# Extending the demand system approach to asset pricing

Thomas Gehrig[1,3] · Leopold Sögner[2,3] · Arne Westerkamp[4]

## Abstract

This article introduces a shrinkage procedure which allows to improve upon the parametric portfolio approach introduced in Brandt et al (Review of Financial Studies 22(9): 3411–3477, 2009) and more general factor conditional frameworks. We analyze optimal investment decisions for constant absolute and constant relative risk aversion. In both preference classes, especially out-of-sample performance of the optimal strategies is rather volatile. In order to reduce parameter and model uncertainty, we augment the optimal strategies by a shrinkage device that pulls the portfolio weights toward a predetermined policy portfolio. Our theoretical approach thereby extends the demand systems approach of Koijen and Yogo (Journal of Political Economy, 127(4):1475–1515, 2019) to more general classes of preferences and provides conditions for the existence of equilibrium. As a side product, we establish that the characteristics-based parametric portfolio approach of Brandt et al. (Review of Financial Studies 22(9): 3411–3477, 2009) can only be justified as optimal investments under exceedingly strong assumptions. In empirical US data, our shrinkage approach outperforms the parametric approach and the naive 1/N-strategy over quite a wide range of levels of absolute and relative risk aversion.

**Keywords** Parametric portfolio policy · Expected utility · Risk aversion

✉ Leopold Sögner
soegner@ihs.ac.at

Thomas Gehrig
thomas.gehrig@univie.ac.at

Arne Westerkamp
arne.westerkamp@iqam.com

[1] Faculty of Business, Economics, and Statistics, University of Vienna, Oskar- Morgenstern-Platz 1, A-1090 Vienna, Austria

[2] Institute for Advanced Studies, Josefstädter Straße 39, A-1080 Vienna, Austria

[3] Vienna Graduate School of Finance (VGSF), Vienna, Austria

[4] IQAM Invest GmbH, Wollzeile 36-38, 1010 Vienna, Austria

🖄 Springer

## 1 Introduction

The concept of characteristics-based parametric portfolio policies introduced by Brandt et al. (2009) provides an attractive reduction technique for determining optimal portfolio choices. By directly applying optimal portfolio weights as functions of observable characteristics, the cumbersome determination of the effects of underlying economic and financial variables on a multitude of moments of the return distributions can be avoided, which according to classical portfolio theory drive optimal portfolio weights. However, without providing a micro-foundation of parametric portfolio choice Brandt et al. (2009) cannot identify the limitations of their approach.

A potential micro-foundation is provided by Koijen and Yogo (2019). In their widely acclaimed contribution, these authors develop an asset pricing model with flexible heterogeneity in asset demand across investors. However, they provide it for a special utility function only, logarithmic utility and hence relative risk aversion of 1. Their framework is especially useful in modeling non-atomic investors such as large institutions and pension funds. In their framework, however, short selling constraints are required, such that optimal portfolio choice reduces to characteristics-based demand, when returns exhibit a factor structure. This allows them to construct and apply an instrumental variable estimator in order to deal with the endogeneity of demand and asset prices. Finally, these authors illustrate the power of their approach on US stock market data and investor holding data from 1980-2017.

Therefore, this article extends the approach of Koijen and Yogo (2019) in order to allow more general preferences in order to provide for a richer and robust micro-foundation of Brandt et al. (2009) for potential applications in portfolio choice. First, in contrast to Koijen and Yogo (2019) we allow for general constant relative risk aversion (relative risk aversion parameter $\gamma \in \mathbb{R}_{>0}$) rather than imposing log-linear utility ($\gamma = 1$). Second, we extend the analysis to the case of constant absolute risk aversion. Doing so allows us to connect the demand system approach directly to the parametric portfolio approach of Brandt et al. (2009). Third, we add the analysis in the absence of short selling constraints in order to analyze and evaluate the empirical relevance of this restriction. In this regard, we identify potential problems of the parametric portfolio policy for low enough levels of constant relative risk aversion in recent US data. Fourth, we show how a shrinkage device can be included in a simple way to "stabilize" the investment strategies and to improve performance in empirical data. Fifth, we show the existence of equilibrium in an economy with heterogeneous agents specifically for the cases of constant absolute risk aversion (CARA, also considered almost recently in Koijen et al. 2023) and constant relative risk aversion (CRRA) preferences. This result still holds if a subset of the agents, which is not necessarily proper, apply the shrinkage device proposed in this article.

Finally, we illustrate the performance of those extensions (in terms of the certainty equivalent and the Sharpe ratio) at the hand of US stock market data.

The basic insights are the following:

- We find that **parametric portfolio policies** (see Brandt et al. 2009) can be derived as optimal portfolio policies only under very restrictive assumptions. Typically, optimal portfolio investments differ from solutions to the characteristics-based approach.
- The case of **constant absolute risk aversion** generates relatively simple solutions because of the absence of wealth effects. We demonstrate that our optimal strategies with shrinkage outperform parametric portfolio policies and a simple 1/*N* investment strategy (see, e.g., DeMiguel et al. 2009).
- In the case of **constant relative risk aversion**, technical pitfalls have to be avoided by imposing restrictions on domains or adapting objective functions for the region of large losses. The necessity of such restrictions is demonstrated empirically at the example of S&P 500 stocks for the USA in the period from 1995 to 2013, especially for low levels of relative risk aversion. Overall, we find that the performance of the "constant relative risk aversion adaptions" is relatively poor for low levels of relative risk aversion $\gamma$. However, the performance is improving for higher levels, both in sample and out of sample. We observe that for moderate and higher $\gamma$ our optimal strategies with shrinkage outperform parametric portfolio policies and a simple 1/*N* investment strategy. For large $\gamma$ the differences in the performance become small.

The demand systems approach can be interpreted as a reduction technique to explain asset prices as a function of a few exogenous characteristics.

Such a reduction technique is expected to reduce numerical complexity and to enhance robustness. Obviously the validity of such a procedure depends on the true underlying economic structure.

Our insights are particularly useful for popular machine learning algorithms (see, e.g., Nagel 2021), since they allow to fuse prior economic knowledge with big data on asset prices and further underlying information sources. Our analysis identifies potential, and empirically relevant pitfalls, and provides solutions to such challenges for algorithmic portfolio optimization. In particular, and in contrast with Nagel (2021), where ridge regression is used to predict returns, we propose an algorithm that allows to shrink toward some specific portfolio weights such as the 1/*N*-portfolio.

The paper is organized as follows: Section 2 provides a literature review. Section 3 presents the basic model. Section 4 presents asset demand based on constant absolute risk aversion (CARA preferences) and develops the conditions for the parametric portfolio policy as an optimal solution to the portfolio investment problem. Section 5 analyzes CRRA-preferences. Section 6 presents asset prices derived in general asset market equilibrium for both, the CARA and the CRRA preferences discussed in the sections before. Section 7 presents an empirical evaluation of the pricing theories at a sample of one-hundred S&P 500 stocks. This chapter also provides robust empirical evidence of potential pitfalls for the unchecked parametric portfolio approach. Section 8 concludes. Appendix contains a section on the properties of the empirical data, while Supplementary Material provides further technical details.

## 2 Literature and relations to machine learning

As already stated in Introduction, in our approach we obtain optimal portfolio weights given some characteristics (abbreviated $\mathbf{x}_{it}$ in the later parts of this article). In the following sections, we also investigate whether these optimal rules are equal to or at least approximately correspond to the parametric portfolio approach of Brandt et al. (2009). In addition, we observe that the optimal rules show poor out-of-sample performance, at least in the empirical data set considered in this article. To improve on this issue, a quadratic penalty function will be included.

Our paper is not the first to discuss issues related to the missing micro-foundations of the parametric portfolio policy approach. Ammann et al. (2016) show that the parametric portfolio policy approach implies unrealistically large amounts of implied short sales and provide conditions to render the approach more empirically appealing, and more in line with the empirical findings of Medeiros et al. (2014). Our contribution complements these earlier studies by providing a micro-foundation for the parametric portfolio policy approach in a factor setting.

We adopt this approach to a S&P 500 sub-sample of 100 assets for the period of 1979-2013 and compare it to the optimal solution implied by the micro-founded model. Other closely related work is Hjalmarsson and Manchev (2012), who consider the special case of mean–variance preferences. We also compare the results with the ad hoc heuristics of the $1/N$-rule (see, e.g., DeMiguel et al. 2009) [1].

Further reduction techniques and methods to stabilize and improve estimates and/or forecasts are tools recently provided in machine learning literature (for an overview, see, e.g., Nagel 2021). For example, in Nagel (2021)[Chapter 4] ridge regression is applied to improve the forecasting performance of a predictive regression model, where a quite large set of exploratory variables is used to predict asset returns. Then, these forecasts are used for portfolio allocation.

Also Kelly et al. (2021) use ridge regression techniques to forecast asset returns by using a large set of predictors. The authors also connect ridge regression to the Moore–Penrose pseudo-inverse (which corresponds to the case where the shrinkage parameter becomes small). In addition, the authors consider the case where the number of regression parameters becomes large and use random matrix theory to obtain asymptotic results (further theoretical results are provided in Hastie et al. 2022). In their empirical analysis, CRSP-data were used. The authors show that using a bulk of "plausibly relevant predictors" in combination with "rich nonlinear models" improves return forecasting and portfolio returns. Nonparametric regression in combination with shrinkage is applied to portfolio allocation in Freyberger et al. (2020).

Alternatively, neural networks—in particular reinforcement learning—can be used to directly optimize the objective function of an investor (see, e.g., Cong et al. 2020). The parametric portfolio approach of Brandt and Santa-Clara (2006); Brandt et al. (2009) can be seen as special case of this machine learning approach (by considering a small number of predictors as well as a linear dependence structure).

---

[1] An overview on reduction techniques is e.g. provided in Thös (2019).

In this article, we augment our objective function (that is, either CARA or CRRA expected utility) by a quadratic penalty term. In contrast with Kelly et al. (2021), Nagel (2021) and a lot of other 'machine learning in finance papers' cited there, the number of predictors remains small in our analysis. The main reason for that is to keep our approach comparable to Brandt et al. (2009) and DeMiguel et al. (2009), that is, whether we can improve on $1/N$ or parametric portfolio policies already with a small number of prediction variables.

We show that for CARA utility our optimization problem exactly corresponds to the optimization problem observed in the case of ridge regression. For constant relative risk aversion, we show that by using a second-order Taylor series approximation of the utility function the optimization problem corresponds to a ridge regression problem. Our approach allows to shrink the portfolio weights toward weights chosen by the investor (such as the equally weighted portfolio). We observe that in our empirical data the implementation of the characteristics-based approach of Koijen and Yogo (2019) requires the application of shrinkage methods to stabilize and improve out-of-sample performance.

As is well known in the literature (see, e.g., James et al. 2017, p. 226), the ridge regression estimator corresponds to the posterior mean of the vector of regression parameters in a Bayesian regression model with normally distributed noise and a normal prior on the regression parameters (e.g., $\breve{\mathbf{w}}_t$ and covariance matrix $\frac{1}{c_p}\mathbf{I}_n$ in Section 5). In our analysis, the vector of regression parameters corresponds to our portfolio weights. The ridge regression methodology easily allows to integrate a-priori information on portfolio weights. The stronger the prior on these weights the more we shrink toward the a-priori weights chosen by the investor. One prominent example is the equally weighted portfolio discussed in DeMiguel et al. (2009). Hence, in contrast with the machine learning approaches discussed above, our approach directly allows to integrate a-priori information on investment weights.

## 3 Model and assumptions

We consider an economy in discrete time $t \in \mathbb{N}$. Denote the one-period return (or yield) of security $i$ from period $t$ to $t+1$ as $r_{it+1}$ and the gross returns as $R_{it+1} := 1 + r_{it+1}, i = 1, \ldots, N$, where $N$ is the number of risky assets traded.[2]

In the case a risk-free asset is traded, we apply the index $i = 0$, its return is $r_{ft+1}$, and the total number of assets is $n = N + 1$ (in sums the summation index 0 is used for the risk-free asset) otherwise $n = N$. Denote the share price of asset $i$ in period $t$ by $P_{it}$ and the number of traded shares by $S_{it}$. Accordingly, the market value of equity of asset $i$ is given by $P_{it}S_{it}$ and aggregate market capitalization reads $\sum_{i=1}^{N} P_{it}S_{it}$, while $\mathbf{P}_t := \left(P_{1t}, \ldots, P_{Nt}\right)^\top$ denotes the vector of share prices.

---

[2] In more general terms, an index set of traded risky securities $\mathbb{I}_t^r$ in period $t$ can be defined. For example, the set $\mathbb{I}_t^r$ contains S&P 500 or CRSP-identifiers. To simplify the analysis, the number of risk assets is kept fixed in the following.

For a given set of weights $w_{it} \in \mathbb{R}$, $i = 0, 1, \ldots, N$, the portfolio return is $r_{pt+1} := \sum_{i=(N-n)+1}^{N} w_{it} r_{it+1}$, with $R_{pt+1} := 1 + r_{pt+1}$ denoting the portfolio's gross return.[3]

We collect observed characteristics in $\mathbf{x}_{it} \in \mathbb{R}^k$, $i = (N-n)+1, \ldots, N$, where $\mathbf{x}_{it}$ could contain endogenous, predetermined and or exogenous variables. For the risk-free asset $\mathbf{x}_{ft}$ is known in period $t - 1$ ($\mathcal{F}_{t-1}$-measureable in mathematical terms[4]).

In particular, we assume that market equity (in the empirical data a stationary transformation of market equity) of asset $i$, that is, $P_{it} S_{it}$, is contained in $\mathbf{x}_{it}$. Following Koijen and Yogo (2019), let $\check{\mathbf{x}}_{it}$ contain these observed variables as well as unobserved variables. We explicitly assume that prior investment weights or amounts invested are not contained in $\check{\mathbf{x}}_{it}$. Let

$$\mathbf{y}_{it} := \begin{pmatrix} \check{\mathbf{x}}_{it} \\ vech(\check{\mathbf{x}}_{it} \check{\mathbf{x}}_{it}^\top) \\ \vdots \end{pmatrix} \in \mathbb{R}^{k_y} \tag{1}$$

collect terms obtained from raising the elements of $\check{\mathbf{x}}_{it}$ by $j = 1, 2, \ldots$. Then, we assume that returns follow from

$$\underbrace{\begin{pmatrix} R_{0t} \\ R_{1t} \\ \vdots \\ R_{Nt} \end{pmatrix}}_{\mathbf{R}_t} = \underbrace{\begin{pmatrix} a_f \\ a_1 \\ \vdots \\ a_N \end{pmatrix}}_{\mathbf{a}} + \underbrace{\begin{pmatrix} \mathbf{A}_f & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1 & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{A}_N \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} \mathbf{y}_{ft} \\ \mathbf{y}_{1t} \\ \vdots \\ \mathbf{y}_N \end{pmatrix}}_{\mathbf{y}_t \in \mathbb{R}^{nk}} + \underbrace{\begin{pmatrix} 0 \\ \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Nt} \end{pmatrix}}_{\tilde{\varepsilon}_t} . \tag{2}$$

$\mathbf{A}_j$, $j = (N-n)+1, \ldots, n$, are $1 \times k_y$-dimensional matrices. In the case, a risk-free asset is traded $\varepsilon_{ft} = 0$, and $R_{ft} = a_f + \mathbf{A}_f \mathbf{y}_{ft}$ already known in period $t - 1$. The vector of noise terms $\tilde{\varepsilon}_t$ contains the $N$-dimensional subvector $\varepsilon_t$ affecting the risky assets. Its expectation is zero and covariance matrix $\boldsymbol{\Sigma}$.

To slightly simplify the analysis and in contrast with Koijen and Yogo (2019), we did not impose a factor structure on the covariance matrix $\boldsymbol{\Sigma}$; however, this simplifying assumption can be relaxed in a straightforward way. [The above paragraphs imply that Assumption 1 of Koijen and Yogo (2019) holds by our model assumptions.] To simplify the notation, an additional index for an investor is only included if necessary. Next we impose

---

[3] If no risk-free asset is traded $w_{ft} = 0$. For vectors and matrices, we apply boldface notation. That is, $\mathbf{x} \in \mathbb{R}^a$ denotes an a-dimensional column vector, while $\mathbf{X} \in \mathbb{R}^{a \times b}$ denotes a matrix with $a$ rows and $b$ columns. $x_{it,j} = [\mathbf{x}_{it}]$ abbreviates the $j$th coordinate of the vector $\mathbf{x}_{it}$. $\mathbf{1}_{N \times 1}$ (for short $\mathbf{1}_{N \times 1}$) and $\mathbf{0}_{N \times 1} =_{N \times 1}$ denote $N$-dimensional column vectors of ones and zeros, respectively. $vech(\mathbf{A})$ transforms the lower triangular part of an $n \times n$ matrix $\mathbf{A}$ into a $n(n+1)/2$-dimensional column vector. $\mathbf{1}_{(A)}$ denotes an indicator function. $u'(x)$ and $v'(x)$ denote the first derivatives of the functions $u(\cdot)$ and $v(\cdot)$ evaluated at $x \in \mathbb{R}$. Given a filtered probability space with filtration $(\mathcal{F}_t)_{t \in \mathbb{N}_0}$, the random variables observed in the periods $1, \ldots, s \leq t$ are $\mathcal{F}_t$-measurable. The $\mathcal{F}_t$-conditional expectation is abbreviated by $\mathbb{E}_t(\mathbf{R}_{t+1})$; the $\mathcal{F}_t$-conditional (co-)variance is $\mathbb{V}_t(\mathbf{R}_{t+1})$.

[4] For definitions, see, e.g., Davidson and MacKinnon (1993).

**Assumption 1** (i) $\mathbf{R}_t$, $\mathbf{y}_t$, and $\boldsymbol{\varepsilon}_t$ are jointly stationary and ergodic.[5] The first and the second moments exist.

(ii) $\mathbf{y}_t$ has full rank covariance matrix.

(iii) The noise term process $(\boldsymbol{\varepsilon}_t)$ follows a martingale difference sequence.

(iv) The covariance matrix $\boldsymbol{\Sigma}$ is finite, symmetric and positive semi-definite.

By (ii), we exclude constant characteristics and colinearities between $\mathbf{y}_t$. That is, no characteristic is redundant. The stronger assumption of a positive definite (conditional) covariance matrix of risky returns is imposed in Section 4 only to obtain a unique optimal investment strategy. Part (i) is important for the empirical implementation of the model, since it avoids technical problems with possibly non-stationary regressors. By (iii), the conditional expectation is affine in $\mathbf{y}_{\text{it}}$. That is, the conditional expectation of the return of asset $i$ is $a_i + \mathbf{A}_i \mathbf{y}_t$.

We consider a sequence of myopic investment problems. There are no trading costs. In each period $t$, $t = 1, 2, \ldots$, an investor is endowed with wealth $e_t > 0$. This wealth can be invested into $n$ alternatives. Portfolio optimization traditionally involves the optimal determination of those weights $w_{\text{it}}$ (or amounts invested into asset $i$, $\phi_{\text{it}}$,) with respect to a utility function, potential endowment and trading constraints. $\phi_{\text{it}} = e_t w_{\text{it}}$ is the amount invested in monetary units in asset $i$, while $\mathbf{w}_t := (w_{1t}, \ldots, w_{Nt})^\top \in \mathbb{R}^N$ and $\boldsymbol{\phi}_t := (\phi_{1t}, \ldots, \phi_{N,t})^\top \in \mathbb{R}^N$ are the investments (investment weights) into the risky assets in the following. Let $\mathcal{W} \subset \mathbb{R}^n$ and $\mathcal{W}_\phi \subset \mathbb{R}^n$ denote the sets of feasible strategies. Hence, $(w_{ft}, \mathbf{w}_t)^\top \in \mathcal{W}$ or equivalently $(\phi_{ft}, \boldsymbol{\phi}_t)^\top \in \mathcal{W}_\phi$ [if no risk-free asset is traded the $\mathcal{W}$ and $\mathcal{W}_\phi$ are such that $w_{ft} = 0$ and $\phi_{ft} = 0$]. Preferences of a typical (or representative) investor are specified by the expected utility (conditional on the information in period $t$) over gross portfolio returns $R_{pt+1} = \sum_{i=(N-n)+1}^{n} w_{\text{it}} R_{\text{it}+1}$, resulting in the optimization problem

$$\max_{(w_{ft}, \mathbf{w}_t)^\top \mathcal{W}} \mathbb{E}_t\big(u(e_t R_{pt+1})\big) = \max_{(w_{ft}, \mathbf{w}_t)^\top \in \mathcal{W}} \mathbb{E}_t\big(u(E_{t+1})\big)$$

$$= \max_{(w_{ft}, \mathbf{w}_t)^\top \in \mathcal{W}} \mathbb{E}_t\left(u\left(e_t\left(1 + \sum_{i=(N-n)+1}^{N} w_{\text{it}} r_{\text{it}+1}\right)\right)\right), \tag{3}$$

where $u(\cdot)$ is a strictly monotone increasing Bernoulli utility function defined on the domain $\mathbb{D} \subset \mathbb{R}$ and $e_t$ the wealth invested in period $t$. We assume that $e_t$, $t = 1, 2, \ldots$, are already given or fixed before any portfolio optimization is performed. Hence, in the optimization problem (3), the $e_t$ invested are deterministic. Given a vector of investment weights $\mathbf{w}_t := (w_{1t}, \ldots, w_{Nt})^\top \in \mathbb{R}^N$ into the risky assets, the $t + 1$ period wealth is $E_{t+1} = e_t \mathbf{w}_t^\top \mathbf{R}_{t+1}$ if no risk-free asset is traded. If a risk-free asset is traded (or depositions and lending in cash are allowed), its gross return will be $R_{ft+1} \geq 0$; $w_{ft}$ is the corresponding proportion of the wealth invested into the risk-free asset at period $t$. In the case a risk-free asset is traded $E_{t+1} = e_t \mathbf{w}_t^\top \mathbf{R}_{t+1} + e_t w_{ft} R_{fT+1}$. To jointly consider both cases, we write $E_{t+1} = e_t \mathbf{w}_t^\top \mathbf{R}_{t+1} + e_t w_{ft} R_{fT+1}$, and assume that $w_{ft} = 0$ if no risk-free asset is traded. As already stated above, note that the next

---

[5] For definitions, see, e.g., Klenke (2008), Chapter 20.

period's amount invested, $e_{t+1}$, need not be equal to the realization of $E_{t+1}$. By contrast, $e_{t+1} > 0$ is some non-random real number.

The model presented in the above paragraphs is closely related to the model of Koijen and Yogo (2019). In contrast, however, we allow for the case without short selling constraints.[6] To simplify notation, we do not consider time varying $\mathbf{A}$, $\mathbf{a}$, and $\boldsymbol{\Sigma}$.[7]

Since both, Koijen and Yogo (2019) and our model, result in optimal strategies which are affine in some variables considered, we relate our approach to the parametric portfolio approach proposed in Brandt et al. (2009), where strategies considered are typically affine in some characteristics. Note that the parametric portfolio policy of Brandt et al. (2009) reduces the dimensionality of the optimization problem by modeling a small number of drivers of the portfolio weights directly.[8] Often the number of the drivers $\tilde{\mathbf{x}}_{it}$ is very low (e.g., 3 in our empirical setting below), and only investments into risky assets are considered ($w_{ft} = 0$). Typically, $\tilde{\mathbf{x}}_{it}$ is a vector of standardized variables. That is, we consider exogenous or predetermined variables $\chi_{it}$ contained in $\mathbf{x}_{it}$. For the variables $\chi_{it} \in \mathbf{R}^{k_\chi}$, we subtract the vector of sample means and multiply by the inverse of the diagonal matrix containing the sample standard deviations on the main diagonal, which results in $\tilde{\mathbf{x}}_{it}$.

$\boldsymbol{\theta} = (\theta_1, \dots, \theta_{k_\chi})^\top$ is a $k_\chi$-dimensional parameter vector in the parameter space $\Theta \subset \mathbb{R}^{k_\chi}$; if not otherwise stated $\Theta = \mathbb{R}^{k_\chi}$. $\boldsymbol{\theta}$ is assumed to be constant over time and is chosen such that expression (3) is maximized.

Following Brandt et al. (2009), we focus on the affine parametric portfolio policy

$$w_{it} = \bar{w}_{it} + \frac{1}{N}\boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it} \,, \text{ for all } i = 1, \dots, N \,. \tag{4}$$

In all applications, we work with $\bar{w}_{it} = 1/N$, where $\bar{\mathbf{w}}_t := \left(\bar{w}_{1t}, \dots, \bar{w}_{Nt}\right)^\top$. Some further results on this form of parametric portfolio policies are provided in Supplementary Material S.5.

The vector $\tilde{\mathbf{x}}_{it}$ denoting the parametric portfolio policies does not need to be identical to the observed characteristics driving expected returns. To simplify notation and to provide a fair comparison between parametric strategies and some (approximately) optimal strategies obtained later, in our empirical analysis we set $\mathbf{x}_{it} = \chi_{it}$ (hence also $k = k_\chi$), where the standardized characteristics $\tilde{\mathbf{x}}_{it}$ are assumed to be stationary. Finally, in order to compare the optimal strategies derived from our model with the affine parametric policy (4) we define [9]:

---

[6] To obtain market equilibrium in this less restrictive model setup, we have to impose stronger assumptions on $\mathbf{y}_{it}$ to proof the existence of market equilibrium; see footnote 14.

[7] This extension would be straightforward in the theoretical parts of this article, but the econometric analysis is much more involved.

[8] This approach is in contrast with the standard Markowitz (1952) approach, where optimal portfolio weights typically depend on a large number of first and second moments of the return distribution. For the estimation of the covariance matrix and related problems necessary to empirically implement a Markowitz (1952)-type approach, see, e.g., Ledoit and Wolf (2004, 2017, 2024).

[9] Our definition is different from Koijen and Yogo (2019)[see equation (10) there] where $\ln\left(w_{it}/w_{ft}\right)$ [in our notation] are affine in the firm's characteristics $\mathbf{y}_{it}$, while in our case the strategy is allowed to depend in all $\mathbf{y}_{it}$, $i = 1 \dots, N$. Since a risk-free asset need not be traded, we proceed with the definition provided in (5). To reduce the notational burden, we simply write $\mathbf{w}_t^*$ instead of $\mathbf{w}_t^*\left(\tilde{\mathbf{y}}_t\right)$, etc.

**Definition 1** (i) An investment rule is affine characteristics based (ACB in the following) if the investment weights $w_{it}$ for the risk assets—or the amounts invested $\phi_{it}$ into the risky assets—are affine in $\mathbf{y}_t$ defined in (1) and (2), $i = 1, \ldots, N$. That is, for the investments into the risk assets we have

$$w_{it} = \pi_{it} + \mathbf{\Pi}_{it}\mathbf{y}_t, \quad \text{for all } i = 1, \ldots, N . \tag{5}$$

(ii) Consider an observed and exogenous or predetermined subvector of $\mathbf{y}_{it}$, denoted $\chi_{it}$ and let $\tilde{\mathbf{x}}_{it}$ abbreviate the standardized $\chi_{it}$. A portfolio strategy is called affine characteristics-based in observed variables (ACBOV) if the investment weights for the risk assets, $w_{it}^{\sharp}$, or the amounts invested into the risky assets $\phi_{it}$, are affine in $\tilde{\mathbf{x}}_{it}$.

(iii) We call a strategy implementable by an affine parametric policy if $\bar{w}_{it} = \pi_{it}$ and $\mathbf{\Pi}_{it}\mathbf{x}_t = \frac{1}{N}\theta^\top \tilde{\mathbf{x}}_{it}$.

By using Definition 1, an ACBOV strategy is also ACB, while ACBOV demands for $w_{it}$ or $\phi_{it}$ to depend only on (a subvector of) $\mathbf{y}_{it}$ contained in $\mathbf{y}_t$. ACB or ACBOV strategies need not be parametric in a very narrow sense, since even if the investment weights are affine in $\mathbf{y}_t$ or $\tilde{\mathbf{x}}_{it}$, a matrix like $\mathbf{\Pi}_{it}$ need not follow from solving an optimization problem in some parameter vector $\theta$. However, if $\mathbf{\Pi}_{it}\mathbf{x}_t$ is equal to $\frac{1}{N}\theta^\top \tilde{\mathbf{x}}_{it}$, then the optimal strategy can be *implemented by using a parametric policy*. In more detail, $\mathbf{\Pi}_{it}\mathbf{y}_t = \frac{1}{N}\theta^\top \tilde{\mathbf{x}}_{it}$ demands for $\mathbf{B}\chi_{it}$, where $\mathbf{B}$ is a submatrix of $\mathbf{\Pi}_{it}$ (equal for all $i$ and $t$). Then, by using the population mean and the population standard deviation of $\chi_{it}$, $\mathbf{B}\chi_{it}$ can be expressed by means of $\frac{1}{N}\theta^\top \tilde{\mathbf{x}}_{it}$ plus a constant term.

The question whether an optimal strategy can be implemented by the reduction strategy (4) is discussed Sections 4 and 5. For an investor, the question arises whether an optimal strategy after performing parameter estimation is really better than a simple reduction strategy or the "$1/N$-rule." This question will be investigated in our empirical Section 7.

**Remark 1** Ferson and Siegel (2001) investigate unconditional minimum variance portfolios. In their work, the corresponding moments are obtained by conditioning on random variables, similar to our variables $\mathbf{x}_t$. In addition, Hjalmarsson and Manchev (2012) show that if the return generating process is linear in the lagged, de-meaned predictor variables ($\mathbf{x}_{it}$ in our notation), the optimal parametric portfolio weighting policy (i.e., the $\theta$s) can be derived analytically but only for the case of mean–variance preferences. [Compare also to discussion about in optimal $\theta$ in the Supplementary Material S.5.4].

Next we develop asset demand for the cases of constant absolute risk aversion (Section 4) and then for constant relative risk aversion (Section 5).

## 4 Parametric portfolio policies with constant absolute risk aversion

In this section, we explore constant absolute risk aversion by applying a Bernoulli utility function $u(x) = -\exp(-\rho x)$, $x \in \mathbb{R}$, where the parameter $\rho > 0$ expresses constant relative risk aversion, defined by $\frac{u''(x)}{u'(x)} = \rho$. The domain $\mathbb{D}$ of this function is the real axis.

For the CARA case, it is easier to work with the amounts invested $\boldsymbol{\phi}_t$. The weights of investments into the $N$ risky assets follow from $\boldsymbol{\phi}_t$ and $e_t$, that is, $\mathbf{w}_t = \frac{1}{\mathbf{1}_N^\top \boldsymbol{\phi}_t} \boldsymbol{\phi}_t$.

The portfolio vector of risky investments is $\boldsymbol{\phi}_t = \left(\phi_{1t}, \ldots, \phi_{Nt}\right)^\top \in \mathbb{R}^N$, where $\phi_{it}$ is the money amount invested into risky asset $i$ at period $t$. The amount invested in the risk-free asset is $\phi_{ft} = e_t - \boldsymbol{\phi}_t^\top \mathbf{1}_N$ if a risk-free asset is traded, and $\phi_{ft} = 0$, $\forall t$, otherwise. Hence, the value of the portfolio in period $t+1$ is a random variable and given by $E_{t+1} = e_t \left(w_{ft} R_{ft} + \sum_{i=1}^N w_{it} R_{it+1}\right) = \phi_{ft} R_{ft+1} + \sum_{i=1}^N \phi_{it} R_{it+1} = \sum_{i=1}^N \phi_{it} R_{it+1}$ $+ \left(e_t - \sum_{i=1}^N \phi_{it}\right) R_{ft+1} = \boldsymbol{\phi}_t^\top \mathbf{R}_{t+1} + \left(e_t - \sum_{i=1}^N \boldsymbol{\phi}_t^\top \mathbf{1}_N\right) R_{ft}$, where $\mathbf{R}_{t+1}$ denotes the vector of risky returns and $\boldsymbol{\phi}_t \in \Theta = \mathbb{R}^N$. In this section, we impose:

**Assumption 2** $\mathbf{R}_{t+1}$ conditional on $\mathbf{y}_t$ (or the observed variables $\mathbf{x}_{it}$, $i = 1, \ldots, N$) is multivariate normal with mean parameter $\mathbb{E}_t\left(\mathbf{R}_{t+1}\right)$ and conditional covariance $\boldsymbol{\Sigma}_t = \mathbb{V}_t\left(\mathbf{R}_{t+1}\right)$ satisfies $0 < \mathbb{V}_t\left(\mathbf{R}_{t+1}\right) < \infty$ [i.e., the conditional covariance matrix is finite and regular].

We first analyze optimal investment strategies and then apply a shrinkage procedure.

### 4.1 Optimal strategy

Using the assumption of normally distributed innovations in the absence of transactions costs, we can derive conditional expected utility

$$\mathbb{E}_t(-\exp(-\rho E_{t+1})) = -\exp\left[-\rho e_t - \rho \boldsymbol{\phi}_t^\top \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) + \frac{\rho^2}{2} \boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1}) \boldsymbol{\phi}_t\right]. \tag{6}$$

Maximizing (6) yields the vector of optimal amounts invested into the risky assets

$$\boldsymbol{\phi}_t^*(\mathbf{x}_t) = \left(\underbrace{\frac{1}{\rho}\left(\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\right)^{-1}}_{\mathbf{B}_t}\right)^{-1} \underbrace{\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft}\mathbf{1}_N\right)}_{\mathbf{b}_t}. \tag{7}$$

The remaining wealth $\phi_{ft} = e_t - \mathbf{1}_N^\top \boldsymbol{\phi}_t^* \in \mathbb{R}$ is invested into the risk-free asset. In case when a risk-free asset is not available, we can establish the following result:[10]

$$\boldsymbol{\phi}_t^+(\mathbf{x}_t) = \frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - \frac{\rho\left(\frac{1}{\rho}\mathbf{1}_N^\top\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbb{E}_t(\mathbf{R}_{t+1}) - e_t\right)}{\mathbf{1}_N^\top\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\mathbf{1}_N\right). \quad (8)$$

From (7) and (8) we conclude:

**Observation 1** (i) The optimal strategies $\boldsymbol{\phi}_t^*$ and $\boldsymbol{\phi}_t^+$ exist and are unique. The optimal $\boldsymbol{\phi}_t^*$ does not depend on the initial wealth $e_t$. The total amount invested into risky assets $\boldsymbol{\phi}_t^{*\top}\mathbf{1}_N$ depends on $\mathbf{y}_t \in \mathbb{R}^{Nk_y}$ (or $\mathbf{x}_t \in \mathbb{R}^{Nk}$). The amount invested into the risk-free asset follows from $\phi_{ft} = e_t - \boldsymbol{\phi}_t^{*\top}\mathbf{1}_N$. Given that $\boldsymbol{\phi}_t^{*\top}\mathbf{1}_N \geq e_t$, for the problem without risk-free asset we get $\phi_{ft} = 0$ and $\boldsymbol{\phi}_t^{+\top}\mathbf{1}_N = e_t$.

(ii) Suppose that Assumption 1 holds, then $\boldsymbol{\phi}_{it}^*$ is affine in $\mathbf{y}_t$ and the strategy is ACB. If the conditional expectation of the returns remains affine also for a subvector of $\mathbf{y}_{it}$, for example the observed characteristics $\mathbf{x}_{it}$, then this strategy is also ACBOV.

(iii) Since the weights depend on $\mathbf{y}_{it}$, $i = 1, \ldots, N$, the investment weights are—in general—not of the structure described in (4).

Equations (1),(7) and the assumption of an unrestricted covariance matrix $\boldsymbol{\Sigma}$ show that the optimal investments into a risky asset $i$ depend on $\mathbf{y}_{1t}, \ldots, \mathbf{y}_{Nt}$. We get dependence on $\mathbf{y}_{it}$ only in the case of a diagonal covariance matrix. Hence, our assumption on the covariance matrix implies that in general the optimal investment strategy cannot by supported by the reduction strategy (4).

This is an important difference to Koijen and Yogo (2019), where a factor structure for the covariance matrix is assumed. Then, the Woodbury matrix identity and some algebra (see Koijen and Yogo 2019, equation(A.6)) allow to derive an optimal strategy which is affine in $\mathbf{y}_{it}$. Based on some empirical literature (see, e.g., Barigozzi and Brownlees 2019, who demonstrate that after adjusting for factors some correlation (some network effects) are present) we decided to work with an unrestricted covariance matrix. The result discussed in this paragraph also remains valid for the CARA case with shrinkage as well as for the CRRA approximation provided in Section 5.[11]

## 4.2 CARA utility and shrinkage

Let us now analyze the general case with or without short selling and apply a shrinkage procedure. By maximizing expected utility (6), we get:

---

[10] For details, see Supplementary Material S.5.1.

[11] Supplementary Material S.1.1 introduces trading cost. Supplementary Material S.5.2 shows that that in this case optimal strategies are path dependent and therefore neither ACB nor ACBOV.

**Observation 2** (i) In sample, good performance in terms of the certainty equivalent (see equation (19) presented later) and the Sharpe ratio for our empirical data set (see Section A) is observed. (ii) The out-of-sample performance is quite poor. The reason for this is that especially without short selling constraints the optimal strategy $\boldsymbol{\phi}_t^* = \mathbf{B}_t^{-1}\mathbf{b}_t$ is very risky (see also Table 8 in Supplementary Material S.3.2).

Figure 1 plots realized returns $R_{pt+1}$ for $\boldsymbol{\phi}_t^*$, the parametric policy $\boldsymbol{\phi}_t^{\sharp}$ and the $1/N$-strategy $\boldsymbol{\phi}_t^{1/N}$ [since $e_t = 1$, $\boldsymbol{\phi}_t^{\sharp} = \mathbf{w}_t^{\sharp}$ and $\boldsymbol{\phi}_t^{1/N} = \mathbf{w}_t^{1/N}$; $w_{ft}^{\sharp} = w_{ft}^{1/N} = 0$]. Note that the vertical axes have different scales and the variation in the returns becomes very large with $\boldsymbol{\phi}_t^*$. In our empirical data set, this results in poor out-of-sample performance. To circumvent this problem, we augment the optimization problem by a shrinkage device. In terms of econometrics, we consider a ridge regression problem[12],

while in terms of finance we add an object close to a quadratic cost term. Although also a cost function as described in Section S.1.1 can be used as some kind of punishment function, we want to exclude path-dependence [as discussed in Supplementary Material S.5.2] and shrink $\boldsymbol{\phi}_t$ or the weights $\mathbf{w}_t$ toward some specific values $\check{\boldsymbol{\phi}}_t$ or weights $\check{\mathbf{w}}_t$, respectively. This is close in spirit to approaches like Black and Litterman (1992) that anchor an optimization in some pre-specified portfolio. This portfolio could be an $1/N$-portfolio as in our empirical example. But it could also be the market portfolio motivated by CAPM equilibrium considerations or a long-term strategic allocation specific to an (institutional) investor.

Hence, we consider a positive definite $N \times N$ matrix $\mathbf{C}_t$ and the punishment term $-\frac{1}{2}(\boldsymbol{\phi}_t - \check{\boldsymbol{\phi}}_t)^\top \mathbf{C}_t (\boldsymbol{\phi}_t - \check{\boldsymbol{\phi}}_t)$. With $\check{\boldsymbol{\phi}}_t = \check{\mathbf{w}}_t = \mathbf{0}_{N\times1}$, we shrink to zero, while with $\check{\boldsymbol{\phi}}_t = e_t \frac{1}{N}\mathbf{1}_{N\times1}$ shrink toward the $1/N$-portfolio.

By using transformed expected utility (6), the (possible) short selling constraints $\phi_{\text{it}} \geq 0$, and the shrinkage device, we get $b_{0t} := \rho e_t$, $\mathbf{b}_t := \rho(\mathbb{E}_t(\mathbf{R}_{t+1}) - \mathbf{1}_N R_{ft+1})^\top$, $\mathbf{B}_t := \frac{\rho^2}{2}\mathbb{V}_t(\mathbf{R}_{t+1} - \mathbf{1}_N R_{ft+1}) = \frac{\rho^2}{2}\mathbb{V}_t(\mathbf{R}_{t+1})$, and the Lagrangian

$$L(\boldsymbol{\phi}_t, \lambda_{1t}, \dots, \lambda_{nt}) = b_{0t} + \mathbf{b}_t\boldsymbol{\phi}_t - \frac{1}{2}\boldsymbol{\phi}_t^\top \mathbf{B}_t\boldsymbol{\phi}_t + \sum_{i=1}^{n} \lambda_{\text{it}}\phi_{\text{it}} - \frac{1}{2}(\boldsymbol{\phi}_t - \check{\boldsymbol{\phi}}_t)^\top \mathbf{C}_t(\boldsymbol{\phi}_t - \check{\boldsymbol{\phi}}_t).$$

Let $\lambda_t := (\lambda_{1t}, \dots, \lambda_{Nt})^\top$. Taking first partial derivatives with respect to $\boldsymbol{\phi}_t$ and $\lambda_t$, we get the Kuhn–Tucker conditions

---

[12] Ridge regression was proposed to consider multi-colinearity in regression problems and has become more prominent as a shrinkage device in more recent machine learning literature (see, e.g., Hastie et al. 2009, Chapter 3.4) or Nagel (2021) for applications in asset pricing models. We applied a quadratic punishment term because of its trace-ability. "Linear punishment" can be included by working with $\ell_1$-distance. This corresponds to the LASSO, where optimal weights can be obtained by applying least angle regression (see Hastie et al. 2009). A mixture of linear and quadratic punishment terms results in the elastic net, see Zou and Hastie (2005) and Chapter 3.4 in Hastie et al. (2009). To obtain closed-form solutions, we proceed with the ridge regression.

$$\frac{\partial L(\boldsymbol{\phi}_t, \boldsymbol{\lambda}_t)}{\partial \boldsymbol{\phi}_t^\top} = \mathbf{b}_t - \mathbf{B}_t \boldsymbol{\phi}_t + \boldsymbol{\lambda}_t - \mathbf{C}_t(\boldsymbol{\phi}_t - \check{\boldsymbol{\phi}}_t) = \mathbf{0}_N , \tag{9}$$

$\frac{\partial L(\boldsymbol{\phi}_t, \boldsymbol{\lambda}_t)}{\partial \lambda_{it}} = \phi_{it} = 0$ , $i = 1, \ldots, N$ , and the complementary slackness conditions

$$0 = \lambda_{it} \frac{\partial L(\boldsymbol{\phi}_t, \boldsymbol{\lambda}_t)}{\partial \lambda_{it}} = \lambda_{it}\phi_{it} , \ i = 1, \ldots, N .$$
$$\tag{10}$$

$\phi_{ft} = e_t - \sum_{i=1}^{N} \phi_{it}$ in the case a risk-free asset is traded. The second-order conditions are satisfied by the quadratic structure of the optimization problem (see, e.g., Simon and Blume 1994, Chapter 19.3). If no short selling constraints are binding or if we consider an optimization problem without short selling constraints, we obtain $\boldsymbol{\lambda}_t = \mathbf{0}_N$ and

$$\boldsymbol{\phi}_t^* = (\mathbf{B}_t + \mathbf{C}_t)^{-1} (\mathbf{b}_t + \mathbf{C}_t \check{\boldsymbol{\phi}}_t) . \tag{11}$$

Let $\mathbf{C}_t$ be equal to $c_p \mathbf{I}_N$, then (11) yields

$$\boldsymbol{\phi}_t^\flat := \frac{1}{\rho} \left( \mathbb{V}_t \left( (R_{t+1} - R_{ft+1}\mathbf{1}_n)(R_{t+1} - R_{ft+1}\mathbf{1}_n)^\top \right) + c_p I_n \right)^{-1} \\ \left( \mathbb{E}_t (R_{t+1} - R_{ft+1}\mathbf{1}_n) + \rho c_p \check{\boldsymbol{\phi}}_t \right) . \tag{12}$$

Observe that the optimal investments $\boldsymbol{\phi}_t^\flat \in \mathbb{R}^N$ do not depend on the wealth level $e_t$, $\phi_{ft}^\flat = e_t - \sum_{i=1}^{N} \phi_{it}^\flat$. For $c_p = 0$, we arrive at an optimization problem without shrinkage (where $\boldsymbol{\phi}_t^\flat = \boldsymbol{\phi}_t^*$), while the larger $c_p$ the more we shrink toward $\check{\boldsymbol{\phi}}_t = e_t \check{\mathbf{w}}_t$. To see this, for large $c_p$, the terms multiplied by $c_p$ become the dominating terms. Hence, for large $c_p$, $\boldsymbol{\phi}_t^\flat \approx \frac{1}{\rho}(c_p \mathbf{I}_n)^{-1} (\rho c_p \check{\boldsymbol{\phi}}_t) = \check{\boldsymbol{\phi}}_t$. Summing up, we get

**Proposition 1** (Asset Demand with CARA Preferences). *Suppose that Assumptions 1 and 2 hold. Consider an investor with CARA preferences and $c_p \geq 0$. Then, if no short selling constraints are present or if the short selling constraints are not binding, the optimal shrinkage strategy provided in* (12) *is ACB.*

*If the conditional expectation of the returns remains affine also for a subvector of $\mathbf{y}_{it}$, for example, the observed characteristics $\mathbf{x}_{it}$, then the optimal strategy is ACBOB.*

Panel (b) of Figure 1 plots the realized returns when applying $\boldsymbol{\phi}_t^\flat$ with $c_p = 0.2$ and shrinkage to the $1/N$-portfolio; since $e_t = 1$ we get $\check{\boldsymbol{\phi}}_t = \frac{1}{N}e_t \mathbf{1}_N = \frac{1}{N}\mathbf{1}_N$. When looking at the scale of the ordinate, we observe that the variation in the returns

$R_{pt}$ decreases a lot. In the case of binding short selling constraints, the system of inequalities (9) can be transformed to a linear programming problem. However, we observed that due to a high number of assets the optimal weights under short selling constraints can hardly be obtained by applying standard linear programming methods.

Hence, we applied numerical tools to obtain the optimal investments $\phi_{it}^{b,\geq 0}$ described by (9). Here the MATLAB function `fminsearch` is used, where we start the optimization routine from $\max\{0, \phi_{it}^b\}$, $i = 1, \ldots, N$.

## 5 Constant relative risk aversion

Let us now focus on constant relative risk aversion (CRRA). First, this section demonstrates potential pitfalls arising from parametric portfolio policies (that is, applying (4)) and a Bernoulli utility function defined in $\mathbb{R}_{>0}$. For CRRA, the Bernoulli utility function is $v(x) := \frac{x^{1-\gamma}}{1-\gamma}$ for $\gamma > 0$, $\gamma \neq 1$ and $\ln x$ for $\gamma = 1$. The domain $\mathbb{D}$ of $v(x)$ is the positive half-line $\mathbb{R}_{>0}$. Given Assumption 1 and the second-order condition [see (S-4) in the Supplementary Material], expected utility is strictly concave in $\theta$. However, for CRRA preferences in a simple binary model, examples can be constructed where the portfolio returns $R_{pt}$ do not remain in the domain $\mathbb{D} = \mathbb{R}_{>0}$ or where for a concave utility function, the first derivative always stays positive (or negative), such that only a supremum exists. Hence, no optimal $\theta \in \mathbb{R}^k$ exists in these cases, see equation (S-4) and Gehrig et al. (2018)]. Therefore, we obtain

**Observation 3** For an investor with CRRA preferences, an optimal $\theta \in \mathbb{R}^k$ solving the parametric portfolio optimization problem (S-3) need not exist.

In the next steps, we investigate whether Observation 3 is also relevant for real-world data. To do this, let us now apply the parametric portfolio policy approach in its original version of Brandt et al. (2009) to US stocks that are particularly relevant for institutional investors, namely S&P 500 stocks; [see Section 7.1 and Appendix A]. Our observations cover the time span from 04/1979 to 12/2013, which amounts to $T = 415$ and $N = 100$.

Consider for example the strategy defined in (4). Since relative risk aversion is only defined on the domain of positive gross returns, we need to check the underlying data and potentially develop a strategy of how to deal with negative gross returns.

In order to analyze whether negative portfolio returns are observed in the underlying empirical data, we pick some $\theta \in \mathbb{R}^3$ and check whether $R_{pt+1}$ becomes negative. And indeed, it turns out that in all the cases considered we observe negative $R_{pt+1}$ for large $\theta$ (in absolute terms), one large coordinate of $\theta$ turned out to be sufficient for negative gross returns.

As demonstrated and discussed in more detail in Supplementary Material S.4, we extend the domain to the real line by applying the utility function

$$v_\flat\left(e_t R_{pt+1}\right) := \begin{cases} v\left(e_t R_{pt+1}\right) & , \text{ for } R_{pt+1} \geq \psi_R, \\ \left(v\left(e_t \psi_R\right) - v'\left(e_t \psi_R\right)\right) + v'\left(e_t \psi_R\right) e_t R_{pt+1} & , \text{ for } R_{pt+1} < \psi_R, \end{cases}$$

$$(13)$$

where $\psi_R > 0$. With (13), we apply $v\left(e_t R_{pt+1}\right)$ for all $R_{pt+1} \geq \psi_R$. At $R_{pt+1} = \psi_R$, we get $v_\flat\left(e_t \psi_R\right) = v\left(e_t \psi_R\right) = \left(v\left(e_t \psi_R\right) - v'\left(e_t \psi_R\right)\right) + v'\left(e_t \psi_R\right) e_t \psi_R$. Observe that $v_\flat'\left(e_t \psi_R\right) = v'\left(e_t \psi_R\right)$ is equal to the slope of the line described by $\left(e_t v\left(\psi_R\right) - 1 \cdot v'\left(e_t \psi_R\right)\right) + 1 \cdot v'\left(e_t \psi_R\right) R_{pt+1}$.

Using these insights, we consider the optimization problem (3), where preferences are described by the *approximate CRRA utility function* $v_\flat$. We assume that a risk-free asset and $N$ risky assets are traded; the portfolio weights are $\mathbf{w}_t = \boldsymbol{\phi}_t / e_t$ for the investments in the risky securities and $w_{ft} = \phi_{ft} / e_t$ for the risk-free asset. Hence, $n = N + 1$.

By a Taylor series approximation of expected utility at $w_{ft} = 1$ and $\mathbf{w}_t = \left(w_{1t}, \ldots, w_{nt}\right)^\top = \mathbf{0}_N$, we obtain

$$\mathbb{E}_t\left(v_\flat(E_{t+1})\right) = \mathbb{E}_t\left(v_\flat(e_t R_{pt+1})\right) = \mathbb{E}_t\left(v_\flat(e_t(R_{ft+1} + \mathbf{w}_t(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N)))\right)$$

$$\approx \underbrace{v_\flat\left(e_{t+1} R_{ft+1}\right)}_{=:\alpha_{0t}} + \underbrace{e_t v_\flat'(e_t R_{ft+1}) \mathbb{E}_t\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)}_{=:\alpha_t} \mathbf{w}_t$$

$$- \frac{1}{2}\mathbf{w}_t^\top \underbrace{\left(-v_\flat''\left(e_{t+1} R_{ft+1}\right) e_t^2 \mathbb{E}_t\left(\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)^\top\right)\right)}_{:=\mathcal{A}_t} \mathbf{w}_t .$$

$$(14)$$

In the following optimization problem, we also allow for short selling constraints, more specifically $w_{it} \geq 0$. (see also Koijen and Yogo 2019, for a model with log-utility).

In our empirical data (see Sect. A for more details), especially, the out-of-sample performance of the approximately optimal strategy $\mathbf{w}_t = \mathcal{A}_t^{-1}\boldsymbol{\alpha}_t$ is very poor and very risky. Hence, similar to Sect. 4 we proceed with a shrinkage device. We consider a positive definite $N \times N$ matrix $\mathbf{C}_t$ and the term $-\frac{1}{2}\left(\mathbf{w}_t - \check{\mathbf{w}}_t\right)^\top \mathbf{C}_t\left(\mathbf{w}_t - \check{\mathbf{w}}_t\right)$, where with $\check{\mathbf{w}}_t = \mathbf{0}_{N \times 1}$ we shrink to zero, while with $\check{\mathbf{w}}_t = \frac{1}{N}\mathbf{1}_{N \times 1}$ shrink toward the $1/N$ portfolio [13].

By using the expected utility approximation (14), the short selling constraints and the shrinkage device, we get the Lagrangian

$$L(\mathbf{w}_t, \lambda_{1t}, \ldots, \lambda_{Nt}) = \alpha_0 + \boldsymbol{\alpha}_t \mathbf{w}_t - \frac{1}{2}\mathbf{w}_t^\top \mathcal{A}_t \mathbf{w}_t + \sum_{i=1}^N \lambda_{it} w_{it} - \frac{1}{2}\left(\mathbf{w}_t - \check{\mathbf{w}}_t\right)^\top \mathbf{C}_t\left(\mathbf{w}_t - \check{\mathbf{w}}_t\right) .$$

---

[13] Stabilizing conditions on the weights on $w_{it}$, $i = 1, \ldots, N$, can be included. That is, $\underline{w} \leq \sum_{i=1}^N w_{it} \leq \bar{w}$. This results on two further inequality constraints which can be included in a straightforward way. This is also implemented in our MATLAB code. By using these constraints, only the out-of-sample performance remains poor [without shrinkage].

Taking first partial derivatives with respect to $\mathbf{w}_t$ and $\lambda_t$, we get the Kuhn–Tucker conditions

$$\frac{\partial L(\mathbf{w}_t, \lambda_t)}{\partial \mathbf{w}_t^\top} = \boldsymbol{\alpha}_t - \mathcal{A}_t \mathbf{w}_t + \lambda_t - \mathbf{C}_t\left(\mathbf{w}_t - \check{\mathbf{w}}_t\right) = \mathbf{0}_N \,, \tag{15}$$

$$\frac{\partial L(\mathbf{w}_t, \lambda_t)}{\partial \lambda_{it}} = w_{it} = 0 \,, \ i = 1, \ldots, N \,, \ \text{and the complementary slackness conditions}$$

$$0 = \lambda_{it} \frac{\partial L(\mathbf{w}_t, \lambda_t)}{\partial \lambda_{it}} = \lambda_{it} w_{it} \,, \ i = 1, \ldots, N \,. \tag{16}$$

The second-order conditions are satisfied by the quadratic structure of the optimization problem. If no short selling constraints are binding or if we consider an optimization problem without short selling constraints, we obtain $\lambda_t = \mathbf{0}_N$ and

$$\begin{aligned}
\mathbf{w}_t &= \left(\mathcal{A}_t + \mathbf{C}_t\right)^{-1}\left(\boldsymbol{\alpha}_t + \mathbf{C}_t\check{\mathbf{w}}_t\right) \\
&= \left(\mathcal{A}_t + \mathbf{C}_t\right)^{-1}\left(\boldsymbol{\alpha}_t + \frac{e_t R_{ft+1} v_\flat'(e_t R_{ft+1})}{e_t R_{ft+1} v_\flat'(e_t R_{ft+1})} \mathbf{C}_t\check{\mathbf{w}}_t\right).
\end{aligned} \tag{17}$$

Let $M_{APR}(e_{t+1}R_{ft+1})$ denote the relative Arrow–Pratt measure evaluated at $e_{t+1}R_{ft+1}$. Since $R_{ft+1} \geq 1$ and usually close to one, we get $\frac{R_{ft+1}}{M_{APR}(e_{t+1}R_{ft+1})} \approx \frac{1}{M_{APR}(e_t R_{ft+1})}$. In realistic scenarios, $\Psi_R$ can be chosen such that $e_{t+1}R_{ft+1} > \Psi_R > 0$. In this case, we Taylor expand at the classical CRRA branch of the Bernoulli utility function $v_\flat$, that is, $\frac{x^{1-\gamma}}{1-\gamma}$.

In the following, let $\mathbf{C}_t$ be a diagonal matrix such that $\mathbf{C}_t = \left(-v_\flat''\left(e_{t+1}R_{ft+1}\right)e_t^2\right)c_p\mathbf{I}_N$, where $\mathbf{I}_N$ denotes the $N$-dimensional identity matrix and $c_p \geq 0$. Recall that $v_\flat'' \leq 0$ and $v_\flat''(x) < 0$ for $x > \Psi$. Then, the approximation $\frac{R_{ft+1}}{M_{APR}(e_{t+1}R_{ft+1})} \approx \frac{1}{M_{APR}(e_{t+1}R_{ft+1})}$ and (17) result in

$$\begin{aligned}
&\frac{1}{\gamma}\left(\mathbb{E}_t\left(\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)^\top\right) + c_p\mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right) + \frac{1}{e_t R_{ft+1} v_\flat'(e_t R_{ft+1})}\mathbf{C}_t\check{\mathbf{w}}_t\right) \\
&= \frac{1}{\gamma}\left(\mathbb{E}_t\left(\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)^\top\right) + c_p\mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right) + \frac{\left(-v_\flat''\left(e_{t+1}R_{ft+1}\right)e_t^2\right)}{e_t R_{ft+1} v_\flat'(e_t R_{ft+1})}c_p\mathbf{I}_N\check{\mathbf{w}}_t\right) \\
&\approx \frac{1}{\gamma}\underbrace{\left(\mathbb{E}_t\left(\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)^\top\right) + c_p\mathbf{I}_N\right)^{-1}}_{\mathcal{B}_t}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right) + \gamma c_p\check{\mathbf{w}}_t\right) =: \mathbf{w}_t^\flat.
\end{aligned} \tag{18}$$

Hence, we get

**Proposition 2** (Asset Demand with CRRA-Preferences). (*i*) *Suppose that Assumption* 1 *holds,* $c_p \geq 0$ *and either* $\boldsymbol{\Sigma}_t$ *is positive definite or* $c_p > 0$. *Consider an investor maximizing expected utility with Bernoulli utility function* $v_\flat(\cdot)$. *Then, if no short selling constraints are present or if the short selling constraints are not binding, the optimal shrinkage strategy* (18) *is ACB.*

(*ii*) *If the conditional expectation of the returns remains affine also for a subvector of* $\mathbf{y}_{it}$, *for example, the observed characteristics* $\mathbf{x}_{it}$, *then the optimal strategy is also ACBOV.*

(*iii*) *If the term* $\mathcal{B}_t$ *is diagonal, the weights only depend on* $\mathbf{x}_{it}$; *the optimal strategy can be supported by a reduction strategy described by* (4).

Note that a diagonal $\mathcal{B}_t$ and the equality $w_{it}^\flat = \bar{w}_{it} + \boldsymbol{\theta}^\top \mathbf{x}_{it}$ are still a strong requirements. Having derived demand functions under different preference specifications, we will next analyze the implications for equilibrium asset pricing.

## 6 Equilibrium

Koijen and Yogo (2019) prove the existence of a (unique) equilibrium price vector in the economy they consider. Recall, in Koijen and Yogo (2019) all agents are log-utility investors, where heterogeneity in the characteristics as well as in the parameters related to these characteristics can be present. Short selling constraints are given for all agents; the main results relate to cases where these constraints are not binding.
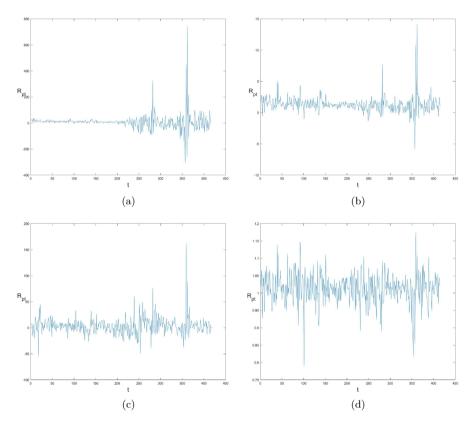
Related to this issue, we consider $J > 0$ agents either with CARA or CRRA preferences (also the risk aversion parameters can be different). Asset demand for agent $j$ is given $\boldsymbol{\phi}_t^{\flat\flat,j}$, where $\boldsymbol{\phi}_t^{\flat\flat,j} = \mathbf{w}_t^{\flat,j} E_t^j$ for CRRA preferences and $\boldsymbol{\phi}_t^{\flat\flat,j} = \boldsymbol{\phi}_t^{\flat,j}$ for CARA preferences. In contrast with Koijen and Yogo (2019), we assume that $\mathbf{y}_{it}$ only contains endogenous variables which are affine in $P_{it}$. No higher-order terms such as $(P_{it}S_{it})^v$, $v > 1$ are included. Market clearing demands for $P_{it}S_{it} = \sum_{j=1}^J \boldsymbol{\phi}_{it}^{\flat\flat,j}$, $i = 1, \ldots, N$. Since $\boldsymbol{\phi}_t^{\flat\flat,j}$ is affine in $\mathbf{P}_t$ if no short selling constraints are present or binding, we determine a unique equilibrium price vector [14].

Hence, we get

---

[14] Since market clearing conditions are affine linear in the prices, finding an equilibrium price vector corresponds to solving $N$ linear equation. By contrast, Koijen and Yogo (2019) allow for higher-order terms $(P_{it}S_{it})^v$, $v > 1$. Due to short selling constraints, lower and upper bounds for the strategies can be obtained in a straightforward way which also allows to apply Brouwer's fixed point theorem on a compact strategy space. Since we also consider the case without short selling constraints, we do not obtain a compact set which would allow us to proceed with standard fixed point arguments.

**Table 1** CARA Investment Strategies

| Abbreviation | Investment Strategy |
|---|---|
| $\phi_t^{\flat}$ | optimal strategy with naive covariance estimator with shrinkage |
| $\phi_{t,LW}^{\flat}$ | optimal strategy with Ledoit and Wolf (2004) covariance estimator with shrinkage |
| $\phi_t^{\flat,\geq 0}$ | optimal strategy with naive covariance estimator, without short-selling, with shrinkage |
| $\phi_{t,LW}^{\flat,\geq 0}$ | optimal strategy with Ledoit and Wolf (2004) covariance estimator, without short-selling, with shrinkage |
| $\phi_t^{1/N}$ | 1/N-portfolio as e.g. considered in DeMiguel et al. (2009) |
| $\phi_t^{\sharp}$ | parametric portfolio policy |
| $\phi_t^{\sharp,\geq 0}$ | parametric portfolio policy without short-selling |



**Fig. 1** Returns $R_{pt+1}$ against $t$, calibration period $t = 1, \ldots, 200$, $T = 416$, $N = 100$ and $n = 101$. $\gamma = 0.5$ and CARA utility. Subfigure (a) applies $\phi_t^*$ [that is $\phi_t^{\flat}$ with $c_p = 0$] where the Ledoit and Wolf (2004) estimator is used to obtain an estimate of the covariance matrix, That is, $\phi_t^{\flat}$ obtained in (12) where the shrinkage parameter $c_p = 0$. Scaling of vertical axis $[-400, 800]$, (b) applies $\phi_t^*$ obtained in (1) where the Ledoit and Wolf (2004) estimator is used to obtain the covariance matrix, shrinkage parameter $c_p = 0.2$. Scaling of vertical axis $[-10, 15]$, (c) parametric portfolio policy, scaling of vertical axis $[-100, 200]$, (d) 1/N-strategy, scaling of vertical axis $[0.75, 1.2]$

**Proposition 3** (Market Equilibrium). *Consider an economy with $J > 0$ investors. Each investor j is either a CARA or CRRA (in more detail, $w_t^\flat$ is applied) expected utility maximizer. Suppose that Assumption 1 holds and $c_p^j \geq 0$. Suppose that either no short selling constraints are present or no short selling constraint is binding.*

*For each CRRA utility maximizer, either $\Sigma_t$ is positive definite or $c_p > 0$. For each CARA utility maximizer Assumption 2 holds.*

*Suppose that $\mathbf{y}_{it}^j$ (or observable subvector $\mathbf{x}_{it}^j$) only contains endogenous variables which are affine functions of $P_{it}$. Then a unique equilibrium price vector exits.*

**Remark 2** Note that the optimal strategies for the CARA (see $\boldsymbol{\phi}_t^*(\mathbf{x}_t)$ in (7) and the CRRA (see $w_b^\flat$ with $c_p = 0$ in (18)) are nested in model considered in Proposition 3. Hence, we obtain equilibrium also for models without shrinkage (i.e., $c_p^j = 0$, $j = 1, \dots, J$).

Equipped with this theoretical foundations, we can now evaluate the empirical performance of the demand systems approach for both preference classes in the next section.

# 7 Empirical results

## 7.1 Comparison of strategies for the CARA case

Let us now compare investment strategies at the hand of US stock prices. Specifically, we consider the following strategies:

While the optimal strategies exploit second moments, the parametric portfolio strategies estimate optimal portfolio weights directly as a function of the characteristics without estimating variances and covariances. The $\frac{1}{N}$-strategy corresponds to a simple investment heuristic that abstracts from any information about second moments or any other characteristics. Supplementary Material S.3.2 describes how the conditional expectations $\mathbb{E}_t(R_{it+1})$ [including the characteristics $\mathbf{x}_{it}$] and variances $\mathbb{V}_t(R_{it+1})$ are estimated. In contrast with Koijen and Yogo (2019) (but following a large finance literature), we run predictive regressions to estimate $\mathbb{E}_t(R_{it+1})$.

Our sample consists of $N = 100$ S&P stocks with monthly data from April 1979 to December 2013 (for more details on the data see Appendix A). The 100 firms considered were traded continuously during this time span. We decided to work with these 100 firms to avoid further problems and effects arising from missing data (e.g., how to treat missing data in the estimation of the covariance matrix [15]).

---

[15] As pointed out by the referee, our results are expected to contain some survivor-ship bias (see, e.g., Carpenter and Lynch 1999; Carhart et al. 2015). In our analysis, we claim that all the biases for various strategies considered are approximately the same, such that comparisons of the performance measures presented in the article still make sense.

The wealth invested per period is $e_t = 1$. Since our main focus is on the risky assets and to exclude impacts arising from changes in the risk-free rate, we assume a constant risk-free rate of $R_{ft} = 1.001$.

The $k = k_\chi = 3$ macro-variables $\chi_{it}$ of the parametric policy are:

- $\chi_{it,1}$ is the natural logarithm (ln) of one plus the firm's book-to-market ratio,
- $\chi_{it,2}$ is the natural logarithm of the firm's market equity,
- $\chi_{it,3}$ is a momentum variable obtained from the compound returns from the periods $t - 13$ to $t - 2$.

As already stated in Section 3, we assume that the standardized $\chi_{it}$ provides us with a stationary process of observed characteristics $\mathbf{x}_{it}$. In particular, $\mathbf{x}_{it}$ is the subvector of $\mathbf{y}_{it}$ used to obtained the amounts invested $\boldsymbol{\phi}_t^\flat$. In addition, in the empirical implementation we work with constant, i.e., not time-varying, model parameters. This of course simplifies the econometric analysis. In addition, by this assumption we investigate whether our relatively simple shrinkage strategies can already improve over $1/N$ or parametric strategies when working with constant model parameters.

The observations from $t = 1, \dots, 200$ are used to estimate the model parameters (training sample).

For in-sample and out-of-sample comparisons, we use the observations $t = 1, \dots, 200$ and $t = 201, \dots, 415$, respectively [16].

In addition, we consider the $1/N$-portfolio as e.g. considered in DeMiguel et al. (2009); this portfolio is denoted $\boldsymbol{\phi}_t^{1/N}$. In the case short selling constraints we apply the notation $\boldsymbol{\phi}_t^{\flat,\geq 0}$ for the optimal strategy (including shrinkage) and $\boldsymbol{\phi}_t^{\sharp,\geq 0}$ for the parametric policy.

We are particularly interested in the performance of the optimal investment strategy relative to the characteristics-based portfolio choice. Hence, we do not calculate some distance measure investigating the closeness of the weights or some parameters of the parametric policy and some optimal strategy, but directly use (estimates of) expected utility for comparison. In particular, we translate utility numbers to monetary units and use the certainty equivalent. For the exponential utility function $u(x) = -e^{\rho x}$, the certainty equivalent $\mathscr{C}$ is the (smallest) value $x$ where $\mathbb{E}\big(u(E_{t+1})\big) = u(x)$. We estimate the certainty equivalent by means of

$$\widehat{\mathscr{C}} = u^{-1}\left(\widehat{\mathbb{E}}\big(u(E_{t+1})\big)\right) = u^{-1}\left(\frac{1}{T_\mathbb{J}} \sum_{t \in \mathbb{T}_\mathbb{J}} u(E_{t+1})\right), \tag{19}$$

where $\mathbb{T}_\mathbb{J}$ is the set of time points used for the evaluation of the strategy and $T_\mathbb{J}$ is the number of time points contained in this set. [17]

In addition to estimates of the certainty equivalent, we calculated (estimates of) the Sharpe ratio by the average excess returns over the sample standard deviation of

---

[16] It is insightful to compare both, simulated and empirical data; results with simulated data are provided in Supplementary Material S.3.

[17] In Tables 2 and 3, the set $\mathbb{T}_\mathbb{J} = \{1, \dots, 200\}$ (in sample) or $\mathbb{T}_\mathbb{J} = \{201, \dots, 415\}$ (out of sample) where $t = 1, \dots, 200$ is used for parameter estimation.

**Table 2** CARA Utility (In Sample): Investment strategies defined in Table 1. Empirical data. Training sample $t = 1, \ldots, 200$, Evaluation in sample; $t = 1, \ldots, 200$. Shrinkage parameter $c_p = 0.2$

| $\rho = 0.25$ | $\phi_t^{\flat}$ | $\phi_{t,LW}^{\flat}$ | $\phi_t^{\flat,\geq 0}$ | $\phi_{t,LW}^{\flat,\geq 0}$ | $\phi_t^{1/N}$ | $\phi_t^{\sharp}$ | $\phi_t^{\sharp,\geq 0}$ |
|---|---|---|---|---|---|---|---|
| $\widehat{\mathscr{C}}$ | 1.3997 | 1.4014 | 1.3137 | 1.3148 | 1.0153 | -34.2976 | 1.0153 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0380 | 0.0382 | 0.0356 | 0.0349 | 0.0023 | 2.11E+04 | 0.0023 |
| mean($E_t$) | 1.4657 | 1.4683 | 1.3728 | 1.3724 | 1.0156 | 3.0080 | 1.0156 |
| sd($E_t$) | 0.7234 | 0.7286 | 0.6926 | 0.6845 | 0.0420 | 11.5086 | 0.0420 |
| Sharpe ratio | 0.6424 | 0.6414 | 0.5368 | 0.5426 | 0.3465 | 0.1744 | 0.3466 |
| mean($w_{it} < 0$) | 0.1888 | 0.1888 | 0.0000 | 0.0000 | 0.0000 | 0.4837 | 0.0000 |
| $sd\left(w_{it} < 0\right)$ | 0.0477 | 0.0479 | 0.0000 | 0.0000 | 0.0000 | 0.0290 | 0.0000 |
| mean$\left(w_{it} < -1\right)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.4443 | 0.0000 |
| $sd\left(w_{it} < -1\right)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0262 | 0.0000 |
| $\rho = 0.5$ | $\phi_t^{\flat}$ | $\phi_{t,LW}^{\flat}$ | $\phi_t^{\flat,\geq 0}$ | $\phi_{t,LW}^{\flat,\geq 0}$ | $\phi_t^{1/N}$ | $\phi_t^{\sharp}$ | $\phi_t^{\sharp,\geq 0}$ |
| $\widehat{\mathscr{C}}$ | 1.2023 | 1.2031 | 1.1350 | 1.1410 | 1.0151 | 0.3416 | 1.0151 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0153 | 0.0153 | 0.0133 | 0.0129 | 0.0018 | 0.3325 | 0.0018 |
| mean($E_t$) | 1.2373 | 1.2386 | 1.1601 | 1.1655 | 1.0156 | 1.3093 | 1.0156 |
| sd($E_t$) | 0.3722 | 0.3748 | 0.3148 | 0.3128 | 0.0420 | 1.6562 | 0.0420 |
| Sharpe ratio | 0.6350 | 0.6339 | 0.5259 | 0.5055 | 0.3465 | 0.1861 | 0.3466 |
| mean($w_{it} < 0$) | 0.1832 | 0.1816 | 0.0000 | 0.0000 | 0.0000 | 0.4806 | 0.0000 |
| $sd\left(w_{it} < 0\right)$ | 0.0478 | 0.0479 | 0.0000 | 0.0000 | 0.0000 | 0.0291 | 0.0000 |
| mean$\left(w_{it} < -1\right)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.2514 | 0.0000 |
| $sd\left(w_{it} < -1\right)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0190 | 0.0000 |
| $\rho = 1$ | $\phi_t^{\flat}$ | $\phi_{t,LW}^{\flat}$ | $\phi_t^{\flat,\geq 0}$ | $\phi_{t,LW}^{\flat,\geq 0}$ | $\phi_t^{1/N}$ | $\phi_t^{\sharp}$ | $\phi_t^{\sharp,\geq 0}$ |
| $\widehat{\mathscr{C}}$ | 1.1035 | 1.1039 | 1.0725 | 1.0702 | 1.0147 | 1.0016 | 1.0147 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0049 | 0.0049 | 0.0041 | 0.0042 | 0.0011 | 0.0143 | 0.0011 |
| mean($E_t$) | 1.1231 | 1.1238 | 1.0860 | 1.0839 | 1.0156 | 1.1076 | 1.0156 |
| sd($E_t$) | 0.1966 | 0.1980 | 0.1643 | 0.1655 | 0.0420 | 0.4494 | 0.0420 |
| Sharpe ratio | 0.6211 | 0.6201 | 0.5013 | 0.5176 | 0.3465 | 0.2371 | 0.3466 |
| mean($w_{it} < 0$) | 0.1687 | 0.1693 | 0.0000 | 0.0000 | 0.0000 | 0.4799 | 0.0000 |
| $sd\left(w_{it} < 0\right)$ | 0.0477 | 0.0476 | 0.0000 | 0.0000 | 0.0000 | 0.0339 | 0.0000 |
| mean$\left(w_{it} < -1\right)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0325 | 0.0000 |
| $sd\left(w_{it} < -1\right)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0071 | 0.0000 |
| $\rho = 2$ | $\phi_t^{\flat}$ | $\phi_{t,LW}^{\flat}$ | $\phi_t^{\flat,\geq 0}$ | $\phi_{t,LW}^{\flat,\geq 0}$ | $\phi_t^{1/N}$ | $\phi_t^{\sharp}$ | $\phi_t^{\sharp,\geq 0}$ |
| $\widehat{\mathscr{C}}$ | 1.0539 | 1.0541 | 1.0342 | 1.0350 | 1.0138 | 1.0213 | 1.0138 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0010 | 0.0010 | 0.0008 | 0.0008 | 0.0004 | 0.0015 | 0.0004 |
| mean($E_t$) | 1.0660 | 1.0663 | 1.0419 | 1.0428 | 1.0156 | 1.0459 | 1.0156 |
| sd($E_t$) | 0.1089 | 0.1096 | 0.0870 | 0.0878 | 0.0420 | 0.1548 | 0.0420 |
| Sharpe ratio | 0.5970 | 0.5961 | 0.4767 | 0.4695 | 0.3465 | 0.2899 | 0.3465 |
| mean($w_{it} < 0$) | 0.1469 | 0.1463 | 0.0000 | 0.0000 | 0.0000 | 0.4558 | 0.0000 |
| $sd\left(w_{it} < 0\right)$ | 0.0442 | 0.0445 | 0.0000 | 0.0000 | 0.0000 | 0.0378 | 0.0000 |
| mean$\left(w_{it} < -1\right)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

**Table 2** (continued)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $sd(w_{it} < -1)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $\rho = 5$ | $\phi_t^{\flat}$ | $\phi_{t,LW}^{\flat}$ | $\phi_t^{\flat,\geq 0}$ | $\phi_{t,LW}^{\flat,\geq 0}$ | $\phi_t^{1/N}$ | $\phi_t^{\sharp}$ | $\phi_t^{\sharp,\geq 0}$ |
| $\widehat{\mathscr{C}}$ | 1.0234 | 1.0235 | 1.0149 | 1.0143 | 1.0109 | 1.0036 | 1.0109 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 2.80E-5 | 2.81E-5 | 2.16E-5 | 2.13E-5 | 2.17E-5 | 2.80E-5 | 2.17E-5 |
| $mean(E_t)$ | 1.0318 | 1.0319 | 1.0199 | 1.0191 | 1.0156 | 1.0130 | 1.0156 |
| $sd(E_t)$ | 0.0564 | 0.0567 | 0.0440 | 0.0429 | 0.0420 | 0.0630 | 0.0420 |
| Sharpe ratio | 0.5453 | 0.5444 | 0.4210 | 0.4299 | 0.3465 | 0.1903 | 0.3465 |
| $mean(w_{it} < 0)$ | 0.0973 | 0.0990 | 0.0000 | 0.0000 | 0.0000 | 0.3755 | 0.0000 |
| $sd(w_{it} < 0)$ | 0.0376 | 0.0362 | 0.0000 | 0.0000 | 0.0000 | 0.0459 | 0.0000 |
| $mean(w_{it} < -1)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $sd(w_{it} < -1)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

the excess returns (when using the corresponding evaluation sample). For the optimal strategies provided in (12) and (18) for the CARA and the CRRA case, respectively, the tuning hyper-parameter $c_p$ has to be specified. We already know e.g. from Observation 1 that for $c_p = 0$ the performance is poor and the strategy is very risky. To choose $c_p$, we used a (coarse) gird 0, 0.0.1, 0.2, 0.3, 0.4, 0.5, 1 and evaluate the performance of the shrinkage strategy. It turned out that shrinkage to $1/N$ with $c_p = 0.2$ performs quite well in our out-of-sample data set.

Tables 2 and 3 present the results for the CARA utility case for different levels of constant absolute risk aversion, spanning the ranges from 0.25, 0.5, 1,2 to 5.[18]

These tables show estimates of the certainty equivalent $\widehat{\mathscr{C}}$ and its standard deviation, the average wealth obtained, $mean(E_t)$, and its standard deviation, the Sharpe Ratio, and the proportions of weights $< 0$ and $< -1$. The average gross return in formal term $mean(R_{pt}) = \frac{mean(E_{t+1})}{e_t}$. Since $e_t = 1$, we get $mean(R_{pt}) = mean(E_t)$ and $mean(r_{pt}) = mean(E_t) - 1$. Hence, only $mean(E_t)$ is presented.

*In-sample results:* By considering the estimates of the certainty equivalents $\widehat{\mathscr{C}}$, we observe that the optimal strategies $\phi_t^{\flat}$ show the best performance. Due to short selling constraints, we obtained $\phi_t^{\flat,\geq 0} < \phi_t^{\flat}$. The difference caused by working with different estimation methods of the covariance matrix is small. The performance of the parametric policy $\phi_t^{\sharp}$ is poor for low degrees of risk aversion, but becomes closer to the performance of the optimal strategy for $\rho \geq 1$. By imposing short selling constraints $\phi_t^{\sharp,\geq 0}$ is almost equal to the results with $1/N$-portfolio (our optimization routine is started with $\bar{w}_i = 1/N$ and $\theta = 0$; differences in the numbers between $\phi_t^{\sharp,\geq 0}$ and $\phi_t^{1/N}$ are only observable when looking at further digits after the comma). Supplementary Material S.3 provides results for simulated

**Table 3** CARA Utility (Out of Sample): Investment strategies defined in Table 1. Training sample $t = 1, \ldots, 200$, evaluation out of sample; $t = 201, \ldots, 415$. shrinkage parameter $c_p = 0.2$

| $\rho = 0.25$ | $\phi_t^\flat$ | $\phi_{t,LW}^\flat$ | $\phi_t^{\flat,\geq0}$ | $\phi_{t,LW}^{\flat,\geq0}$ | $\phi_t^{1/N}$ | $\phi_t^\sharp$ | $\phi_t^{\sharp,\geq0}$ |
|---|---|---|---|---|---|---|---|
| $\widehat{\mathscr{C}}$ | 1.1418 | 1.1376 | 0.9996 | 0.9925 | 1.0107 | -28.7653 | 1.0107 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0724 | 0.0743 | 0.1278 | 0.1308 | 0.0024 | 4.61E+03 | 0.0024 |
| mean($E_t$) | 1.3547 | 1.3588 | 1.4290 | 1.4325 | 1.0109 | 4.4794 | 1.0109 |
| sd($E_t$) | 1.3359 | 1.3624 | 1.8231 | 1.8443 | 0.0445 | 20.0580 | 0.0445 |
| Sharpe ratio | 0.2648 | 0.2626 | 0.2348 | 0.2340 | 0.2231 | 0.1734 | 0.2232 |
| mean($w_{it} < 0$) | 0.3300 | 0.3284 | 0.0000 | 0.0000 | 0.0000 | 0.4534 | 0.0000 |
| sd$\left(w_{it} < 0\right)$ | 0.0474 | 0.0482 | 0.0000 | 0.0000 | 0.0000 | 0.0280 | 0.0000 |
| mean$\left(w_{it} < -1\right)$ | 0.0268 | 0.0268 | 0.0000 | 0.0000 | 0.0000 | 0.4232 | 0.0000 |
| sd$\left(w_{it} < -1\right)$ | 0.0174 | 0.0172 | 0.0000 | 0.0000 | 0.0000 | 0.0271 | 0.0000 |
| $\rho = 0.5$ | $\phi_t^\flat$ | $\phi_{t,LW}^\flat$ | $\phi_t^{\flat,\geq0}$ | $\phi_{t,LW}^{\flat,\geq0}$ | $\phi_t^{1/N}$ | $\phi_t^\sharp$ | $\phi_t^{\sharp,\geq0}$ |
| $\widehat{\mathscr{C}}$ | 1.0706 | 1.0683 | 1.0125 | 1.0332 | 1.0104 | 0.0316 | 1.0104 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0288 | 0.0295 | 0.0436 | 0.0322 | 0.0018 | 0.2771 | 0.0018 |
| mean($E_t$) | 1.1805 | 1.1825 | 1.1571 | 1.1469 | 1.0109 | 1.5106 | 1.0109 |
| sd($E_t$) | 0.6779 | 0.6911 | 0.7165 | 0.6544 | 0.0445 | 2.8799 | 0.0445 |
| Sharpe ratio | 0.2648 | 0.2627 | 0.2229 | 0.2179 | 0.2231 | 0.1770 | 0.2232 |
| mean($w_{it} < 0$) | 0.3253 | 0.3234 | 0.0000 | 0.0000 | 0.0000 | 0.4501 | 0.0000 |
| sd$\left(w_{it} < 0\right)$ | 0.0477 | 0.0478 | 0.0000 | 0.0000 | 0.0000 | 0.0276 | 0.0000 |
| mean$\left(w_{it} < -1\right)$ | 0.0029 | 0.0030 | 0.0000 | 0.0000 | 0.0000 | 0.2669 | 0.0000 |
| sd$\left(w_{it} < -1\right)$ | 0.0046 | 0.0046 | 0.0000 | 0.0000 | 0.0000 | 0.0220 | 0.0000 |
| $\rho = 1$ | $\phi_t^\flat$ | $\phi_{t,LW}^\flat$ | $\phi_t^{\flat,\geq0}$ | $\phi_{t,LW}^{\flat,\geq0}$ | $\phi_t^{1/N}$ | $\phi_t^\sharp$ | $\phi_t^{\sharp,\geq0}$ |
| $\widehat{\mathscr{C}}$ | 1.0348 | 1.0336 | 1.0165 | 1.0274 | 1.0099 | 0.9376 | 1.0099 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0091 | 0.0094 | 0.0104 | 0.0093 | 0.0011 | 0.0185 | 0.0011 |
| mean($E_t$) | 1.0934 | 1.0944 | 1.0786 | 1.0830 | 1.0109 | 1.1399 | 1.0109 |
| sd($E_t$) | 0.3489 | 0.3556 | 0.3471 | 0.3332 | 0.0445 | 0.7083 | 0.0445 |
| Sharpe ratio | 0.2649 | 0.2628 | 0.2460 | 0.2236 | 0.2231 | 0.1961 | 0.2232 |
| mean($w_{it} < 0$) | 0.3156 | 0.3133 | 0.0000 | 0.0000 | 0.0000 | 0.4412 | 0.0000 |
| sd$\left(w_{it} < 0\right)$ | 0.0479 | 0.0470 | 0.0000 | 0.0000 | 0.0000 | 0.0198 | 0.0000 |
| mean$\left(w_{it} < -1\right)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0349 | 0.0000 |
| sd$\left(w_{it} < -1\right)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0091 | 0.0000 |
| $\rho = 2$ | $\phi_t^\flat$ | $\phi_{t,LW}^\flat$ | $\phi_t^{\flat,\geq0}$ | $\phi_{t,LW}^{\flat,\geq0}$ | $\phi_t^{1/N}$ | $\phi_t^\sharp$ | $\phi_t^{\sharp,\geq0}$ |
| $\widehat{\mathscr{C}}$ | 1.0167 | 1.0160 | 1.0087 | 1.0099 | 1.0089 | 1.0027 | 1.0089 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0018 | 0.0019 | 0.0019 | 0.0019 | 0.0004 | 0.0021 | 0.0004 |
| mean($E_t$) | 1.0499 | 1.0504 | 1.0404 | 1.0411 | 1.0109 | 1.0527 | 1.0109 |
| sd($E_t$) | 0.1846 | 0.1879 | 0.1739 | 0.1732 | 0.0445 | 0.2432 | 0.0445 |
| Sharpe ratio | 0.2648 | 0.2627 | 0.2313 | 0.2267 | 0.2231 | 0.2124 | 0.2231 |
| mean($w_{it} < 0$) | 0.2963 | 0.2932 | 0.0000 | 0.0000 | 0.0000 | 0.4240 | 0.0000 |
| sd$\left(w_{it} < 0\right)$ | 0.0476 | 0.0470 | 0.0000 | 0.0000 | 0.0000 | 0.0227 | 0.0000 |
| mean$\left(w_{it} < -1\right)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

**Table 3** (continued)

| | $\phi_t^{\flat}$ | $\phi_{t,LW}^{\flat}$ | $\phi_t^{\flat,\geq 0}$ | $\phi_{t,LW}^{\flat,\geq 0}$ | $\phi_t^{1/N}$ | $\phi_t^{\sharp}$ | $\phi_t^{\sharp,\geq 0}$ |
|---|---|---|---|---|---|---|---|
| $sd(w_{it} < -1)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $\rho = 5$ | $\phi_t^{\flat}$ | $\phi_{t,LW}^{\flat}$ | $\phi_t^{\flat,\geq 0}$ | $\phi_{t,LW}^{\flat,\geq 0}$ | $\phi_t^{1/N}$ | $\phi_t^{\sharp}$ | $\phi_t^{\sharp,\geq 0}$ |
| $\widehat{\mathcal{C}}$ | 1.0049 | 1.0045 | 1.0039 | 1.0040 | 1.0057 | 0.9933 | 1.0057 |
| $\widehat{sd}\left(\widehat{\mathcal{C}}\right)$ | 4.67E-5 | 4.79E-5 | 4.17E-5 | 4.13E-5 | 2.22E-5 | 4.86E-5 | 2.22E-5 |
| $mean(E_t)$ | 1.0238 | 1.0239 | 1.0196 | 1.0188 | 1.0109 | 1.0169 | 1.0109 |
| $sd(E_t)$ | 0.0863 | 0.0877 | 0.0787 | 0.0744 | 0.0445 | 0.1070 | 0.0445 |
| Sharpe ratio | 0.2636 | 0.2616 | 0.2389 | 0.2363 | 0.2231 | 0.1483 | 0.2231 |
| $mean(w_{it} < 0)$ | 0.2418 | 0.2426 | 0.0000 | 0.0000 | 0.0000 | 0.3809 | 0.0000 |
| $sd(w_{it} < 0)$ | 0.0478 | 0.0479 | 0.0000 | 0.0000 | 0.0000 | 0.0520 | 0.0000 |
| $mean(w_{it} < -1)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $sd(w_{it} < -1)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

data, where we observe that in the case the true parameter values are known, the optimal strategies show superior performance.

*Out-of-sample results* are then presented in Table 3. In this case, our optimal shrinkage strategies still have in almost all cases a slightly higher (estimate of the) certainty equivalent. For $\rho = 5$, we observe the highest certainty equivalent for the $1/N$-strategy. For $\rho \geq 2$, the performance of the strategies considered is quite similar.

Finally, both tables present the proportion of the portfolio weights smaller than zero and the proportion of weights smaller than $-1$. Note that especially for parametric portfolio policies with $\rho < 1$, the proportion of weights smaller than minus one is very high.

As expected, the in-sample performance is better than the out-of-sample performance (compare the estimates of the certainty equivalents $\widehat{\mathcal{C}}$ in Tables 2 and 3). At least parts of this difference can be attributed to the in-sample and out-of-sample forecast of the conditions means $\mathbb{E}_t(R_{it+1})$, which are equal to the fitted values in a linear regression of the returns on the variables $\mathbf{x}_{it}$. We calculated the in-sample and the out-of-sample coefficient of determination $R^2$, where we observe that approximately 2.1% of the variation in the returns $R_{it}$ can be explained by $\mathbf{x}_{it}$ in-sample, while out of sample we get $R^2 = -3.31\%$. This has a first-order effect since $\widehat{\mathbb{E}}_t(R_{it+1})$ directly enters into the optimal shrinkage strategy (12), but also has a second-order effect via the estimation of the covariance matrix $\mathbb{V}_t$. Note that the total effect, even if returns and characteristics are jointly stationary (see Assumption 1), also contains a sampling effect.[19]

Summing up, we get

---

[19] We thank the anonymous reviewer for that remark.

**Observation 4** (i) In sample: Not surprisingly, the optimal shrinkage strategy shows the best performance, followed by the $\frac{1}{N}$-strategy and the parametric strategy. For larger $\rho$, the performances of the alternative strategies as measured by the certainty equivalent perform quite similarly.

(ii) Out of sample: The optimal shrinkage strategies show the best performance followed by the $1/N$-strategy. Only for a large degree of absolute risk aversion, the performances of the strategies considered are roughly the same across strategies.

(iii) For small values of absolute risk aversion parametric portfolio policies imply a large amount of short selling in and out of sample.

(iv) In sample, the informational content contained in the variance–covariance matrix relative to the $1/N$-portfolio, as measured by the certainty equivalent, is decreasing the in level of absolute risk aversion. It is particularly high for risk aversion below 1. It is always negative for parametric portfolio policies.

### 7.2 Comparison of strategies for the CRRA case

The approximately optimal investment weights $\mathbf{w}_t^\flat \in \mathbb{R}^N$ do not depend on the wealth level $e_t$, $w_{ft} = 1 - \sum_{i=1}^N w_{it}$. For $c_p = 0$, we arrive at an optimization problem without shrinkage, while the larger $c_p$ the more we shrink toward $\check{\mathbf{w}}_t$. To implement (18), the conditional expectations $\mathbb{E}_t\big(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\big)$ and $\mathbb{E}_t\Big(\big(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\big)\big(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\big)^\top\Big)$ can be estimated in the same way as we did it in the CARA case.

Numerical tools are used to obtain the optimal weights $\mathbf{w}_t^{\flat,\geq 0}$ in the case of short selling constraints. The certainty equivalent for $v_\flat(x)$ is obtained by replacing the Bernoulli utility function $u(x)$ by $v_\flat(x)$ in (19).

**Table 4** CRRA investment strategies

| Abbreviation | Investment strategy |
| --- | --- |
| $\mathbf{w}_t^\flat$ | approximately optimal strategy with naive covariance estimator with shrinkage |
| $\mathbf{w}_{t,LW}^\flat$ | approximately optimal strategy with Ledoit and Wolf (2004) covariance estimator with shrinkage |
| $\mathbf{w}_t^{\flat,\geq 0}$ | approximately optimal strategy with naive covariance estimator, without short-selling, with shrinkage |
| $\mathbf{w}_{t,LW}^{\flat,\geq 0}$ | approximately optimal strategy with Ledoit and Wolf (2004) covariance estimator, without short-selling, with shrinkage |
| $\mathbf{w}_t^{1/N}$ | $1/N$-portfolio as e.g. considered in DeMiguel et al. (2009) |
| $\mathbf{w}_t^\sharp$ | parametric portfolio policy |
| $\mathbf{w}_t^{\sharp,\geq 0}$ | parametric portfolio policy without short-selling |

**Table 5** Approximate CRRA Utility (In Sample): Strategies defined in Table 4. Empirical data. Training sample $t = 1, \ldots, 200$. Evaluation in sample; $t = 1, \ldots, 200$. Shrinkage parameter $c_p = 0.2$

| $\gamma = 0.25$ | $\mathbf{w}_t^\flat$ | $\mathbf{w}_{t,LW}^\flat$ | $\mathbf{w}_t^{\flat,\geq 0}$ | $\mathbf{w}_{t,LW}^{\flat,\geq 0}$ | $\mathbf{w}_t^{1/N}$ | $\mathbf{w}_t^\sharp$ | $\mathbf{w}_t^{\sharp,\geq 0}$ |
|---|---|---|---|---|---|---|---|
| $\widehat{\mathscr{C}}$ | 1.1656 | 1.1668 | 1.0340 | 1.0227 | 1.0153 | 1.0519 | 1.0153 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.1335 | 0.1332 | 0.0031 | 0.0020 | 0.0022 | 0.0128 | 0.0022 |
| mean($E_t$) | 1.4138 | 1.4160 | 1.0344 | 1.0229 | 1.0156 | 1.0594 | 1.0156 |
| sd($E_t$) | 0.6464 | 0.6509 | 0.0584 | 0.0385 | 0.0420 | 0.2329 | 0.0420 |
| Sharpe Ratio | 0.6387 | 0.6376 | 0.5717 | 0.5694 | 0.3465 | 0.2507 | 0.3465 |
| mean($w_{it} < 0$) | 0.1882 | 0.1881 | 0.0000 | 0.4546 | 0.0000 | 0.4546 | 0.0000 |
| sd($w_{it} < 0$) | 0.0478 | 0.0479 | 0.0000 | 0.0297 | 0.0000 | 0.0297 | 0.0000 |
| mean($w_{it} < -1$) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| sd($w_{it} < -1$) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $\gamma = 0.5$ | $\mathbf{w}_t^\flat$ | $\mathbf{w}_{t,LW}^\flat$ | $\mathbf{w}_t^{\flat,\geq 0}$ | $\mathbf{w}_{t,LW}^{\flat,\geq 0}$ | $\mathbf{w}_t^{1/N}$ | $\mathbf{w}_t^\sharp$ | $\mathbf{w}_t^{\sharp,\geq 0}$ |
| $\widehat{\mathscr{C}}$ | 0.6065 | 0.6131 | 1.0157 | 1.0320 | 1.0151 | 1.0434 | 1.0151 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.2460 | 0.2443 | 0.0012 | 0.0020 | 0.0015 | 0.0091 | 0.0015 |
| mean($E_t$) | 1.2110 | 1.2120 | 1.0160 | 1.0328 | 1.0156 | 1.0592 | 1.0156 |
| sd($E_t$) | 0.3331 | 0.3354 | 0.0324 | 0.0561 | 0.0420 | 0.2321 | 0.0420 |
| Sharpe Ratio | 0.6303 | 0.6293 | 0.4622 | 0.5673 | 0.3465 | 0.2510 | 0.3465 |
| mean($w_{it} < 0$) | 0.1813 | 0.1799 | 0.0000 | 0.0000 | 0.0000 | 0.4552 | 0.0000 |
| sd($w_{it} < 0$) | 0.0478 | 0.0481 | 0.0000 | 0.0000 | 0.0000 | 0.0297 | 0.0000 |
| mean($w_{it} < -1$) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| sd($w_{it} < -1$) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $\gamma = 1$ | $\mathbf{w}_t^\flat$ | $\mathbf{w}_{t,LW}^\flat$ | $\mathbf{w}_t^{\flat,\geq 0}$ | $\mathbf{w}_{t,LW}^{\flat,\geq 0}$ | $\mathbf{w}_t^{1/N}$ | $\mathbf{w}_t^\sharp$ | $\mathbf{w}_t^{\sharp,\geq 0}$ |
| $\widehat{\mathscr{C}}$ | 1.0934 | 1.0937 | 1.0130 | 1.0287 | 1.0147 | 1.0341 | 1.0147 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0142 | 0.0142 | 0.0017 | 0.0036 | 0.0030 | 0.0193 | 0.0030 |
| mean($E_t$) | 1.1095 | 1.1101 | 1.0132 | 1.0300 | 1.0156 | 1.0650 | 1.0156 |
| sd($E_t$) | 0.1765 | 0.1776 | 0.0236 | 0.0509 | 0.0420 | 0.2354 | 0.0420 |
| Sharpe Ratio | 0.6150 | 0.6140 | 0.5186 | 0.5698 | 0.3465 | 0.2720 | 0.3465 |
| mean($w_{it} < 0$) | 0.1653 | 0.1652 | 0.0000 | 0.0000 | 0.0000 | 0.4575 | 0.0000 |
| sd($w_{it} < 0$) | 0.0476 | 0.0466 | 0.0000 | 0.0000 | 0.0000 | 0.0351 | 0.0000 |
| mean($w_{it} < -1$) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| sd($w_{it} < -1$) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $\gamma = 2$ | $\mathbf{w}_t^\flat$ | $\mathbf{w}_{t,LW}^\flat$ | $\mathbf{w}_t^{\flat,\geq 0}$ | $\mathbf{w}_{t,LW}^{\flat,\geq 0}$ | $\mathbf{w}_t^{1/N}$ | $\mathbf{w}_t^\sharp$ | $\mathbf{w}_t^{\sharp,\geq 0}$ |
| $\widehat{\mathscr{C}}$ | 1.0490 | 1.0491 | 1.0082 | 1.0216 | 1.0138 | 1.0233 | 1.0138 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0078 | 0.0078 | 0.0011 | 0.0029 | 0.0031 | 0.0113 | 0.0031 |
| mean($E_t$) | 1.0588 | 1.0591 | 1.0084 | 1.0232 | 1.0156 | 1.0447 | 1.0156 |
| sd($E_t$) | 0.0982 | 0.0988 | 0.0163 | 0.0404 | 0.0420 | 0.1430 | 0.0420 |
| Sharpe Ratio | 0.5885 | 0.5876 | 0.4574 | 0.5499 | 0.3465 | 0.3059 | 0.3465 |
| mean($w_{it} < 0$) | 0.1416 | 0.1418 | 0.0000 | 0.0000 | 0.0000 | 0.4377 | 0.0000 |
| sd($w_{it} < 0$) | 0.0428 | 0.0438 | 0.0000 | 0.0000 | 0.0000 | 0.0334 | 0.0000 |
| mean($w_{it} < -1$) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

**Table 5** (continued)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $sd(w_{it} < -1)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0022 | 0.0000 | 0.0000 |
| $\gamma = 5$ | $\mathbf{w}_t^\flat$ | $\mathbf{w}_{t,LW}^\flat$ | $\mathbf{w}_t^{\flat,\geq0}$ | $\mathbf{w}_{t,LW}^{\flat,\geq0}$ | $\mathbf{w}_t^{1/N}$ | $\mathbf{w}_t^\sharp$ | $\mathbf{w}_t^{\sharp,\geq0}$ |
| $\widehat{\mathscr{C}}$ | 1.0214 | 1.0214 | 1.0056 | 1.0109 | 1.0108 | 1.0031 | 1.0108 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0173 | 0.0173 | 0.0028 | 0.0084 | 0.0139 | 0.0171 | 0.0139 |
| mean($E_t$) | 1.0284 | 1.0285 | 1.0059 | 1.0128 | 1.0156 | 1.0124 | 1.0156 |
| $sd(E_t)$ | 0.0514 | 0.0517 | 0.0104 | 0.0277 | 0.0420 | 0.0631 | 0.0420 |
| Sharpe Ratio | 0.5325 | 0.5316 | 0.4668 | 0.4275 | 0.3465 | 0.1801 | 0.3465 |
| mean($w_{it} < 0$) | 0.0870 | 0.0875 | 0.0000 | 0.0000 | 0.0000 | 0.3777 | 0.0000 |
| $sd(w_{it} < 0)$ | 0.0346 | 0.0341 | 0.0000 | 0.0000 | 0.0000 | 0.0454 | 0.0000 |
| mean$(w_{it} < -1)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $sd(w_{it} < -1)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

For $c_p = 0$, in our empirical data $\boldsymbol{\alpha}_t$ and $\mathcal{A}_t$ result in $w_{it}^\flat$ of large absolute value, where (18) results in poor performance. This problem can be expected, since the weights obtained in (18) are derived in a similar way as the investments $\boldsymbol{\phi}_t^*$. A second driver of larger portfolio weights is the parameter of relative risk aversion $\gamma$. The smaller $\gamma$, the larger the weights $\mathbf{w}_t^\flat$ in absolute value for any $c_p \geq 0$. Note that in contrast with the CARA case, our solutions for the CRRA case (with or without short selling constraints) are based on the Taylor series approximation (14) around $\mathbf{w} = \mathbf{0}_{N\times1}$. If the weights $w_{it}^\flat$ are quite far away from the approximation point of the Taylor series, the approximation quality can become poor.

Our shrinkage device also dampens this effect.

Table 4 presents the investment strategies to be compared in the following.

In Tables 5 and 6, we observe that for small $\gamma$ the weights following from (18) become quite large and the performance measured in terms of the certainty equivalent becomes poor (not only out of sample but also in sample). Surprisingly, this effect is stronger for $\gamma = 0.5$ than for 0.25. Some certainty equivalent samples become quite negative, and the sample standard deviation of the certainty equivalents becomes high.

By contrast, when finding an optimal $\theta$ in the case of parametric portfolio policies, no approximation of the expected utility function is used. Hence, also for small $\gamma$ the performance of the parametric portfolio approach is quite satisfactory. By comparing the estimates of the certainty equivalents with the parametric approach to the approximately optimal strategy described in (18), we observe that the parametric approach outperforms the approximately optimal approach for $\gamma = 0.5$. This holds for an in- and an out-of-sample comparisons without short selling constraints. By imposing short selling constraints, the optimal approach slightly dominates the 1/$N$ and the parametric policy.

Using the Sharpe ratio also verifies this result. We observe that when increasing the degree of risk aversion ($\gamma \geq 1$) optimal strategies based on (18) in terms of the points estimate of the certainty equivalent dominate the other strategies.

**Table 6** Approximate CRRA Utility (Out of Sample): Strategies defined in Table 4. Empirical data. Training sample $t = 1, \ldots, 200$. Evaluation out of sample; $t = 201, \ldots, 415$. Shrinkage parameter $c_p = 0.2$

| $\gamma = 0.25$ | $\mathbf{w}_t^\flat$ | $\mathbf{w}_{t,LW}^\flat$ | $\mathbf{w}_t^{\flat,\geq0}$ | $\mathbf{w}_{t,LW}^{\flat,\geq0}$ | $\mathbf{w}_t^{1/N}$ | $\mathbf{w}_t^\sharp$ | $\mathbf{w}_t^{\sharp,\geq0}$ |
|---|---|---|---|---|---|---|---|
| $\widehat{\mathscr{C}}$ | 0.7703 | 0.7563 | 1.0177 | 1.0127 | 1.0107 | 1.0582 | 1.0107 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.1521 | 0.1560 | 0.0036 | 0.0025 | 0.0023 | 0.0208 | 0.0023 |
| mean($E_t$) | 1.2135 | 1.2152 | 1.0183 | 1.0130 | 1.0109 | 1.0797 | 1.0109 |
| sd($E_t$) | 0.7643 | 0.7751 | 0.0701 | 0.0495 | 0.0445 | 0.4032 | 0.0445 |
| Sharpe Ratio | 0.2780 | 0.2764 | 0.2462 | 0.2433 | 0.2231 | 0.1952 | 0.2231 |
| mean($w_{it} < 0$) | 0.3265 | 0.3249 | 0.0000 | 0.4350 | 0.0000 | 0.4350 | 0.0000 |
| sd($w_{it} < 0$) | 0.0461 | 0.0465 | 0.0000 | 0.0240 | 0.0000 | 0.0240 | 0.0000 |
| mean($w_{it} < -1$) | 0.0045 | 0.0045 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| sd($w_{it} < -1$) | 0.0050 | 0.0050 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $\gamma = 0.5$ | $\mathbf{w}_t^\flat$ | $\mathbf{w}_{t,LW}^\flat$ | $\mathbf{w}_t^{\flat,\geq0}$ | $\mathbf{w}_{t,LW}^{\flat,\geq0}$ | $\mathbf{w}_t^{1/N}$ | $\mathbf{w}_t^\sharp$ | $\mathbf{w}_t^{\sharp,\geq0}$ |
| $\widehat{\mathscr{C}}$ | 0.3829 | 0.3223 | 1.0112 | 1.0165 | 1.0104 | 0.9783 | 1.0104 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.1698 | 0.1701 | 0.0014 | 0.0023 | 0.0015 | 0.0385 | 0.0015 |
| mean($E_t$) | 1.1097 | 1.1105 | 1.0117 | 1.0176 | 1.0109 | 1.0794 | 1.0109 |
| sd($E_t$) | 0.3912 | 0.3966 | 0.0412 | 0.0678 | 0.0445 | 0.4012 | 0.0445 |
| Sharpe Ratio | 0.2777 | 0.2761 | 0.2590 | 0.2455 | 0.2231 | 0.1954 | 0.2231 |
| mean($w_{it} < 0$) | 0.3190 | 0.3166 | 0.0000 | 0.0000 | 0.0000 | 0.4353 | 0.0000 |
| sd($w_{it} < 0$) | 0.0461 | 0.0458 | 0.0000 | 0.0000 | 0.0000 | 0.0238 | 0.0000 |
| mean($w_{it} < -1$) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| sd($w_{it} < -1$) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $\gamma = 1$ | $\mathbf{w}_t^\flat$ | $\mathbf{w}_{t,LW}^\flat$ | $\mathbf{w}_t^{\flat,\geq0}$ | $\mathbf{w}_{t,LW}^{\flat,\geq0}$ | $\mathbf{w}_t^{1/N}$ | $\mathbf{w}_t^\sharp$ | $\mathbf{w}_t^{\sharp,\geq0}$ |
| $\widehat{\mathscr{C}}$ | 1.0365 | 1.0363 | 1.0068 | 1.0151 | 1.0099 | 1.0035 | 1.0099 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0149 | 0.0152 | 0.0020 | 0.0043 | 0.0031 | 0.0315 | 0.0031 |
| mean($E_t$) | 1.0577 | 1.0581 | 1.0072 | 1.0170 | 1.0109 | 1.0777 | 1.0109 |
| sd($E_t$) | 0.2048 | 0.2075 | 0.0298 | 0.0632 | 0.0445 | 0.3727 | 0.0445 |
| Sharpe Ratio | 0.2771 | 0.2754 | 0.2088 | 0.2532 | 0.2231 | 0.2058 | 0.2232 |
| mean($w_{it} < 0$) | 0.3024 | 0.2993 | 0.0000 | 0.0000 | 0.0000 | 0.4318 | 0.0000 |
| sd($w_{it} < 0$) | 0.0444 | 0.0439 | 0.0000 | 0.0000 | 0.0000 | 0.0223 | 0.0000 |
| mean($w_{it} < -1$) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| sd($w_{it} < -1$) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $\gamma = 2$ | $\mathbf{w}_t^\flat$ | $\mathbf{w}_{t,LW}^\flat$ | $\mathbf{w}_t^{\flat,\geq0}$ | $\mathbf{w}_{t,LW}^{\flat,\geq0}$ | $\mathbf{w}_t^{1/N}$ | $\mathbf{w}_t^\sharp$ | $\mathbf{w}_t^{\sharp,\geq0}$ |
| $\widehat{\mathscr{C}}$ | 1.0194 | 1.0192 | 1.0062 | 1.0122 | 1.0089 | 1.0033 | 1.0089 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0080 | 0.0081 | 0.0014 | 0.0038 | 0.0031 | 0.0162 | 0.0032 |
| mean($E_t$) | 1.0318 | 1.0320 | 1.0066 | 1.0152 | 1.0109 | 1.0500 | 1.0109 |
| sd($E_t$) | 0.1118 | 0.1131 | 0.0208 | 0.0556 | 0.0445 | 0.2277 | 0.0445 |
| Sharpe Ratio | 0.2753 | 0.2737 | 0.2682 | 0.2553 | 0.2231 | 0.2152 | 0.2232 |
| mean($w_{it} < 0$) | 0.2688 | 0.2680 | 0.0000 | 0.0000 | 0.0000 | 0.4222 | 0.0000 |
| sd($w_{it} < 0$) | 0.0409 | 0.0398 | 0.0000 | 0.0000 | 0.0000 | 0.0214 | 0.0000 |
| mean($w_{it} < -1$) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

**Table 6** (continued)

| | $\mathbf{w}_t^\flat$ | $\mathbf{w}_{t,LW}^\flat$ | $\mathbf{w}_t^{\flat,\geq 0}$ | $\mathbf{w}_{t,LW}^{\flat,\geq 0}$ | $\mathbf{w}_t^{1/N}$ | $\mathbf{w}_t^\sharp$ | $\mathbf{w}_t^{\sharp,\geq 0}$ |
|---|---|---|---|---|---|---|---|
| $sd(w_{it} < -1)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0073 | 0.0000 | 0.0000 |
| $\gamma = 5$ | $\mathbf{w}_t^\flat$ | $\mathbf{w}_{t,LW}^\flat$ | $\mathbf{w}_t^{\flat,\geq 0}$ | $\mathbf{w}_{t,LW}^{\flat,\geq 0}$ | $\mathbf{w}_t^{1/N}$ | $\mathbf{w}_t^\sharp$ | $\mathbf{w}_t^{\sharp,\geq 0}$ |
| $\widehat{\mathscr{C}}$ | 1.0081 | 1.0080 | 1.0032 | 1.0068 | 1.0057 | 0.9921 | 1.0057 |
| $\widehat{sd}\left(\widehat{\mathscr{C}}\right)$ | 0.0169 | 0.0171 | 0.0033 | 0.0103 | 0.0137 | 0.0292 | 0.0137 |
| $\text{mean}(E_t)$ | 1.0162 | 1.0162 | 1.0035 | 1.0101 | 1.0109 | 1.0162 | 1.0109 |
| $sd(E_t)$ | 0.0564 | 0.0569 | 0.0116 | 0.0359 | 0.0445 | 0.1064 | 0.0445 |
| Sharpe Ratio | 0.2694 | 0.2679 | 0.2164 | 0.2524 | 0.2231 | 0.1432 | 0.2231 |
| $\text{mean}(w_{it} < 0)$ | 0.1882 | 0.1907 | 0.0000 | 0.0000 | 0.0000 | 0.3807 | 0.0000 |
| $sd(w_{it} < 0)$ | 0.0358 | 0.0359 | 0.0000 | 0.0000 | 0.0000 | 0.0522 | 0.0000 |
| $\text{mean}(w_{it} < -1)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $sd(w_{it} < -1)$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

In this case, the results without constraints are slightly above the results with constraints.

The results for $\mathbf{w}^{\sharp,\geq 0}$ are very close the results for the 1/*N*-strategy.

**Observation 5** (i) In sample: Similar to the CARA case, the optimal shrinkage strategy shows the best performance for $\gamma \geq 1$. The performances of the alternative strategies as measured by the certainty equivalent perform quite similarly.

(ii) Out of sample: The 1/*N*-strategy shows the best performance for $\gamma \leq 0.5$. For a very small $\gamma$, the best performance with the parametric policy is observed. For $\gamma \geq 1$, the best performance is achieved with the optimal shrinkage strategy; however, the performances for the strategies considered are roughly the same across strategies.

(iii) Similar to the CARA case, for small values of risk aversion parametric portfolio policies imply a large amount of short selling in and out of sample.

## 8 Conclusions

Theory-guided reduction techniques prove particularly helpful for machine learning applications as forcefully argued by Nagel (2021).[20] Such techniques are extremely valuable under conditions that are challenging for the determination of optimal portfolios either because utility frontiers are rather flat or even exhibit multiple optima. Not only in such cases our new shrinkage facility renders portfolio strategies less risky and improves performance when applied to empirical data. In contrast with

---

[20] In the words of Nagel (2021), we provide "an analytical framework that allows to inject a limited amount of economic reasoning when we set up ML [machine learning] tools to tackle asset pricing problems." Nagel (2021)[p. 63].

a bulk of methods proposed in the literature, we do not shrink (some of) the model parameters but shrink the portfolio weights toward some a-priori specified strategy.

We test our approach in simulation exercises on data for the S&P 500, the US market for large caps.[21] For the scenario when parameters actually can be estimated directly, we observe that the simple optimal shrinkage strategy proposed in this article outperforms the parametric portfolio approach of Brandt et al. (2009), and the 1/$N$-strategy, for most levels of absolute and relative risk aversion. Only for CRRA preferences with very low levels of risk aversion, the other strategies are superior. For higher degrees of risk aversion, the performances of the strategies considered are quite similar.

The demand systems approach to asset pricing introduced by Koijen and Yogo (2019) lends itself to numerous applications, such as the intermediary asset pricing theory of He and Krishnamurthy (2013) or asset pricing with frictions more generally. In this article, our shrinkage approach also augments the demand systems approach to CARA and CRRA expected utility. We consider the cases with and without short selling constraints and show the existence of equilibrium.

Another aspect of the demand system approach is its relation to the BSV characteristics-based parametric portfolio approach (see Brandt et al. 2009 ) that has received a lot of interest from empirical researchers because it provides an attractive reduction technique to an otherwise complex optimization problem. From the results obtained in this article, we observe that parametric portfolio strategies can be optimal under rather strong assumptions.

While our work, as a first step, has focused on a quasi-static analysis and evaluation, a promising route for future research would seem as the next step to consist in a dynamic implementation of optimal shrinkage strategies, allowing for time dependence and tuning of the shrinkage parameter. Alternatively different shrinkage portfolios can be investigated. For example, does it make sense to shrink toward the risk-free asset in risky periods, etc. In the current implementation, the model parameters are estimated in the training sample and not adapted in the evaluation sample. It is tempting to experiment with rolling windows or more sophisticated dynamic models in order to improve out-of-sample performance.

## Appendix: Empirical data

In the study, we use the characteristics and returns of all 100 firms that are *continuously* a member of the S&P 500 firms in the time span from 04/1979 to 12/2013. The three characteristics are closely based on Brandt et al. (2009). Market equity, $me_{it}$, is the natural logarithm of the number of shares outstanding (Compustat item cshoq for the primary issue—priusa) times the closing price (prccq). Book-to-Market, $btm_{it}$, is the natural logarithm of (1 + book equity / market equity), where book equity is measured as Shareholders' Equity (seq) and is used six months after the

---

[21] In related work Gehrig et al. (2018), similar evidence extends to CRSP-data for low enough risk aversion.

**Table 7** Pearson correlation coefficients for S&P 500 data

| | $r_{it}$ | $\tilde{x}_{it,1}$ | $\tilde{x}_{it,2}$ | $\tilde{x}_{it,3}$ |
|---|---|---|---|---|
| $r_{it}$ | 1.0000 | 0.0005 | -0.0029 | −0.0065 |
| $\tilde{x}_{it,1}$ | 0.0005 | 1.0000 | −0.4846 | −0.4724 |
| $\tilde{x}_{it,2}$ | −0.0029 | -0.4846 | 1.0000 | −0.4906 |
| $\tilde{x}_{it,3}$ | −0.0065 | -0.4724 | -0.4906 | 1.0000 |

close of the fiscal year to ensure availability of the data. Momentum, $mom_{it}$, is the cumulative return over the time period $t$-13 to $t$-2, expressed as monthly average. Hence, we get $k = 3$ and $\mathbf{x}_{it} = \left(me_{it}, btm_{it}, mom_{it}\right)^{\top}$. To be included in the estimation, a firm must fulfill three conditions at the portfolio formation. It must be a continuous constituent of the S&P 500 (ticker i0003), must have data for all three characteristics, and needs to have return data (trt1m) over the following month. The number of included firms $N$ is always equal to 100. All characteristics $\mathbf{x}_{it}$ are cross-sectionally standardized, resulting in $\tilde{\mathbf{x}}_{it}$.

Table 7 provides correlation coefficients; the first-order autocorrelations of the variables $x_{it,j}$ are 0.9621, 0.9706 and 0.8638. To further investigate the relationship between the returns and the variables $\tilde{\mathbf{x}}_{it}$, we estimated the pooled model

$$r_{it} = a + \mathbf{b}^{\top} \tilde{\mathbf{x}}_{it} + u_{it} , \tag{20}$$

where the noise terms are—in a first step—assumed to be exogenous. The ordinary least squares estimates are

$\hat{a} = 0.1471, \hat{\mathbf{b}} = (0.0104, -0.0021, -0.1178)^{\top}$, where the corresponding $p$-values for $\mathbf{b}$ are all $< 0.01$.

That is, the linear relationship between $r_{it+1}$ and $\tilde{\mathbf{x}}_{it}$ is significant for $\tilde{x}_{it,1}$ and $\tilde{x}_{it,2}$ on a 5% significance level. Since the variables $\tilde{\mathbf{x}}_{it}$ are at least partially jointly determined with the returns, the assumption of exogenous regressors is a strong one. Therefore, we estimated the panel regression model (20) by means of instrumental variables, where we assumed that the noise term $u_{it}$ is uncorrelated with $\tilde{\mathbf{x}}_{is}$, $s < t$. Based on this assumption, we estimate $\mathbf{b}$ by using $\tilde{\mathbf{x}}_{it-1}$ as instruments and obtained the two-stage least squares estimates $\hat{a}_{IV} = 0.1495$ and $\hat{\mathbf{b}}_{IV} = (0.0111, -0.0020, -0.1214)^{\top}$; all p-values are $< 0.001$.

## Declarations

# References

Ammann, M., Coqueret, G., Schade, J.-P.: Characteristics-based portfolio choice with leverage constraints. J. Bank. Finance **70**, 23–37 (2016)

Barigozzi, M., Brownlees, C.: Nets: network estimation for time series. J. Appl. Econom. **34**(3), 347–364 (2019)

Black, F., Litterman, R.: Global portfolio optimization. Financ. anal. J. **48**(5), 28–43 (1992)

Brandt, M.W., Santa-Clara, P.: Dynamic portfolio selection by augmenting the asset space. J. Finance **61**(5), 2187–2217 (2006)

Brandt, M.W., Santa-Clara, P., Valkanov, R.: Parametric portfolio policies: exploiting characteristics in the cross-section of equity returns. Rev. Finan. Stud. **22**(9), 3411–3447 (2009)

Carhart, M.M., Carpenter, J.N., Lynch, A.W., Musto, D.K.: Mutual fund survivorship. Rev. Finan. Stud. **15**(5), 1439–1463 (2015)

Carpenter, J.N., Lynch, A.W.: Survivorship bias and attrition effects in measures of performance persistence. J. Finan. Econ. **54**(3), 337–374 (1999)

Cong, L.W., Tang, K., Wang, J., Zhang, Y.: Alphaportfolio: direct construction through deep reinforcement learning and interpretable ai. Asset Pricing and Valuation eJournal, Capital Markets (2020)

Davidson, R., MacKinnon, J.G.: Estimation and inference in econometrics. Oxford University Press, New York (1993)

DeMiguel, V., Garlappi, L., Uppal, R.: Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? Rev. Finan. Stud. **22**(5), 1915–1953 (2009)

Ferson, W.E., Siegel, A.F.: The efficient use of conditioning information in portfolios. J. Finance **56**(3), 967–982 (2001)

Freyberger, J., Neuhierl, A., Weber, M.: Dissecting characteristics nonparametrically. Rev. Finan. Stud. **33**(5), 2326–2377 (2020)

Gehrig, T., Sögner, L., and Westerkamp, A.: Making parametric portfolio policies work. CEPR Discussion Paper 13193 (2018)

Goeree, J.K., Palfrey, T.R., Holt, C.A.: Risk averse behavior in generalized matching pennies games. Games Econ. Behav. **45**(1), 97–113 (2002)

Hansen, L.P.: Large sample properties of generalized method of moments estimators. Econometrica **50**(4), 1029–1054 (1982)

Harrison, G. W. and Rutström, E. E.: Risk aversion in the laboratory. In *of Research in Experimental Economics. Emerald Group Publishing Limited*, pages 41–196 (2008)

Hastie, T., Montanari, A., Rosset, S., Tibshirani, R.J.: Surprises in high-dimensional ridgeless least squares interpolation. Annals Stat. **50**(2), 949–986 (2022)

Hastie, T., Tibshirani, R., and Friedman, J.: *The elements of statistical learning: data mining, inference, and prediction*. Springer series in statistics. Springer (2009)

He, Z., Krishnamurthy, A.: Intermediary asset pricing. Am. Econ. Rev. **103**, 732–757 (2013)

Hjalmarsson, E., Manchev, P.: Characteristic-based mean-variance portfolio choice. J. Bank. Finance **36**(5), 1392–1401 (2012)

James, G., Witten, D., Hastie, T., and Tibshirani, R.: *An introduction to statistical learning – with applications in R*, volume 103 of *Springer Texts in Statistics*. Springer, New York (2017)

Kelly, B.T., Malamud, S., Zhou, K.: The Virtue of Complexity in Return Prediction, pp. 21–90. Technical report, Swiss Finance Institute Research Paper No (2021)

Klenke, A.: Probability theory - A comprehensive course. Springer (2008)

Koijen, R. S. J., Richmond, R., and Yogo, M.: Which investors matter for equity valuations and expected returns? *University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2019-92, NYU Stern School of Business* (2023)

Koijen, R.S.J., Yogo, M.: A demand system approach to asset pricing. J. Political Econ. **127**(4), 1475–1515 (2019)

Ledoit, O., Wolf, M.: A well-conditioned estimator for large-dimensional covariance matrices. J. Multivar. Anal. **88**(2), 365–411 (2004)

Ledoit, O., Wolf, M.: Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. Revi. Finan. Stud. **30**(12), 4349–4388 (2017)

Ledoit, O. and Wolf, M.: Markowitz portfolios under transaction costs. ECON - Working Papers 420, Department of Economics - University of Zurich (2024)

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

Medeiros, M.C., Passos, A.M., Vasconcelos, G.F.R.: Parametric portfolio selection: evaluating and comparing to Markowitz Portfolios. Brazilian Rev. Finance **12**(2), 257–284 (2014)

Nagel, S.: *Machine Learning in Asset Pricing*. Princeton Lectures in Finance. Princeton University Press (2021)

Simon, C., Blume, L.: Mathematics for Economists. Norton (1994)

Tanaka, T., Camerer, C.F., Nguyen, Q.: Risk and time preferences: linking experimental and household survey data from vietnam. Am. Econ. Rev. **100**(1), 557–71 (2010)

Thös, A.-K.: *Naive Diversification with fewer assets – A risk reduction approach using clustering methods*. PhD Dissertation, Technical University of Kaiserslautern (2019)

Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. R. Stat. Soc.: Series B (Stat. Methodol.) **67**(2), 301–320 (2005)

**Thomas P. Gehrig** is professor of finance at the University of Vienna. He is a member of the Vienna Graduate School of Finance and a research fellow of the CEPR, the ECGI and the Systemic Risk Center at LSE. He is the founding speaker of the Oskar-Morgenstern-Doctoral School and Editor-in-Chief of the Schmalenbach Journal of Business Research. Prior to his Vienna appointment, he held professorial positions at the University of Basel, Northwestern University, University of Freiburg and Rice University. He has been a recipient of a grant of the 5th European Framework Program as well as numerous visiting grants.

**Leopold Sögner** obtained his doctoral degree from the Vienna University of Economics and Business. He joined the Institute for Advanced Studies in 2008 from the Vienna University of Technology and has served as director of graduate studies, head of the research group "Financial Markets and Econometrics" and is currently head of the "Macroeconomics and Business Cycles" research group. In 2018, he was guest professor at NYU Abu Dhabi. Since 2010, he has been full faculty member at the Vienna Graduate School of Finance and has been principal investigator in several research grants.

**Arne Westerkamp** holds a doctorate from the Vienna University of Economics and Business, where he received the WU City of Vienna Best Paper Award. He is currently a research executive at IQAM Invest GmbH, but the views expressed in this paper do not necessarily represent the views of IQAM Invest. His research interests are factor-based multi-asset allocation, sustainable investing and capital structure. He is a guest lecturer in financial economics and has published articles in The Review of Financial Studies, Journal of Risk and China Finance Review International