamma)T}{2\gamma} \overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2 = \infty \\
\Leftrightarrow &\lim_{d\to\infty} -\frac{d}{2} \log\left(1 + \frac{(1-\gamma)T}{\gamma t_{obs}}\right) - \frac{(\gamma-1)T}{2\gamma} \overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2 = \infty.
\end{aligned}$$

We set

$$x_d = -\frac{d}{2} \log\left(1 + \frac{(1-\gamma)T}{\gamma t_{obs}}\right), \qquad y_d = \frac{(\gamma-1)T}{2\gamma} \overline{m}_\lambda ||\mu^{(d)} - r1^{(d)}||_2^2.$$

Hence, as $||\mu - r1||_2^2$ is in $o(d)$, $\lim_{d\to\infty} x_d - y_d = \infty$. ∎

Proof of Theorem 3.6. Using the results of Theorem 3.2, and the expressions (2.3) and (2.4), we obtain

$$\begin{aligned}
&\frac{\Theta_\gamma(\hat{\pi})^{1-\gamma}}{1-\gamma} \exp((1-\gamma)rT) \exp\left(\frac{(1-\gamma)T}{2\gamma} (\mu - r1)^\top \Sigma^{-1} (\mu - r1)\right) \\
&= L_\gamma(\hat{\pi}, \pi^*) \frac{1^{1-\gamma}}{1-\gamma} \exp((1-\gamma)rT) \exp\left(\frac{(1-\gamma)T}{2\gamma} (\mu - r1)^\top \Sigma^{-1} (\mu - r1)\right) \\
\Longleftrightarrow\ &\Theta_\gamma(\hat{\pi}) = L_\gamma(\hat{\pi}, \pi^*)^{\frac{1}{1-\gamma}}
\end{aligned}$$

for $\gamma > 1$, and for $\gamma = 1$ we obtain

$$\begin{aligned}
&\log\left(\Theta_1\left(\hat{\pi}\right)\right) + rT + \frac{T}{2}\left(\mu - r1\right)^\top \Sigma^{-1}\left(\mu - r\right) \\
&\quad = \log\left(1\right) + rT + \frac{T}{2}\left(\mu - r1\right)^\top \Sigma^{-1}\left(\mu - r\right) - L_1\left(\hat{\pi}, \pi^*\right) \\
&\Longleftrightarrow \Theta_1\left(\hat{\pi}\right) = \exp\left(-L_1\left(\hat{\pi}, \pi^*\right)\right).
\end{aligned}$$

Remark B.4. Because of Lemma B.2, Theorem 3.6 remains true for any (possibly random) strategy constant in time, sufficiently integrable and independent of $W(T)$. In particular, this is the case of the constrained strategies considered in section 4. ■

Appendix C. Proofs for section 4. We will use the following lemma to prove Proposition 4.1.

Lemma C.1. *Suppose $w(t,x)$ is a regular solution of the Hamilton–Jacobi–Bellman (HJB) equation*

$$w_t(t,x) + \sup_{\{\nu \in \mathbb{R}^d : ||\nu||_1 \le c\}} \left\{\mathcal{L}^\nu w(t,x)\right\} = 0 \quad \text{with } w(T,x) = U_\gamma(x), \ x \ge 0, \tag{C.1}$$

where $\mathcal{L}^\nu$ is the differential operator given by

$$\mathcal{L}^\nu w(t,x) = \left(r + \nu^\top\left(\mu - r1\right)\right) x w_x + \frac{1}{2}\nu^\top \Sigma \nu x^2 w_{xx}.$$

Then, the stochastic integral

$$\int_0^t w_x\left(s, X^\pi(s)\right) X^\pi(s)\, \pi^\top(s)\, \sigma dW(s) \tag{C.2}$$

is a martingale for any admissible portfolio weight process $\pi \in \mathcal{A}_c$.

Proof. We use the ansatz[14] that the solution of the HJB equation is of the form

$$w(t,x) = \phi(t)\, U_\gamma(x),$$

with

$$\frac{\phi'(t)}{1-\gamma} + \rho\phi(t) = 0, \quad \phi(T) = 1, \quad \text{and} \quad \rho = \sup_{\{\nu \in \mathbb{R}^d : ||\nu||_1 \le c\}} \left\{r + \nu^\top\left(\mu - r1\right) - \frac{\gamma}{2}\nu^\top \Sigma \nu\right\}.$$

To prove that the stochastic integral (C.2) is a true martingale for any admissible portfolio weight process $\pi \in \mathcal{A}_c$, we show that

$$\mathbb{E}\left(\int_0^T \left(w_x\left(s, X^\pi(s)\right) X^\pi(s)\right)^2 \pi^\top(s)\, \Sigma \pi(s)\, ds\right) < \infty.$$

Let $p = 1 - \gamma < 0$; then

$$\left(w_x\left(s, X^\pi(s)\right) X^\pi(s)\right)^2 = \phi^2(t)\left(X^\pi(s)\right)^{2p}.$$

[14] See [28, subsect. 3.6.1] for a related one-dimensional version of the problem.

Since ϕ is continuous on $[0,T]$ and the process $\pi \in \mathcal{A}_c$, we have

$$\mathbb{E}\left(\int_0^T (w_x(s, X^\pi(s)))^2 (X^\pi(s))^2 \pi^\top(s)\Sigma\pi(s)\,ds\right)$$
$$\leq c^2 ||\Sigma||_\infty \max_{t\in[0,T]} |\phi^2(t)|\mathbb{E}\left(\int_0^T (X^\pi(s))^{2p}\,ds\right),$$

where $||\Sigma||_\infty = \max\left\{||\Sigma x||_\infty : x \in \mathbb{R}^d \text{ with } ||x||_\infty = 1\right\}$. Hence, as $(X^\pi(s))^{2p}$ is positive, by Fubini's theorem it remains to show that $\mathbb{E}((X^\pi(s))^{2p})$ is bounded by a continuous function of time and the result will follow:

$$(X^\pi(s))^{2p} = X^{2p}(0)\exp\left(2p\left(\int_0^s \pi_u^\top(\mu - r1) - \frac{1}{2}\pi_u^\top\Sigma\pi_u du + \int_0^s \pi_u^\top \sigma dW_u\right)\right)$$
$$= X^{2p}(0)\exp\left(2p\left(\int_0^s \pi_u^\top(\mu - r1) + \left(p - \frac{1}{2}\right)\pi_u^\top\Sigma\pi_u du\right)\right)\mathcal{E}\left(2p\pi^\top\sigma\right)(s).$$

For $\pi \in \mathcal{A}_c$, the process $\left(\mathcal{E}\left(2p\pi^\top\sigma\right)(t)\right)_{(0\leq t\leq T)}$ is an exponential martingale, as it verifies the Novikov condition. Hence, we define the probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ as the Radon–Nikodym derivative

$$\frac{d\mathbb{Q}}{d\mathbb{P}}|\mathcal{F}_T = \mathcal{E}\left(2p\pi^\top\sigma\right)(T).$$

As $\mathbb{Q}$ is equivalent to $\mathbb{P}$ and $\pi \in \mathcal{A}_c$, we have for $p < 0$,

$$\mathbb{E}\left((X^\pi(s))^{2p}\right) = \mathbb{E}^{\mathbb{Q}}\left(X^{2p}(0)\exp\left(2p\left(\int_0^s \pi_u^\top(\mu - r1) + \left(p - \frac{1}{2}\right)\pi_u^\top\Sigma\pi_u du\right)\right)\right)$$
$$\leq X^{2p}(0)\exp\left(2p\left(\left(p - \frac{1}{2}\right)c^2||\Sigma||_\infty - c||\mu - r1||_\infty\right)s\right).$$

The last term is continuous in the variable s and this concludes the proof. ■

Proof of Proposition 4.1. Let $v(t,x)$ be the value function

$$v(t,x) = \sup_{\pi\in\mathcal{A}_c} \mathbb{E}[U\left(X^{\pi,t,x}(T)\right)].$$

We want to show that the value function of problem (4.1) is equal to the solution of the associated HJB equation $w(t,x)$. Since $w(t,x)$ is a regular solution, we can apply Itô's formula; we have the following decomposition for $\pi \in \mathcal{A}_c$:

$$U_\gamma\left(X^{\pi,t,x}(T)\right) = w\left(T, X^{\pi,t,x}(T)\right)$$
$$= w(t,x) + \int_t^T w_t\left(s, X^{\pi,t,x}(s)\right) + \mathcal{L}^{\pi(s)}w\left(s, X^{\pi,t,x}(s)\right)ds$$
$$+ \int_t^T w_x\left(s, X^{\pi,t,x}(s)\right)X^{\pi,t,x}(s)\pi^\top(s)\sigma dW(s).$$

Note that the first integral is negative, as $w(t,x)$ is the solution of the HJB equation. By Lemma C.1 we know that the stochastic integral is a true martingale. Hence

$$\mathbb{E}\left(U\left(X^{\pi,t,x}(T)\right)\right) \leq w(t,x) \Rightarrow v(t,x) \leq w(t,x) \quad \forall (t,x) \in [0,T]\times[0,\infty).$$

Furthermore, the optimal control of the HJB equation is given by

$$\pi_c^* = \arg\max_{||\nu||_1 \leq c} \left\{ r + \nu^\top (\mu - r1) - \frac{\gamma}{2}\nu^\top \Sigma \nu \right\}.$$

By definition of the value function, we have

$$w(t,x) = \mathbb{E}\left(U\left(X^{\pi_c^*,t,x}(T)\right)\right) \leq v(t,x) \quad \forall (t,x) \in [0,T]\times[0,\infty).$$

This implies $v = w$. Therefore the original optimization problem (4.1) is equivalent to the HJB equation. As the optimal control is deterministic and constant in (C.1), the dynamic problem (4.1) reduces to the static problem (4.2). ■

Proof of Theorem 4.4. The proof of this result is divided into two parts. First, we identify the shape of the optimal constrained weights. Second, we characterize the shrinkage intensity.

Since the constraint is binding, it is equivalent to the minimization of the Lagrangian:

$$\min_\pi \frac{1}{2}\left\|\sqrt{\gamma}\sigma^\top \pi - \frac{1}{\sqrt{\gamma}}\sigma^{-1}(\mu - r1)\right\|_2^2 + a||\pi||_1, \quad a \geq 0.$$

As the matrix σ is diagonal, we can optimize term by term. For each $i = 1,\dots,d$,

$$\min_{\pi_i} \frac{1}{2}\left(\sqrt{\gamma}\sigma_i\pi_i - \frac{\mu_i - r}{\sqrt{\gamma}\sigma_i}\right)^2 + a|\pi_i| = \min_{\pi_i} \frac{\gamma}{2}\sigma_i^2\left(\pi_i - \frac{\mu_i - r}{\gamma\sigma_i^2}\right)^2 + a|\pi_i|.$$

Therefore, for each $i = 1,\dots,d$, the optimal solution $(\pi_c^*)_i$ is the proximal mapping of the previous minimization problem. For the absolute value function $|\cdot|$, the proximal mapping corresponds to the soft-thresholding operator.[15] It is given by computing the stationary point of the objective function for $\pi_i > 0$ and $\pi_i < 0$, and we get

$$(\pi_c^*)_i = \frac{\operatorname{sgn}(\mu_i - r)}{\gamma\sigma_i^2}\left(|\mu_i - r| - a\right)^+ \qquad \text{for each } i = 1,\dots,d.$$

We can now compute the parameter a. The argument follows from a discussion of [27] (section 5.2) on L_1-constrained regression with an orthogonal matrix. We order the absolute excess returns by decreasing order

$$|\mu_1 - r| \geq |\mu_2 - r| \geq \cdots \geq |\mu_d - r|.$$

[15]The notion of proximal mapping is due to [25]. The terminology of "soft" threshold was first introduced in [8].

Let $k = \min_{i=1,\dots,d}\{|\mu_{i+1} - r| \le a\}$; we have that

$$\begin{aligned}
||\pi^*||_1 - c &= \sum_{i=1}^{d} \frac{|\mu_i - r|}{\gamma\sigma_i^2} - \sum_{i=1}^{d} \frac{(|\mu_i - r| - a)^+}{\gamma\sigma_i^2} \\
&= \sum_{i=1}^{d} \frac{|\mu_i - r|}{\gamma\sigma_i^2} I(|\mu_i - r| \le a) + \sum_{i=1}^{d} \frac{a}{\gamma\sigma_i^2} I(|\mu_i - r| > a) \\
&= \sum_{i=k+1}^{d} \frac{|\mu_i - r|}{\gamma\sigma_i^2} + \sum_{i=1}^{k} \frac{a}{\gamma\sigma_i^2}.
\end{aligned}$$

Using the expression of $||\pi^*||_1$ again, we obtain a. ■

Remark C.2. Note that if the volatility is a multiple of the identity matrix, the result follows directly from [12, Thm. 5.2] by considering $\tilde{\sigma} = \sqrt{\gamma}\sigma$ rather than σ in the optimization problem, which is solved for a logarithmic utility function. Hence, the novelty of this result relies on the more general structure of the volatility matrix and not on the type of utility function considered.

Proof of Proposition 4.5. The domain of both maps is $\mathbb{R}^+$. Indeed, c has to be nonnegative for the constrained problem to be well defined.

1. From Lemma 4.3 it is sufficient to consider problem (4.5). By changing the signs of the objective function and (4.4), problem (4.5) is equivalent to the following minimization problem:

$$\begin{cases} \min_{\pi} \dfrac{1}{2}\pi^\top D\pi + \pi^\top b \\ \text{subject to } ||\pi_c||_1 \le c, \end{cases}$$

   where, in our case, $D = \gamma\Sigma$ and $b = -(\mu - r1)$.
   Next, we represent the L_1-constraint as a set of linear inequality constraints. The L_1-constraint is equivalent to the 2^d constraints of the type

$$(\pm 1, \dots, \pm 1)\,\pi \le c.$$

   Let $\mathbf{1} = (1, \dots, 1)^\top$ be the 2^d-dimensional vector of ones. In matrix form, we have

$$A\pi \le c\mathbf{1},$$

   where $A = (A_{ij})$ with $A_{ij} \in \{1, -1\}$, $i = 1, \dots, 2^d$, $j = 1, \dots, d$, and each possible combination of signs appears only once as a row of A. Therefore, the constrained OLS problem is equivalent to the standard quadratic programming problem

$$\begin{cases} \min_{\pi} \dfrac{1}{2}\pi^\top D\pi + \pi^\top b \\ \text{subject to } A\pi \le c\mathbf{1}. \end{cases}$$

   Since the matrix D is positive definite, the solution map $c \mapsto \pi_c^*$ is continuous, by [21, Cor. 3.1].

2. Let $v(x) = K_\gamma \exp((1-\gamma)Tx)$, $x \in \mathbb{R}$. The function v and the mean-variance term function M_γ are both continuous on $\mathbb{R}$ and $\mathbb{R}^d$, respectively. Since the map $c \mapsto \pi_c^*$ is continuous, the map $c \mapsto V_\gamma(\pi_c^*|\mu,\sigma)$ is continuous as a composition of continuous functions:

$$c \mapsto \pi_c^* \mapsto M_\gamma(\pi_c^*) \mapsto v(M_\gamma(\pi_c^*)) = V_\gamma(\pi_c^*|\mu,\sigma).$$

Note that since π_c^* is deterministic the last equality holds.
To show the concavity of the expected utility as a function of c, we define the functions

$$f(\pi) = \left\| \sqrt{\gamma}\sigma^\top \pi - \frac{1}{\sqrt{\gamma}}\sigma^{-1}(\mu - r1) \right\|_2 \quad \text{and} \quad G(\pi) = ||\pi||_1.$$

Both functions are convex on $\mathbb{R}^d$. This implies by [22, sect. 8.3, Prop. 1] that the function $w(c) = f(\pi_c^*)$ is convex in c. Moreover, by (4.4), $M_\gamma(\pi_c^*) = -\frac{1}{2}w(c) + K$ and the map $c \mapsto M_\gamma(\pi_c^*)$ is concave in c. Finally, as $\gamma > 1$ and $K_\gamma < 0$, the function v is concave and increasing. Hence, the expected utility map $c \mapsto V_\gamma(\pi_c^*|\mu,\sigma)$ is concave in c as a composition of a concave and increasing function with a concave function, $V_\gamma(\pi_c^*|\mu,\sigma) = v(M_\gamma(\pi_c^*))$. ■

Proof of Lemma 4.6. The expected utility of any L_1-constrained strategy π_c is well defined because the L_1-norm of π_c is bounded by c. The result then follows from Lemma B.2. ■

Proof of Proposition 4.7. Let $m = -c||\mu^{(d)} - r1^{(d)}||_\infty - \frac{\gamma c^2}{2}||\Sigma^{(d)}||_\infty$. Since $||\pi_c^{(d)}||_1 \le c$,

$$m \le (\pi_c^{(d)})^\top \left(\mu^{(d)} - r1^{(d)}\right) - \frac{\gamma}{2}\left(\pi_c^{(d)}\right)^\top \Sigma^{(d)}\pi_c^{(d)},$$

where $||\Sigma^{(d)}||_\infty = \max\left\{||\Sigma^{(d)}x||_\infty : x \in \mathbb{R}^d \text{ with } ||x||_\infty = 1\right\}$.

Thus, for $\gamma \neq 1$, we have, using Lemma 4.6,

$$V_\gamma\left(\pi_c^{(d)}|\mu^{(d)},\sigma^{(d)}\right) \ge \frac{1}{1-\gamma}X(0)^{1-\gamma}\exp((1-\gamma)T(r+m)).$$

For $\gamma = 1$,

$$V_1\left(\pi_c^{(d)}|\mu^{(d)},\sigma^{(d)}\right) \ge \log(X(0)) + T(r+m). \quad ■$$

Proof Proposition 4.8. By (3.1), the maximum likelihood of the drift $\hat{\mu}$ is an $\mathbb{R}^d$-valued random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. To compute the plug-in constrained strategy, we solve problem (4.5) with $\hat{\mu}$. In Proposition 4.5 we showed that for a fixed value of the drift, the solution map is continuous. Hence, the solution map of the plug-in strategy $(\omega, c) \mapsto \hat{\pi}_c$ is $\mathcal{F}$-measurable and continuous in c a.s.

As in Proposition 4.5, let $v(x) = K_\gamma \exp((1-\gamma)Tx)$, with $K_\gamma < 0$. Our objective is to show that $|v(M_\gamma(\hat{\pi}_c))| \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Indeed, since the map $c \mapsto v(M_\gamma(\hat{\pi}_c))$ is continuous, we can then establish the continuity of the map $c \mapsto \mathbb{E}(v(M_\gamma(\hat{\pi}_c))) = V_\gamma(\hat{\pi}_c|\mu,\sigma)$ by the dominated convergence theorem.

We start by computing a lower bound for the mean-variance term $M_\gamma(\hat{\pi}_c)$ (inequalities hold a.s.):

$$M_\gamma(\hat{\pi}_c) = \hat{\pi}_c^\top(\mu - r1) - \frac{\gamma}{2}\hat{\pi}_c^\top\Sigma\hat{\pi}_c \geq -||\hat{\pi}_c||_1||\mu - r1||_\infty - \frac{\gamma}{2}||\Sigma||_\infty||\hat{\pi}_c||_1^2.$$

Since $||\hat{\pi}_c||_1 \leq ||\hat{\pi}||_1$, we deduce that

$$M_\gamma(\hat{\pi}_c) \geq -||\hat{\pi}||_1||\mu - r1||_\infty - \frac{\gamma}{2}||\Sigma||_\infty||\hat{\pi}||_1^2,$$

and we define

$$Y = -||\hat{\pi}||_1||\mu - r1||_\infty - \frac{\gamma}{2}||\Sigma||_\infty||\hat{\pi}||_1^2. \tag{C.3}$$

Because the function v is increasing and negative, $v(Y) \leq v(M_\gamma(\hat{\pi}_c))$ and $|v(M_\gamma(\hat{\pi}_c))| \leq |v(Y)|$.

To apply the dominated convergence theorem, we need to show that $|v(Y)| \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Note that

$$\mathbb{E}(|v(Y)|) = -K_\gamma\mathbb{E}\left[\exp\left((\gamma - 1)T\left(||\hat{\pi}||_1||\mu - r1||_\infty + \frac{\gamma}{2}||\Sigma||_\infty||\hat{\pi}||_1^2\right)\right)\right]$$

and by the Cauchy–Schwarz inequality in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ we obtain

$$(\mathbb{E}[|v(Y)|])^2 \leq K_\gamma^2\mathbb{E}[\exp(2(\gamma - 1)T||\hat{\pi}||_1||\mu - r1||_\infty)]\,\mathbb{E}\left[\exp\left((\gamma - 1)\gamma T||\Sigma||_\infty||\hat{\pi}||_1^2\right)\right]. \tag{C.4}$$

Therefore we have to show that the two expectations on the right-hand side of the inequality are finite. Recall that $\hat{\pi} \sim \mathcal{N}_d(\pi^*, V_0^2)$. We set $V_0^2 = \tilde{V}\tilde{V}^\top$ with $\tilde{V} \in \mathbb{R}^{d\times d}$. For $\gamma > 1$ the first expectation in (C.4) can be bounded as follows:

$$\begin{aligned}
&\mathbb{E}[\exp(2(\gamma - 1)T||\hat{\pi}||_1||\mu - r1||_\infty)]\\
&= \mathbb{E}\left[\exp\left(2(\gamma - 1)T||\tilde{V}Z + \pi^*||_1||\mu - r1||_\infty\right)\right] \quad \text{with } Z \sim \mathcal{N}_d(0, I_d)\\
&\leq C\mathbb{E}\left[\exp\left(2(\gamma - 1)T||\mu - r1||_\infty||\tilde{V}||_\infty||Z||_1\right)\right] \quad \text{with } C = \exp(2(\gamma - 1)T||\mu - r1||_\infty||\pi^*||_1)\\
&= C\left(\mathbb{E}\left[\exp\left(2(\gamma - 1)T||\mu - r1||_\infty||\tilde{V}||_\infty|z|\right)\right]\right)^d \quad \text{with } z \sim \mathcal{N}(0, 1)\\
&= C\left(2\exp\left(\frac{a^2}{2}\right)\Phi_z(a)\right)^d \quad \text{with } \Phi_z \text{ the CDF of } z \text{ and } a = 2(\gamma - 1)T||\mu - r1||_\infty||\tilde{V}||_\infty\\
&< \infty.
\end{aligned}$$

Since $||\hat{\pi}||_1^2 \leq d||\hat{\pi}||_2^2$, the second expectation in (C.4) can be bounded as follows:

$$\mathbb{E}\left[\exp\left((\gamma - 1)\gamma T||\Sigma||_\infty||\hat{\pi}||_1^2\right)\right] \leq \mathbb{E}\left[\exp\left((\gamma - 1)\gamma Td||\Sigma||_\infty||\hat{\pi}||_2^2\right)\right],$$

which is finite by Lemma B.3 because $\left(V_0^2\right)^{-1} - (\gamma - 1)\,\gamma T d ||\Sigma||_\infty I_d$ is positive definite as $\left(V_0^2\right)^{-1} = \gamma^2 t_{obs}\Sigma$. This condition is equivalent to $t_{obs}\gamma\Sigma + (1-\gamma)\,Td||\Sigma||_\infty||I_d$ being positive definite. ■

Remark C.3. If $||\hat{\pi}_c||_2 \le ||\hat{\pi}||_2$ a.s., then

$$\hat{\pi}_c^\top \Sigma \hat{\pi}_c \le \lambda_{max}(\Sigma)\,||\hat{\pi}_c||_2^2 \le \lambda_{max}(\Sigma)\,||\hat{\pi}||_2^2 \quad \text{a.s.,}$$

with $\lambda_{max}(\Sigma) = \max\left\{||\Sigma x||_2 : x \in \mathbb{R}^d \text{ with } ||x||_2 = 1\right\}$. Therefore, we can set the lower bound in (C.3) as

$$Y = -||\hat{\pi}||_1||\mu - r1||_\infty - \frac{\gamma}{2}\lambda_{max}(\Sigma)\,||\hat{\pi}||_2^2.$$

Then $|v(Y)| \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ if $t_{obs}\gamma\Sigma + (1-\gamma)\,T\lambda_{max}(\Sigma)\,I_d$ is positive definite.

Corollary C.4. *Under the assumptions of Proposition 4.8, we have the following convergence properties of the loss factors:*

$$L_\gamma(\pi_c^*, \pi^*) \to 1 \quad \text{as } c \to \infty,$$
$$L_\gamma(\hat{\pi}_c, \pi_c^*) \to L_\gamma(\hat{\pi}, \pi^*) \quad \text{and} \quad L_\gamma(\hat{\pi}_c, \pi^*) \to L_\gamma(\hat{\pi}, \pi^*) \quad \text{as } c \to \infty.$$

Proof Corollary C.4. Note that $\pi_c^* \to \pi^*$ as $c \to \infty$. Since π_c^* and π^* are deterministic, by Lemma B.1, $V_\gamma(\pi_c^*|\mu,\sigma) = v(M_\gamma(\pi_c^*))$ and $V_\gamma(\pi^*|\mu,\sigma) = v(M_\gamma(\pi^*))$, where $v(x) = K_\gamma \exp((1-\gamma)\,Tx)$ is continuous. Hence,

$$V_\gamma(\pi_c^*|\mu,\sigma) \to V_\gamma(\pi^*|\mu,\sigma) \quad \text{as } c \to \infty$$

and

$$L_\gamma(\pi_c^*, \pi^*) \to 1 \quad \text{as } c \to \infty.$$

For the constrained plug-in strategy, $\hat{\pi}_c \to \hat{\pi}$ as $c \to \infty$, and the proof boils down to showing that

$$V_\gamma(\hat{\pi}_c|\mu,\sigma) = \mathbb{E}(v(M_\gamma(\hat{\pi}_c))) \to \mathbb{E}(v(M_\gamma(\hat{\pi}))) = V_\gamma(\hat{\pi}|\mu,\sigma) \quad \text{as } c \to \infty.$$

This is true, as we have shown in the proof of Proposition 4.8 that, for all $c \ge 0$, $|v(M_\gamma(\hat{\pi}_c))| \le |v(Y)| \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, where Y is given in (C.3). Therefore, the dominated convergence theorem can be applied to the sequence of random variables $v_n = v(M_\gamma(\hat{\pi}_{c_n}))$, where c_n is any sequence converging to $+\infty$. ■

Appendix D. Number of paths for the Monte-Carlo method. Suppose that we are to estimate $\mu_Y = \mathbb{E}(Y)$ by $\bar{Y}_M = \frac{1}{M}\sum_{i=1}^M Y^i$, where $Y^i, i = 1, \dots, M$, are i.i.d. realizations of Y. By the central limit theorem, the approximate $(1-\epsilon)$-confidence interval (for $0 < \epsilon < 1$) for μ_Y is given by

$$\left[\bar{Y}_M - a\frac{\sigma_Y}{\sqrt{M}}, \bar{Y}_M + a\frac{\sigma_Y}{\sqrt{M}}\right],$$

where $\Phi(a) = 1 - \frac{\epsilon}{2}$ and $\sigma_Y^2 = \mathbf{Var}(Y)$. Now suppose that we want to estimate $\mu_Z = \mathbb{E}(Z)$, with $Z = kY$, k being a constant. In this case, the accuracy of the Monte-Carlo estimator of μ_Z, is given by $\frac{|k|\sigma_Y}{\sqrt{M}}$. Indeed, the $(1-\epsilon)$-confidence interval is

$$\left[\bar{Z}_M - a\frac{\sigma_Z}{\sqrt{M}}, \bar{Z}_M + a\frac{\sigma_Z}{\sqrt{M}}\right] = \left[k\bar{Y}_M - a\frac{|k|\sigma_Y}{\sqrt{M}}, k\bar{Y}_M + a\frac{|k|\sigma_Y}{\sqrt{M}}\right].$$

Therefore, the constant k is shrinking or widening the confidence interval of the Monte-Carlo estimation, depending on whether it is smaller or bigger than one. Furthermore,

$$\mathbb{P}\left(\mu_Z \in \left[k\bar{Y}_M - a\frac{|k|\sigma_Y}{\sqrt{M}}, k\bar{Y}_M + a\frac{|k|\sigma_Y}{\sqrt{M}}\right]\right) = \mathbb{P}\left(\mu_Y \in \left[\bar{Y}_M - a\frac{\sigma_Y}{\sqrt{M}}, \bar{Y}_M + a\frac{\sigma_Y}{\sqrt{M}}\right]\right).$$

For k large, μ_Z is a larger quantity than μ_Y, as $\mu_Z = k\mu_Y$. For a fixed number of realizations, we just estimate a bigger quantity with a bigger error.

If we want to estimate the expected utility for a plug-in strategy, we sample M i.i.d. realizations $Y^1, \dots, Y^M$ of

$$Y_\gamma = \exp\left((1-\gamma)T\left(\hat{\pi}^\top(\mu - r1) - \frac{\gamma}{2}\hat{\pi}^\top\Sigma\hat{\pi}\right)\right).$$

Next, for K_γ given in (2.4), we define the Monte-Carlo estimator of the expected utility by

$$\bar{V}_{\gamma,M}(\hat{\pi}|\mu,\sigma) = K_\gamma\bar{Y}$$

with

$$\bar{Y} = \frac{1}{M}\sum_{i=1}^{M} Y^i.$$

The variance of the Monte-Carlo is then given by

$$\mathbf{Var}\left(\bar{V}_{\gamma,M}\right) = K_\gamma^2\mathbf{Var}(Y_\gamma)/M.$$

The constant and the random variable depend nonlinearly on the parameter γ, and the variance of $\bar{V}_{\gamma,M}$ is nonmonotonic in γ. Hence, the number of realizations to reach a given level of accuracy varies greatly with γ.

As we do not know $\mathbf{Var}(Y_\gamma)$, we have to replace it in the confidence interval by an estimator. We choose

$$s_M^2 = \frac{1}{M-1}\sum_{i=1}^{M}\left(Y^i - \bar{Y}\right)^2.$$

s_M^2 is an unbiased estimator of $\mathbf{Var}(Y_\gamma)$. In our case, we take $M = 5000$, so that the level of accuracy is such that, for all $\gamma \in [1,7]$,

$$\frac{K_\gamma s_M}{\sqrt{M}} \approx \bar{V}_{\gamma,M}(\hat{\pi}_c|\mu,\sigma)/100.$$

The same type of issue occurs for the computation of the loss. We define the loss factor by

$$V_\gamma\left(\hat{\pi}|\mu,\sigma\right) = L_\gamma\left(\hat{\pi},\pi^*\right) V_\gamma\left(\pi^*|\mu,\sigma\right)$$

or, equivalently, by

$$L_\gamma\left(\hat{\pi},\pi^*\right) = \frac{V_\gamma\left(\hat{\pi}|\mu,\sigma\right)}{V_\gamma\left(\pi^*|\mu,\sigma\right)}.$$

Then to compute this loss numerically, one would use the estimate

$$\bar{L}_{\gamma,M}\left(\hat{\pi},\pi^*\right) = \frac{\bar{V}_{\gamma,M}\left(\hat{\pi}|\mu,\sigma\right)}{V_\gamma\left(\pi^*|\mu,\sigma\right)}.$$

Again, here the accuracy of $\bar{L}_{\gamma,M}\left(\hat{\pi},\pi^*\right)$ is very different from the accuracy of the estimation of the expected utility. As $|V_\gamma\left(\pi^*|\mu,\sigma\right)| \ll 1$,

$$\frac{1}{|V_\gamma\left(\pi^*|\mu,\sigma\right)|}\sqrt{\mathbf{Var}\left(\bar{V}_{\gamma,M}\right)} \gg \sqrt{\mathbf{Var}\left(\bar{V}_{\gamma,M}\right)},$$

and the necessary number of steps would have to be changed accordingly. It could be that, even if the expected utility is accurately estimated, the loss factor is not. Hence, we compute the logarithm of $\bar{L}_{\gamma,M}\left(\hat{\pi},\pi^*\right)$. In this case, the logarithm of the ratio is equal to the difference of the logarithms, and the expected utility $\bar{V}_{\gamma,M}\left(\hat{\pi}|\mu,\sigma\right)$ is estimated independently of $V_\gamma\left(\pi^*|\mu,\sigma\right)$.

REFERENCES

[1] F. Bach, S. D. Ahipasaoglu, and A. d'Aspremont, *Convex Relaxations for Subset Selection*, preprint, arXiv:1006.3601 [math.OC], 2010.

[2] N. Bäuerle, S. P. Urban, and L. A. M. Veraart, *The relaxed investor with partial information*, SIAM J. Financial Math., 3 (2012), pp. 304–327.

[3] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris, *Sparse and stable Markowitz portfolios*, Proc. Nat. Acad. Sci. USA, 106 (2009), pp. 12267–12272.

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

[5] M. H. A. Davis and A. R. Norman, *Portfolio selection with transaction costs*, Math. Oper. Res., 15 (1990), pp. 676–713.

[6] V. DeMiguel, L. Garlappi, F. Nogales, and R. Uppal, *A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms*, Management Sci., 55 (2009), pp. 798–812.

[7] V. DeMiguel, L. Garlappi, and R. Uppal, *Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy?*, Rev. Financial Stud., 22 (2009), pp. 1915–1953.

[8] L. Donoho and I. M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455.

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

[10] N. El Karoui, *High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation*, Ann. Statist., 38 (2010), pp. 3487–3566.

[11] J. Fan, J. Zhang, and K. Yu, *Vast portfolio selection with gross-exposure constraints*, J. Amer. Statist. Assoc., 107 (2012), pp. 592–606.

[12] A. Gandy and L. A. M. Veraart, *The effect of estimation in high-dimensional portfolios*, Math. Finance, 23 (2013), pp. 531–559.
[13] M. A. Goldstein, P. Irvine, E. Kandel, and Z. Wiener, *Brokerage commissions and institutional trading patterns*, Rev. Financial Stud., 22 (2009), pp. 5175–5212.
[14] R. Jagannathan and T. Ma, *Risk reduction in large portfolios: Why imposing the wrong constraints helps*, J. Finance, 58 (2003), pp. 1651–1684.
[15] P. Jorion, *Bayes-Stein estimation for portfolio analysis*, J. Financial Quant. Anal., 21 (1986), pp. 279–292.
[16] R. Kan and G. Zhou, *Optimal portfolio choice with parameter uncertainty*, J. Financial Quant. Anal., 42 (2007), pp. 621–656.
[17] I. Karatzas and S. E. Shreve, *Methods of Mathematical Finance*, Springer-Verlag, New York, 1998.
[18] D. B. Keim and A. Madhavan, *The cost of institutional equity trades*, Financial Analysts J., 54 (1998), pp. 50–69.
[19] A. Kempf, O. Korn, and S. Sassning, *Portfolio optimization using forward-looking information*, Rev. Finance, to appear. Published online March 7, 2014.
[20] O. Ledoit and M. Wolf, *Honey, I shrunk the sample covariance matrix*, J. Portfolio Management, 30 (2004), pp. 110–119.
[21] G. M. Lee, N. N. Tam, and N. D. Yen, *Continuity of the solution map in quadratic programs under linear perturbations*, J. Optim. Theory Appl., 129 (2006), pp. 415–423.
[22] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, New York, 1969.
[23] R. C. Merton, *Optimum consumption and portfolio rules in a continuous-time model*, J. Econom. Theory, 3 (1971), pp. 373–413.
[24] R. C. Merton, *On estimating the expected return on the market: An exploratory investigation*, J. Financial Econom., 8 (1980), pp. 323–361.
[25] J. J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.
[26] B. K. Natarajan, *Sparse approximate solutions to linear systems*, SIAM J. Comput., 24 (1995), pp. 227–234.
[27] M. R. Osborne, B. Presnell, and B. A. Turlach, *On the lasso and its dual*, J. Comput. Graph. Statist., 9 (2000), pp. 319–337.
[28] H. Pham, *Continuous-Time Stochastic Control and Optimization with Financial Applications*, Springer, New York, 2009.
[29] L. C. G. Rogers, *The relaxed investor and parameter uncertainty*, Finance Stoch., 5 (2001), pp. 131–154.
[30] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B, 58 (1996), pp. 267–288.
[31] J. Tu and G. Zhou, *Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies*, J. Financial Econom., 99 (2011), pp. 204–215.