DISCUSSION PAPER SERIES

DP17743

EXTENDING THE DEMAND SYSTEM APPROACH TO ASSET PRICING

Thomas Gehrig and Leopold Sögner

ASSET PRICING

ISSN 0265-8003

EXTENDING THE DEMAND SYSTEM APPROACH TO ASSET PRICING

Thomas Gehrig and Leopold Sögner

Discussion Paper DP17743
Published 12 December 2022
Submitted 03 December 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Asset Pricing

EXTENDING THE DEMAND SYSTEM APPROACH TO ASSET PRICING

Abstract

We extend the demand systems approach of Koijen and Yogo (2019) to more general classes of preferences. Specifically we analyse constant absolute and constant relative risk aversion, provide conditions for the existence of equilibrium, and evaluate equilibrium prices at US-data. We find that constant absolute risk aversion works particularly well at moderate levels of risk aversion. In the case of relative risk aversion, optimal interior portfolio solutions may not even exist. In both preference classes especially out-of-sample predictions are rather volatile. In order to improve out-of-sample performance we augment the optimal strategies by a shrinkage device. As a side product we establish that the characteristics-based parametric portfolio approach of Brandt, Santa Clara and Valkanov (2009) can only be justified as optimal investments under exceedingly strong assumptions. In empirical data the 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.

JEL Classification: C51, G11, G12

Keywords: parametric portfolio approach, Expected utility, Risk aversion, Machine learnings

Thomas Gehrig - thomas.gehrig@univie.ac.at
University of Vienna and CEPR

Leopold Sögner - soegner@ihs.ac.at
IHS Vienna

Acknowledgements
Gehrig gratefully acknowledges the hospitality of the Financial Markets Group at LSE. Without implicating them we are grateful for the comments of Suleyman Basak, Jonathan Berk, Thomas Dangl, Richard Franz, Jens Jackwerth, Christian Julliard, Ian Martin, Roland Mestel, Otto Randl, Andrea Tarelli, Rossen Volkanov and Josef Zechner and the participants of the AWG in Graz, ESSFM in Gerzensee as well as the Finance Brown Bag Seminar at WU.

Extending the Demand System Approach to Asset Pricing

Thomas Gehrig* Leopold Sögner† Arne Westerkamp‡

December 3, 2022§

Abstract

We extend the demand systems approach of Koijen and Yogo (2019) to more general classes of preferences. Specifically we analyse constant absolute and constant relative risk aversion, provide conditions for the existence of equilibrium, and evaluate equilibrium prices at US-data. We find that constant absolute risk aversion works particularly well at moderate levels of risk aversion. In the case of relative risk aversion, optimal interior portfolio solutions may not even exist. In both preference classes especially out-of-sample predictions are rather volatile. In order to improve out-of-sample performance we augment the optimal strategies by a shrinkage device. As a side product we establish that the characteristics-based parametric portfolio approach of Brandt, Santa Clara and Valkanov (RFS 2009) can only be justified as optimal investments under exceedingly strong assumptions. In empirical data the 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.

JEL: C51, G11, G12.

*Thomas Gehrig (thomas.gehrig@univie.ac.at), Tel. +43 1 4277 38071, Faculty of Business, Economics, and Statistics, University of Vienna Oskar- Morgenstern-Platz 1 A-1090 Vienna, Austria.

†Leopold Sögner (soegner@ihs.ac.at), Tel. +43 1 59991 182, Department of Economics and Finance, Institute for Advanced Studies, Josefstädter Straße 39, A-1080 Vienna, Austria. T. Gehrig and L. Sögner have a further affiliation with the Vienna Graduate School of Finance (VGSF).

‡Arne Westerkamp (arne.westerkamp@iqam.com), Tel. +43 1 31336 5070, IQAM Research Center, Wollzeile 36-38, 1010 Wien. The views expressed herein are solely those of the authors and do not necessarily represent the views of IQAM Invest.

§Gehrig gratefully acknowledges the hospitality of the Financial Markets Groupt at LSE. Without implicating them we are grateful for the comments of Suleyman Basak, Jonathan Berk, Thomas Dangl, Richard Franz, Jens Jackwerth, Christian Julliard, Ian Martin, Roland Mestel, Otto Randl, Andrea Tarelli, Rossen Volkanov and Josef Zechner and the participants of the AWG in Graz, ESSFM in Gerzensee as well as the Finance Brown Bag Seminar at WU.

1 Introduction

In their widely acclaimed contribution Koijen and Yogo (2019) develop an asset pricing model with flexible heterogeneity in asset demand across investors. Their framework is especially useful in modelling non-atomic investors such as large institutions and pension funds. In their framework, with log-utility and short-selling constraints in place, optimal portfolio choice reduces to characteristics-based demand, when returns exhibit a factor structure, which 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.

We extend the approach of Koijen and Yogo (2019) in various dimensions. First, 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 absence of short-selling constraints in order to analyse and evaluate the empirical relevance of this restriction. 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), and constant relative risk aversion (CRRA) preferences. This result still holds if a – not-necessarily proper – subset of the agents apply the shrinkage device proposed in this article. Finally, we illustrate the performance of those extensions at the hand of US stock market data on asset prices.[1]

The basic insights from extending the demand systems approach to asset pricing 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

[1] In the absence of individual holdings data, in contrast to Koijen and Yogo (2019) our empirical analysis focuses on asset pricing only since we cannot identify demand.

absence of wealth effects. We demonstrate that our optimal strategies with shrinkage outperform parametric portfolio policies and a simple $1/N$ investment strategy.

- 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 US in the period from 1995-2013 especially for low levels of relative risk aversion. Overall we find that the performance of the "constant relative risk aversion-adaptions" are relatively poor for low levels of relative risk aversion γ. However, the performance is improving for higher levels, both in-sample as well as out-of-sample. We observe that for moderate and higher γ our optimal strategies with shrinkage outperform parametric portfolio policies and a simple $1/N$ investment strategy. For large γ 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 to Nagel (2021), where ridge regression is used to predict returns, we propose an algorithm that allows to shrink towards 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 analyses CRRA-preferences. Section 6 presents asset prices derived in general asset market equilibrium for both, the CARA as well as the CRRA-preferences discussed in the sections before. Section 7 presents an empirical evaluation of the pricing theories at a sample of onehundred S&P 500 stocks. This chapter also provides robust empirical evidence of potential pitfalls for the unchecked parametric portfolio approach. Section 8 concludes. The Appendix contains a section on the properties of the empirical data and a detailed derivation of the CARA model in Section 4.

2 Literature and Relations to Machine Learning

As already stated in the 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).[2]

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

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

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 was used. The authors show that using a bulk of "plausibly relevant predictors" in combination with "rich non-linear models", improves return forecasting and portfolio returns. Non-parametric 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) 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).

In this article we augment our objective function (that is, either CARA or CRRA expected utility) by a quadratic penalty term. In contrast to 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. 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 towards 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 well known to 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 towards 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 to the machine learning approaches discussed above, our approach directly allows to integrate a-priori information on investment weights.

3 Model and Assumptions

We follow Koijen and Yogo (2019) and consider an economy with discrete time t. 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 $R_{it+1} := 1+r_{it+1}$. The index set of traded risky securities (e.g. stocks) in period t is abbreviated by $\mathbb{I}_t$.[3] N_t, $1 \leq N_t \leq N < \infty$, is the number of risky elements in $\mathbb{I}_t$. In the case a risk-free asset is traded we apply the index $i = f$, its return is r_{ft+1}, and the total number of assets is $n_t = N_t + 1$ in $\mathbb{I}_t$; in sums the summation index 0 is used for the risk-free asset. 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 y $P_{it}S_{it}$ and aggregate market capitalization reads $\sum_{i\in\mathbb{I}_t} P_{it}S_{it}$. Denote the vector of share prices by $\mathbf{P}_t := (P_{1t}, \ldots, P_{N_t t})^\top$.

If the number of securities is constant, then $N_t = N$ (and $n_t = n$) for all $t = 1, \ldots, T$. For a given set of weights $w_{it} \in \mathbb{R}$, the portfolio return is $r_{pt+1} := \sum_{i\in\mathbb{I}_t} w_{it}r_{it+1}$, with $R_{pt+1} := 1 + r_{pt+1}$ denoting the portfolio's gross return.[4]

We collect observed characteristics in $\mathbf{x}_{it} \in \mathbb{R}^k$, $i = 1, \ldots, n_t$, where $\mathbf{x}_{it}$ could contain endogenous, predetermined and or exogenous variables. [5] 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

[3]For example, the set $\mathbb{I}_t$ contains S&P 500 or CRSP-identifiers.

[4]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 jth 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 $\mathbb{V}_t(\mathbf{R}_{t+1})$.

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

explicitly assume that prior investment weights or amounts invested are not contained in $\breve{\mathbf{x}}_{it}$. Let

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

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

$$\underbrace{\begin{pmatrix} R_{1t} \\ \vdots \\ R_{n_t t} \end{pmatrix}}_{\mathbf{R}_t} = \underbrace{\begin{pmatrix} a_{01t} \\ \vdots \\ a_{0n_t t} \end{pmatrix}}_{\mathbf{a}_{0t}} + \underbrace{\begin{pmatrix} \mathbf{A}_{1t} & \mathbf{0} & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{n_t t} \end{pmatrix}}_{\mathbf{A}_t} \underbrace{\begin{pmatrix} \mathbf{y}_{1t} \\ \vdots \\ \mathbf{y}_{n_t} \end{pmatrix}}_{\mathbf{y}_t \in \mathbb{R}^{n_t k}} + \underbrace{\begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{n_t t} \end{pmatrix}}_{\tilde{\boldsymbol{\varepsilon}}_t} .$$

$\mathbf{A}_{jt}$, $j = 1, \ldots, n_t$, are $1 \times k_y$-dimensional matrices. In the case a risk-free asset is traded $\mathbf{A}_f = \mathbf{0}_{1\times k_y}$, $\mathbf{x}_{ft} = \mathbf{0}_{k\times 1}$, $\varepsilon_{ft} = 0$, and $R_{ft} = a_{0ft}$. The vector of noise terms $\tilde{\boldsymbol{\varepsilon}}_t$, contains the N_t-dimensional subvector $\boldsymbol{\varepsilon}_t$ affecting the risky assets. Its expectation is zero and covariance matrix $\boldsymbol{\Sigma}_t$. To slightly simplify the analysis and in contrast to Koijen and Yogo (2019) we did not impose a factor structure on the covariance matrix $\boldsymbol{\Sigma}_t$, 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.] Koijen and Yogo (2019) as well as this article allows for investor dependent $\mathbf{A}_{it}$, a_{it} and $\mathbf{y}_{it}$. However to simplify the notation an additional index for an investor is only included if necessary. Next we impose

Assumption 1. *(i) $\mathbf{R}_t$, $\mathbf{y}_t$, and $\boldsymbol{\varepsilon}_t$ are jointly stationary and ergodic.*[6] *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, such that the conditional expectation of the return of asset i is $a_{it} + \mathbf{A}_{it}\mathbf{y}_{it}$.* The covariance matrix $\boldsymbol{\Sigma}_t$ 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

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

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 $\mathbf{y}_{it}$.

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_t alternatives, in addition, a risk-free asset can but need not be traded. Portfolio optimization traditionally involves the optimal determination of those weights w_{it} (or amounts invested into asset i, ϕ_{it},) with respect to a utility function, potential endowment and trading constraints. $\phi_{it} = e_t w_{it}$ is the amount invested in monetary units in asset i, while $\mathbf{w}_t := (w_{1t}, \ldots, w_{N_t,t})^\top \in \mathbb{R}^N$ and $\boldsymbol{\phi}_t := (\phi_{1t}, \ldots, \phi_{N_t,t})^\top \in \mathbb{R}^{N_t}$ are the investments (investment weights) into the risky assets in the following. Let $\mathcal{W} \subset \mathbb{R}^{n_t}$ and $\mathcal{W}_\phi \subset \mathbb{R}^{n_t}$ 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 $R_{pt+1} = \sum_{i=(N_t-n_t)+1}^{n_t} w_{it} R_{it+1}$, resulting in the optimization problem

$$\max_{(w_{ft},\mathbf{w}_t)^\top \mathcal{W}} \mathbb{E}_t\left(u\left(e_t R_{pt+1}\right)\right) = \max_{(w_{ft},\mathbf{w}_t)^\top \in \mathcal{W}} \mathbb{E}_t\left(u\left(E_{t+1}\right)\right) = \max_{(w_{ft},\mathbf{w}_t)^\top \in \mathcal{W}} \mathbb{E}_t\left(u\left(e_t\left(1 + \sum_{i \in \mathbb{I}_t} w_{it} r_{it+1}\right)\right)\right) , \quad (1)$$

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 (1), the e_t invested are deterministic. Let us define an investment rule as *characteristics based* if w_{it} or the amounts invested are affine in $\mathbf{y}_{it}$, $i = 1, \ldots, N_t$. That is, for investment weights we have[7]

$$w_{it} = \pi_{it} + \boldsymbol{\Pi}_{it} \mathbf{y}_t \ , \quad \text{for all } i = 1, \ldots, N_t \ . \quad (2)$$

Given a vector of investment weights $\mathbf{w}_t := (w_{1t}, \ldots, w_{N_t t})^\top \in \mathbb{R}^{N_t}$ 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

[7]This definition is different from Koijen and Yogo (2019)[see equation (10) there] where $\ln(w_{it}/w_{ft})$ [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 $\mathbf{y}_{it}$, $i = 1 \ldots, N_t$. Since a risk-free asset need not be traded we proceed with the definition provided in (2).
To reduce the notational burden we simply write $\mathbf{w}_t^*$ instead of $\mathbf{w}_t^*(\tilde{\mathbf{y}}_t)$, etc.

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 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.

In contrast to the standard Markowitz (1952) approach, where optimal portfolio weights typically depend on a large number of first and second moments of the return distribution, the parametric portfolio policy of Brandt et al. (2009) reduces the dimensionality of the optimization problem by modelling a small number of drivers of the portfolio weights directly.[8] Often the dimensionality of the drivers $\tilde{x}_{it}$ is very low (e.g. 3 in our empirical setting below) and only investments into risky assets are considered ($w_{ft} = 0$). Specifically, the weights are modelled as functions $w_{it} = f_i(N_t, \tilde{x}_{it}; \boldsymbol{\theta})$. Typically, $\tilde{x}_{it}$ is a vector of standardized variables. That is, for observed variables $\boldsymbol{\chi}_t \in \mathbf{R}^{k_\chi}$,

$$\tilde{\mathbf{x}}_{it} := \left(diag \left(\frac{1}{N_{t-1} - 1} \sum_{i \in \mathbb{I}_t} \left(\boldsymbol{\chi}_{it} - \frac{1}{N_t} \sum_{i \in \mathbb{I}_t} \boldsymbol{\chi}_{it} \right)^2 \right) \right)^{-0.5} \left(\boldsymbol{\chi}_{it} - \frac{1}{N_t} \sum_{i \in \mathbb{I}_t} \boldsymbol{\chi}_{it} \right) . \tag{3}$$

$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{k_\chi})^\top$ is a k_χ-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 (1) is maximized. Following Brandt et al. (2009) we also focus our attention on the linear function

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

Equation (4) results in the parametric portfolio strategy $\phi_{it}^\sharp := e_t w_{it}$ such that $\boldsymbol{\phi}_t^\sharp := (e_t w_{it}, \ldots, e_t w_{N_t,t})^\top$. In all applications we work with $\bar{w}_{it} = 1/N_t$, where $\bar{\mathbf{w}}_t := (w_{1t}, \ldots, w_{N_t t})^\top$. Some further results on parametric portfolio policies are provided in Appendix A.

Finally, the vector $\tilde{x}_{it}$ used for parametric portfolio policies and the observed characteristics driving

[8] 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).

expected returns need not to be the same. To simplify notation and to provide a fair comparison between parametric strategies and some (approximately) optimal strategies obtained later we set $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it}$ (hence also $k = k_\chi$), where the standardized characteristics $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it}$ are assumed to be stationary.

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* $\boldsymbol{\theta}$*s) can be derived analytically but only for the case of mean-variance preferences. [Compare also to discussion about in optimal* $\boldsymbol{\theta}$ *in the Appendix F.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.[9] The number of risky assets is $N_t = N$.

For the CARA case it is easier to work with the amounts invested $\boldsymbol{\phi}_t$. The weights of investments into the risky assets follow from $\boldsymbol{\phi}_t$ and e_t, that is $\mathbf{w}_t = \frac{1}{\mathbf{1}_N^\top \boldsymbol{\phi}_t}\boldsymbol{\phi}_t$. In addition to N risky assets we also consider the case where a risk-free asset is traded [the risk-free asset has cross-sectional index $i = f$, in the case $n = N + 1$]. The portfolio vector of risky-assets is $\boldsymbol{\phi}_t = (\phi_{1t}, \ldots, \phi_{Nt})^\top \in \mathbb{R}^N$, where ϕ_{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:

[9] Problems considered in Observation 6(b) do not show up. For example, in companion work (Gehrig et al., 2018) we have explored behavioural utility functions.

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 $\mathbb{E}_t(\mathbf{R}_{t+1})$ and conditional covariance $\boldsymbol{\Sigma}_t = \mathbb{V}_t(\mathbf{R}_{t+1})$ satisfies $0 < \mathbb{V}_t(\mathbf{R}_{t+1}) < \infty$ [i.e. the conditional covariance matrix is finite and regular].

We first analyse 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

$$\mathbb{E}_t(-\exp(-\rho E_{t+1})) \quad = \quad -\exp\left[-\rho e_t - \rho \boldsymbol{\phi}_t^{\top}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - \mathbf{1}_N R_{ft+1}\right) + \frac{\rho^2}{2}\boldsymbol{\phi}_t^{\top}\mathbb{V}_t(\mathbf{R}_{t+1})\boldsymbol{\phi}_t\right]. \tag{5}$$

Maximizing (5) yields the vector of optimal amounts invested into the risky assets

$$\boldsymbol{\phi}_t^*(\mathbf{x}_t) \quad = \quad \Bigg(\underbrace{\frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right)^{-1}}_{\mathbf{B}_t}\Bigg)^{-1} \underbrace{\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - R_{ft}\mathbf{1}_N\right)}_{\mathbf{b}_t}. \tag{6}$$

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]

$$\boldsymbol{\phi}_t^{+}(\mathbf{x}_t) \quad = \quad \frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - \frac{\rho\left(\frac{1}{\rho}\mathbf{1}_N^{\top}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbb{E}_t(\mathbf{R}_{t+1}) - e_t\right)}{\mathbf{1}_N^{\top}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\mathbf{1}_N\right). \tag{7}$$

From (6) and (7) 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 ${\boldsymbol{\phi}_t^+}^{\top}\mathbf{1}_N = e_t$.*

[10] For details see Appendix F.1.

(ii) Suppose that Assumption 1 holds, then ϕ^*_{it} is affine in $\mathbf{y}_t$ and the strategy is a characteristics based strategy. 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 characteristics based.

(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). Hence, the optimal strategy is in general not a parametric strategy.

In Appendix A.3 we introduce linear and quadratic cost. In Appendix F.2 we observe that optimal strategies are path dependent and therefore not parametric as defined in the main text.

4.2 CARA Utility and Shrinkage

Let us now analyse the general case with or without short-selling and apply a shrinkage procedure. By maximizing expected utility (5), we get:

Observation 2. *(i) In-sample, good performance in terms of the certainty equivalent (see equation (18) presented later) and the Sharpe ratio for our empirical data set (see Section C) 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 D.2 in Appendix D.2).*

Figure 1 plots realized returns R_{pt+1} for $\boldsymbol{\phi}^*_t$, the parametric strategy $\boldsymbol{\phi}^\sharp_t$ and the $1/N$-strategy $\boldsymbol{\phi}^{1/N}_t$ [since $e_t = 1$, $\boldsymbol{\phi}^\sharp_t = \mathbf{w}^\sharp_t$ and $\boldsymbol{\phi}^{1/N}_t = \mathbf{w}^{1/N}_t$; $w^\sharp_{ft} = w^{1/N}_{ft} = 0$]. Note that the vertical axes have different scales and the variation of 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[11], while in terms of finance we add an object close to a quadratic cost term. Although also a cost function as described in Section A.3 can be used as some kind of punishment function, we want to exclude path-dependence [as discussed in Appendix F.2] and shrink $\boldsymbol{\phi}_t$ or the weights $\mathbf{w}_t$ towards some specific values $\breve{\boldsymbol{\phi}}_t$ or weights

[11] 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 ℓ_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.

$\breve{\mathbf{w}}_t$, respectively. Hence, we consider a positive definite $N \times N$ matrix $\mathbf{C}_t$ and the punishment term $-\frac{1}{2}\left(\boldsymbol{\phi}_t - \breve{\boldsymbol{\phi}}_t\right)^\top \mathbf{C}_t \left(\boldsymbol{\phi}_t - \breve{\boldsymbol{\phi}}_t\right)$. With $\breve{\boldsymbol{\phi}}_t = \breve{\mathbf{w}}_t = \mathbf{0}_{N\times 1}$ we shrink to zero, while with $\breve{\boldsymbol{\phi}}_t = e_t \frac{1}{N}\mathbf{1}_{N\times 1}$ shrink towards the $1/N$-portfolio. By using transformed expected utility (5), the (possible) short-selling constraints $\phi_{it} \geq 0$, and the shrinkage device, we get $b_{0t} := \rho e_t$, $\mathbf{b}_t := \rho\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right)^\top$, $\mathbf{B}_t := \frac{\rho^2}{2}\mathbb{V}_t(\mathbf{R}_{t+1} - \mathbf{1}_N R_{ft+1}) = \frac{\rho^2}{2}\mathbb{V}_t(\mathbf{R}_{t+1})$, and the Lagrangian

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

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

$$\begin{aligned}
\frac{\partial L(\boldsymbol{\phi}_t, \boldsymbol{\lambda}_t)}{\partial \boldsymbol{\phi}_t^\top} &= \mathbf{b}_t - \mathbf{B}_t\boldsymbol{\phi}_t + \boldsymbol{\lambda}_t - \mathbf{C}_t\left(\boldsymbol{\phi}_t - \breve{\boldsymbol{\phi}}_t\right) = \mathbf{0}_N \ , \qquad (8)\\
\frac{\partial L(\boldsymbol{\phi}_t, \boldsymbol{\lambda}_t)}{\partial \lambda_{it}} &= \phi_{it} = 0 \ , \ i = 1, \ldots, N \ , \text{ and the complementary slackness conditions}\\
0 &= \lambda_{it}\frac{\partial L(\boldsymbol{\phi}_t, \boldsymbol{\lambda}_t)}{\partial \lambda_{it}} = \lambda_{it}\phi_{it} \ , \ i = 1, \ldots, N \ . \qquad (9)
\end{aligned}$$

$\phi_{ft} = e_t - \sum_{i=1}^{N}\phi_{it}$ 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

$$\boldsymbol{\phi}_t^* = \left(\mathbf{B}_t + \mathbf{C}_t\right)^{-1}\left(\mathbf{b}_t + \mathbf{C}_t\breve{\boldsymbol{\phi}}_t\right) \ . \qquad (10)$$

Let $\mathbf{C}_t$ be equal to $c_p\mathbf{I}_N$, then (10) yields

$$\boldsymbol{\phi}_t^\flat := \frac{1}{\rho}\left(\mathbb{V}_t\left(\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_n\right)\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_n\right)^\top\right) + c_p\mathbf{I}_n\right)^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_n\right) + \rho c_p\breve{\boldsymbol{\phi}}_t\right) . \qquad (11)$$

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 problem without shrinkage (where $\boldsymbol{\phi}_t^\flat = \boldsymbol{\phi}_t^*$), while the larger c_p the more we shrink towards $\check{\boldsymbol{\phi}}_t = e_t \check{\mathbf{w}}_t$. To see this, for large c_p, the terms multiplied by c_p become the dominating terms. Hence, for large c_p, $\boldsymbol{\phi}_t^\flat \approx \frac{1}{\rho} \left(c_p \mathbf{I}_n\right)^{-1} \left(\rho c_p \check{\boldsymbol{\phi}}_t\right) = \check{\boldsymbol{\phi}}_t$. Summing up, we get

Proposition 1 (Asset Demand with CARA-Preferences)**.** *Suppose that the 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 (11) is a characteristics based strategy. 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 characteristics based.*

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 of the returns R_{pt} decreases a lot. In the case of binding short-selling constraints the system of inequalities (8) 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 $\phi_{it}^{\flat;\geq 0}$ described by (8). Here the Matlab function `fminsearch` is used, where we start the optimisation routine from $\max\{0, \phi_{it}^\flat\}$, $i = 1, \ldots, N$.

5 Constant Relative Risk Aversion

Let us now focus on constant relative risk aversion (CRRA). This case demonstrates potential pitfalls arising from parametric portfolio policies 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 (22) in the Appendix], expected utility is strictly concave in $\boldsymbol{\theta}$. 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 $\boldsymbol{\theta} \in \mathbb{R}^k$ exists in these cases [see equation (22) and Gehrig et al. (2018)]. Therefore, we obtain

Observation 3. *For an investor with CRRA preferences an optimal $\boldsymbol{\theta} \in \mathbb{R}^k$ solving the parametric portfolio optimization problem (21) 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 C]. 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 $\boldsymbol{\theta} \in \mathbb{R}^3$ 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 $\boldsymbol{\theta}$ (in absolute terms), one large coordinate of $\boldsymbol{\theta}$ turned out to be sufficient for negative gross returns. Hence, (i) and (ii) of Observation 6 [see Appendix A.1] become relevant in real world data. As demonstrated and discussed in more detail in Appendix E we extend the domain to the real line by applying the utility function

$$v_\flat\left(e_t R_{pt+1}\right) \quad := \quad \begin{cases} v\left(e_t R_{pt+1}\right) & \text{, for } R_{pt+1} \geq \psi_R, \\ \left(v\left(e_t \psi_R\right) - \delta v'\left(e_t \psi_R\right)\right) + \delta v'\left(e_t \psi_R\right) e_t R_{pt+1} & \text{, for } R_{pt+1} < \psi_R, \end{cases} \tag{12}$$

where $\psi_R > 0$. With (12) 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 $v_\flat\left(e_t \psi_R\right) = v\left(e_t \psi_R\right) = \left(v\left(e_t \psi_R\right) - \delta v'\left(e_t \psi_R\right)\right) + \delta v'\left(e_t \psi_R\right) e_t \psi_R$. For $\delta = 1$, we 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 (1), 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

$\mathbf{w}_t = (w_{1t}, \ldots, w_{nt})^\top = \mathbf{0}_N$ we obtain

$$
\begin{aligned}
\mathbb{E}_t\left(v_\flat(E_{t+1})\right) &= \mathbb{E}_t\left(v_\flat(e_t R_{pt+1})\right) = \mathbb{E}_t\left(v_\flat(e_t\left(R_{ft+1} + \mathbf{w}_t\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)\right))\right) \\
&\approx \underbrace{v_\flat\left(e_{t+1}R_{ft+1}\right)}_{=:\alpha_{0t}} + \underbrace{e_t v_\flat'(e_t R_{ft+1})\mathbb{E}_t\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)}_{=:\boldsymbol{\alpha}_t}\mathbf{w}_t \\
&\quad -\frac{1}{2}\mathbf{w}_t^\top \underbrace{\left(-v_\flat''\left(e_{t+1}R_{ft+1}\right)e_t^2\mathbb{E}_t\left(\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)^\top\right)\right)}_{:=\mathcal{A}_t}\mathbf{w}_t \ . \qquad (13)
\end{aligned}
$$

In the following optimization problem we also allow for short-selling constraints, that is $w_{it} \geq 0$. (see, also Koijen and Yogo, 2019, for a model with log-utility). Especially, the out-of-sample performance is very poor and the approximately optimal strategy $\mathbf{w}_t = \mathcal{A}_t^{-1}\boldsymbol{\alpha}_t$ is very risky. Hence, similar to Section 4 we proceed with a shrinkage device. We consider a positive definite $N \times N$ matrix $\mathbf{C}_t$ and the punishment 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 towards the $1/N$ portfolio.[12] By using the expected utility approximation (13), the short-selling constraints and the shrinkage device we get the Lagrangian

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

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

$$
\begin{aligned}
\frac{\partial L(\mathbf{w}_t, \boldsymbol{\lambda}_t)}{\partial \mathbf{w}_t^\top} &= \boldsymbol{\alpha}_t - \mathcal{A}_t\mathbf{w}_t + \boldsymbol{\lambda}_t - \mathbf{C}_t\left(\mathbf{w}_t - \check{\mathbf{w}}_t\right) = \mathbf{0}_N \ , \qquad (14) \\
\frac{\partial L(\mathbf{w}_t, \boldsymbol{\lambda}_t)}{\partial \lambda_{it}} &= w_{it} = 0 \ , \ i = 1, \ldots, N \ , \ \text{and the complementary slackness conditions} \\
0 &= \lambda_{it}\frac{\partial L(\mathbf{w}_t, \boldsymbol{\lambda}_t)}{\partial \lambda_{it}} = \lambda_{it}w_{it} \ , \ i = 1, \ldots, N \ . \qquad (15)
\end{aligned}
$$

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

[12] 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].

constraints we obtain $\boldsymbol{\lambda}_t = \mathbf{0}_N$ and

$$\begin{aligned} \mathbf{w}_t &= (\mathcal{A}_t + \mathbf{C}_t)^{-1} (\boldsymbol{\alpha}_t + \mathbf{C}_t \check{\mathbf{w}}_t) \\ &= (\mathcal{A}_t + \mathbf{C}_t)^{-1} \left(\boldsymbol{\alpha}_t + \frac{e_t R_{ft+1} v'_\flat(e_t R_{ft+1})}{e_t R_{ft+1} v'_\flat(e_t R_{ft+1})} \mathbf{C}_t \check{\mathbf{w}}_t \right) . \end{aligned} \tag{16}$$

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 Ψ_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 (16) result in

$$\begin{aligned} & \frac{1}{\gamma} \left(\mathbb{E}_t \left(\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right) \left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)^\top \right) + c_p \mathbf{I}_N \right)^{-1} \left(\mathbb{E}_t \left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right) + \frac{1}{e_t R_{ft+1} v'_\flat(e_t R_{ft+1})} \mathbf{C}_t \check{\mathbf{w}}_t \right) \\ = \; & \frac{1}{\gamma} \left(\mathbb{E}_t \left(\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right) \left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)^\top \right) + c_p \mathbf{I}_N \right)^{-1} \left(\mathbb{E}_t \left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right) + \frac{\left(-v''_\flat\left(e_{t+1}R_{ft+1}\right) e_t^2\right)}{e_t R_{ft+1} v'_\flat(e_t R_{ft+1})} c_p \mathbf{I}_N \check{\mathbf{w}}_t \right) \\ \approx \; & \frac{1}{\gamma} \Bigg(\underbrace{\mathbb{E}_t \left(\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right) \left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)^\top \right) + c_p \mathbf{I}_N}_{\mathcal{B}_t} \Bigg)^{-1} \left(\mathbb{E}_t \left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right) + \gamma c_p \check{\mathbf{w}}_t \right) =: \mathbf{w}_t^\flat . \end{aligned} \tag{17}$$

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 (17) is a characteristics based strategy.*

(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 characteristics based. (iii) If the term $\mathcal{B}_t$ is diagonal the weights only dependent on $\mathbf{y}_{it}$. A parametric strategy of the form described by (4) can be optimal, in which case $w_{it}^\flat = \bar{w}_{it} + \boldsymbol{\theta}^\top \mathbf{x}_{it}$ has to hold.

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 analyse 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 to 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 $P_{it}S_{it} = \sum_{j=1}^J \phi_{it}^{\flat\flat,j}$, $i = 1, \dots, N$. 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.[13] Hence, we get

Proposition 3 (Market Equilibrium). *Consider an economy with $J > 0$ investors. Each investor j is either a CARA or CRRA (in more detail, $v_\flat$ is applied) expected utility maximizer. Suppose that the 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 $\boldsymbol{\Sigma}_t$ 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.

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.

[13]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.

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:

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

Table 1: CARA Investment 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 $\frac{1}{N}$-estimator corresponds to a simple investment heuristic that abstracts from any information about second moments or any other characteristics. Appendix D.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 to 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 C). 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. 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 $k = 3$ macro-variables $\boldsymbol{\chi}_{it}$ 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 $\boldsymbol{\chi}_{it}$ 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 $t = 1, \ldots, 200$ are used to estimate the model parameters (training sample). For in-sample and out-of-sample comparisons, we use the observations $t = 1, \ldots, 200$ and $t = 201, \ldots, 415$, respectively.[14]

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 strategy.

We are particularly interested in the performance of the optimal investment strategy relative to the characteristics-based portfolio choice. For the exponential utility function $u(x) = -e^{\rho x}$ the certainty equivalent CI is the (smallest) value x where $\mathbb{E}\left(u(E_{t+1})\right) = u(x)$. We estimate the certainty equivalent by means of

$$\widehat{CI} = u^{-1}\left(\widehat{\mathbb{E}}\left(u(E_{t+1})\right)\right) = u^{-1}\left(\frac{1}{T_{\mathbb{J}}}\sum_{t\in\mathbb{T}_{\mathbb{J}}} u(E_{t+1})\right) , \tag{18}$$

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.[15] 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 the excess returns (when using the corresponding evaluation sample).

[14] It is insightful to compare both, simulated and empirical data; results with simulated data are provided in Appendix D.

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

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.[16]

These tables show estimates of the certainty equivalent $\widehat{CI}$ 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 $\widehat{CI}$, we observe that the optimal strategies $\boldsymbol{\phi}_t^\flat$ show the best performance. Due to short-selling constraints we obtained $\boldsymbol{\phi}_t^{\flat,\geq 0} < \boldsymbol{\phi}_t^\flat$. The difference caused by working with different estimation methods of the covariance matrix are small. The performance of the parametric strategy $\boldsymbol{\phi}_t^\sharp$ 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 $\boldsymbol{\phi}_t^{\sharp,\geq 0}$ is almost equal to the results with $1/N$-portfolio (our optimization routine is started with $\bar{w}_i = 1/N$ and $\boldsymbol{\theta} = \mathbf{0}$; differences in the numbers between $\boldsymbol{\phi}_t^{\sharp,\geq 0}$ and $\boldsymbol{\phi}_t^{1/N}$ are only observable when looking at further digits after the comma). Appendix D provides results for simulated data, where we observe that in the case the true parameter values are known the optimal strategies shows 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 are 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. Summing up, we get

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

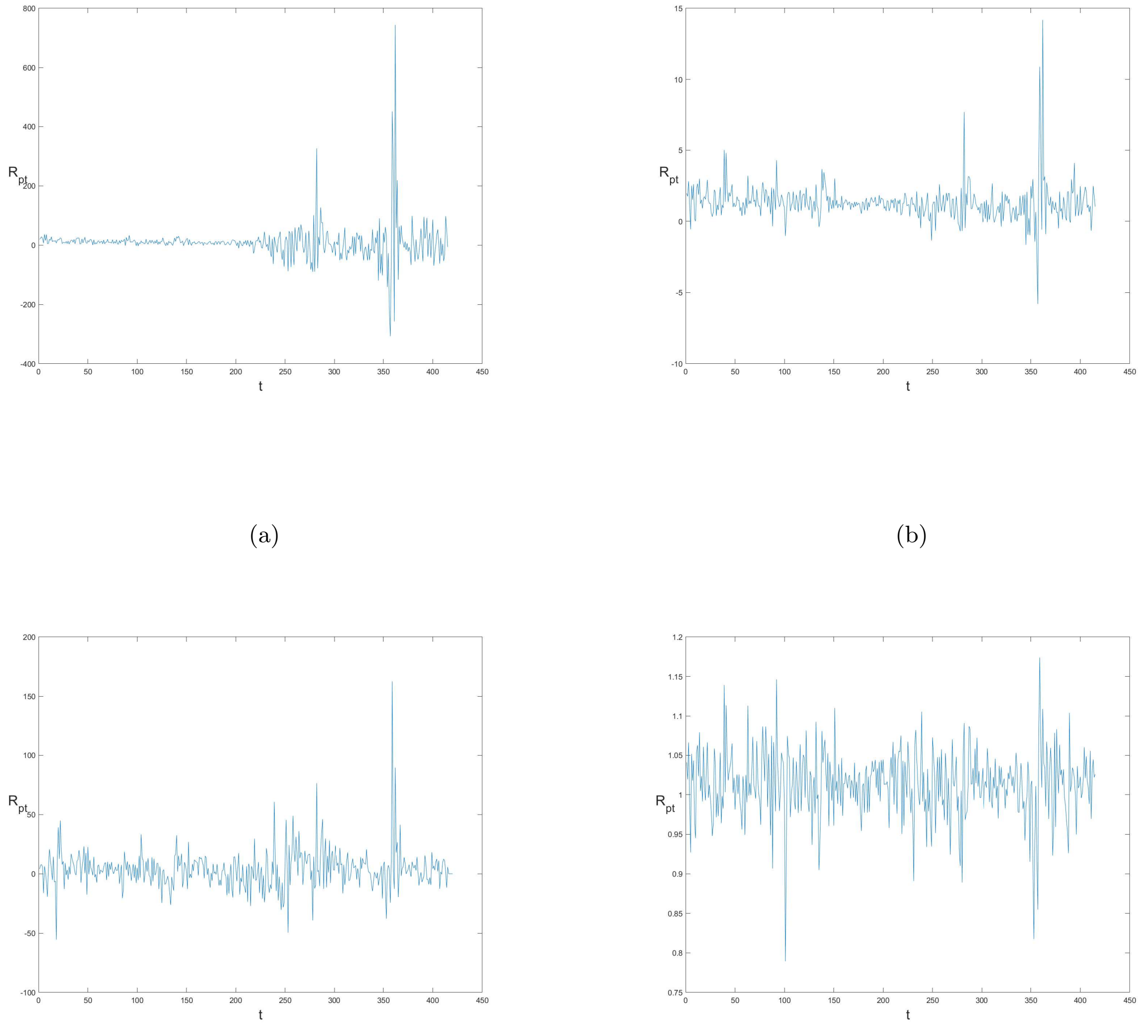
(ii) Out-of-sample: The optimal shrinkage strategies show the best performance followed by the $1/N$-

[16] We span the range of parameters that have been applied in different research environments in the experimental lab such as Goeree et al. (2002), Harrison and Rutström (2008), in the field experiments Tanaka et al. (2010) or in macroeconomic studies such as Hansen (1982).

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 $\frac{1}{N}$-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.



(a) (b)

(c) (d)

Figure 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 $\boldsymbol{\phi}_t^*$ [that is $\boldsymbol{\phi}_t^\flat$ with $c_p = 0$] where the Ledoit and Wolf (2004) estimator is used to obtain an estimate of the covariance matrix, That is, $\boldsymbol{\phi}_t^\flat$ obtained in (11) where the shrinkage parameter $c_p = 0$. Scaling of vertical axis $[-400, 800]$, (b) applies $\boldsymbol{\phi}_t^*$ 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 $[0.75, 1.2]$.

$\rho = 0.25$	$\boldsymbol{\phi}_t^{\flat}$	$\boldsymbol{\phi}_{t,LW}^{\flat}$	$\boldsymbol{\phi}_t^{\flat,\geq 0}$	$\boldsymbol{\phi}_{t,LW}^{\flat,\geq 0}$	$\boldsymbol{\phi}_t^{1/N}$	$\boldsymbol{\phi}_t^{\sharp}$	$\boldsymbol{\phi}_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.3997	1.4014	1.3137	1.3148	1.0153	-34.2976	1.0153
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\,(w_{it} < 0)$	0.0477	0.0479	0.0000	0.0000	0.0000	0.0290	0.0000
$mean\,(w_{it} < -1)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.4443	0.0000
$sd\,(w_{it} < -1)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0262	0.0000
$\rho = 0.5$	$\boldsymbol{\phi}_t^{\flat}$	$\boldsymbol{\phi}_{t,LW}^{\flat}$	$\boldsymbol{\phi}_t^{\flat,\geq 0}$	$\boldsymbol{\phi}_{t,LW}^{\flat,\geq 0}$	$\boldsymbol{\phi}_t^{1/N}$	$\boldsymbol{\phi}_t^{\sharp}$	$\boldsymbol{\phi}_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.2023	1.2031	1.1350	1.1410	1.0151	0.3416	1.0151
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\,(w_{it} < 0)$	0.0478	0.0479	0.0000	0.0000	0.0000	0.0291	0.0000
$mean\,(w_{it} < -1)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.2514	0.0000
$sd\,(w_{it} < -1)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0190	0.0000
$\rho = 1$	$\boldsymbol{\phi}_t^{\flat}$	$\boldsymbol{\phi}_{t,LW}^{\flat}$	$\boldsymbol{\phi}_t^{\flat,\geq 0}$	$\boldsymbol{\phi}_{t,LW}^{\flat,\geq 0}$	$\boldsymbol{\phi}_t^{1/N}$	$\boldsymbol{\phi}_t^{\sharp}$	$\boldsymbol{\phi}_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.1035	1.1039	1.0725	1.0702	1.0147	1.0016	1.0147
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\,(w_{it} < 0)$	0.0477	0.0476	0.0000	0.0000	0.0000	0.0339	0.0000
$mean\,(w_{it} < -1)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0325	0.0000
$sd\,(w_{it} < -1)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0071	0.0000
$\rho = 2$	$\boldsymbol{\phi}_t^{\flat}$	$\boldsymbol{\phi}_{t,LW}^{\flat}$	$\boldsymbol{\phi}_t^{\flat,\geq 0}$	$\boldsymbol{\phi}_{t,LW}^{\flat,\geq 0}$	$\boldsymbol{\phi}_t^{1/N}$	$\boldsymbol{\phi}_t^{\sharp}$	$\boldsymbol{\phi}_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.0539	1.0541	1.0342	1.0350	1.0138	1.0213	1.0138
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\,(w_{it} < 0)$	0.0442	0.0445	0.0000	0.0000	0.0000	0.0378	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
$\rho = 5$	$\boldsymbol{\phi}_t^{\flat}$	$\boldsymbol{\phi}_{t,LW}^{\flat}$	$\boldsymbol{\phi}_t^{\flat,\geq 0}$	$\boldsymbol{\phi}_{t,LW}^{\flat,\geq 0}$	$\boldsymbol{\phi}_t^{1/N}$	$\boldsymbol{\phi}_t^{\sharp}$	$\boldsymbol{\phi}_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.0234	1.0235	1.0149	1.0143	1.0109	1.0036	1.0109
$\widehat{sd}\left(\widehat{CI}\right)$	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

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 $c_p = 0.2$.

$\rho = 0.25$	$\phi_t^\flat$	$\phi_{t,LW}^\flat$	$\phi_t^{\flat,\geq 0}$	$\phi_{t,LW}^{\flat,\geq 0}$	$\phi_t^{1/N}$	$\phi_t^\sharp$	$\phi_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.1418	1.1376	0.9996	0.9925	1.0107	-28.7653	1.0107
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\,(w_{it} < 0)$	0.0474	0.0482	0.0000	0.0000	0.0000	0.0280	0.0000
$mean\,(w_{it} < -1)$	0.0268	0.0268	0.0000	0.0000	0.0000	0.4232	0.0000
$sd\,(w_{it} < -1)$	0.0174	0.0172	0.0000	0.0000	0.0000	0.0271	0.0000
$\rho = 0.5$	$\phi_t^\flat$	$\phi_{t,LW}^\flat$	$\phi_t^{\flat,\geq 0}$	$\phi_{t,LW}^{\flat,\geq 0}$	$\phi_t^{1/N}$	$\phi_t^\sharp$	$\phi_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.0706	1.0683	1.0125	1.0332	1.0104	0.0316	1.0104
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\,(w_{it} < 0)$	0.0477	0.0478	0.0000	0.0000	0.0000	0.0276	0.0000
$mean\,(w_{it} < -1)$	0.0029	0.0030	0.0000	0.0000	0.0000	0.2669	0.0000
$sd\,(w_{it} < -1)$	0.0046	0.0046	0.0000	0.0000	0.0000	0.0220	0.0000
$\rho = 1$	$\phi_t^\flat$	$\phi_{t,LW}^\flat$	$\phi_t^{\flat,\geq 0}$	$\phi_{t,LW}^{\flat,\geq 0}$	$\phi_t^{1/N}$	$\phi_t^\sharp$	$\phi_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.0348	1.0336	1.0165	1.0274	1.0099	0.9376	1.0099
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\,(w_{it} < 0)$	0.0479	0.0470	0.0000	0.0000	0.0000	0.0198	0.0000
$mean\,(w_{it} < -1)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0349	0.0000
$sd\,(w_{it} < -1)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0091	0.0000
$\rho = 2$	$\phi_t^\flat$	$\phi_{t,LW}^\flat$	$\phi_t^{\flat,\geq 0}$	$\phi_{t,LW}^{\flat,\geq 0}$	$\phi_t^{1/N}$	$\phi_t^\sharp$	$\phi_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.0167	1.0160	1.0087	1.0099	1.0089	1.0027	1.0089
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\,(w_{it} < 0)$	0.0476	0.0470	0.0000	0.0000	0.0000	0.0227	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
$\rho = 5$	$\phi_t^\flat$	$\phi_{t,LW}^\flat$	$\phi_t^{\flat,\geq 0}$	$\phi_{t,LW}^{\flat,\geq 0}$	$\phi_t^{1/N}$	$\phi_t^\sharp$	$\phi_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.0049	1.0045	1.0039	1.0040	1.0057	0.9933	1.0057
$\widehat{sd}\left(\widehat{CI}\right)$	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

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

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 towards $\check{\mathbf{w}}_t$. To implement (17), $\mathbb{E}_t\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)$ and $\mathbb{E}_t\left(\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)\left(\mathbf{R}_{t+1} - R_{ft+1}\mathbf{1}_N\right)^\top\right)$ 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 (18).

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 (17) results in poor performance. This problem can be expected, since the weights obtained in (17) 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 γ. The smaller γ, the larger the weights $\mathbf{w}_t^\flat$ in absolute value for any $c_p \geq 0$. Note that in contrast to the CARA case, our solutions for the CRRA case (with or without short-selling constraints) are based on the Taylor series approximation (13) around $\mathbf{w} = \mathbf{0}_{N\times 1}$. If the weights obtained $w_{it}^\flat$ 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.

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 strategy
$\mathbf{w}_t^{\sharp,\geq 0}$	parametric portfolio strategy without short-selling

Table 4: CRRA Investment Strategies

In the following Tables 5 and 6 we observe that for small γ the weights following from (17) 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 $\boldsymbol{\theta}$ in the case of parametric portfolio policies, no approximation of the expected utility function is used. Hence, also for small γ 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 (17), 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 slighly dominates the $1/N$ and the parametric strategy. 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 (17) in terms of the points estimate of the certainty equivalent dominate the other strategies. 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 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 γ the best performance with the parametric strategy is observed. For $\gamma \geq 1$ the best performance is achieved with the optimal 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.*

$\gamma = 0.25$	$\mathbf{w}_t^{\flat}$	$\mathbf{w}_{t,LW}^{\flat}$	$\mathbf{w}_t^{\flat,\geq 0}$	$\mathbf{w}_{t,LW}^{\flat,\geq 0}$	$\mathbf{w}_t^{1/N}$	$\mathbf{w}_t^{\sharp}$	$\mathbf{w}_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.1656	1.1668	1.0340	1.0227	1.0153	1.0519	1.0153
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\left(w_{it} < 0\right)$	0.0478	0.0479	0.0000	0.0297	0.0000	0.0297	0.0000
$mean\left(w_{it} < -1\right)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$sd\left(w_{it} < -1\right)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\gamma = 0.5$	$\mathbf{w}_t^{\flat}$	$\mathbf{w}_{t,LW}^{\flat}$	$\mathbf{w}_t^{\flat,\geq 0}$	$\mathbf{w}_{t,LW}^{\flat,\geq 0}$	$\mathbf{w}_t^{1/N}$	$\mathbf{w}_t^{\sharp}$	$\mathbf{w}_t^{\sharp,\geq 0}$
$\widehat{CI}$	0.6065	0.6131	1.0157	1.0320	1.0151	1.0434	1.0151
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\left(w_{it} < 0\right)$	0.0478	0.0481	0.0000	0.0000	0.0000	0.0297	0.0000
$mean\left(w_{it} < -1\right)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$sd\left(w_{it} < -1\right)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\gamma = 1$	$\mathbf{w}_t^{\flat}$	$\mathbf{w}_{t,LW}^{\flat}$	$\mathbf{w}_t^{\flat,\geq 0}$	$\mathbf{w}_{t,LW}^{\flat,\geq 0}$	$\mathbf{w}_t^{1/N}$	$\mathbf{w}_t^{\sharp}$	$\mathbf{w}_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.0934	1.0937	1.0130	1.0287	1.0147	1.0341	1.0147
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\left(w_{it} < 0\right)$	0.0476	0.0466	0.0000	0.0000	0.0000	0.0351	0.0000
$mean\left(w_{it} < -1\right)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$sd\left(w_{it} < -1\right)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\gamma = 2$	$\mathbf{w}_t^{\flat}$	$\mathbf{w}_{t,LW}^{\flat}$	$\mathbf{w}_t^{\flat,\geq 0}$	$\mathbf{w}_{t,LW}^{\flat,\geq 0}$	$\mathbf{w}_t^{1/N}$	$\mathbf{w}_t^{\sharp}$	$\mathbf{w}_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.0490	1.0491	1.0082	1.0216	1.0138	1.0233	1.0138
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\left(w_{it} < 0\right)$	0.0428	0.0438	0.0000	0.0000	0.0000	0.0334	0.0000
$mean\left(w_{it} < -1\right)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$sd\left(w_{it} < -1\right)$	0.0000	0.0000	0.0000	0.0000	0.0022	0.0000	0.0000
$\gamma = 5$	$\mathbf{w}_t^{\flat}$	$\mathbf{w}_{t,LW}^{\flat}$	$\mathbf{w}_t^{\flat,\geq 0}$	$\mathbf{w}_{t,LW}^{\flat,\geq 0}$	$\mathbf{w}_t^{1/N}$	$\mathbf{w}_t^{\sharp}$	$\mathbf{w}_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.0214	1.0214	1.0056	1.0109	1.0108	1.0031	1.0108
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\left(w_{it} < 0\right)$	0.0346	0.0341	0.0000	0.0000	0.0000	0.0454	0.0000
$mean\left(w_{it} < -1\right)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$sd\left(w_{it} < -1\right)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

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 $c_p = 0.2$.

8 Conclusions

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 we have augmented this 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 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 characteristics-based parametric portfolio strategies can be optimal under rather strong assumptions.

Moreover, theory-guided reduction techniques prove particularly helpful for machine learning applications as forcefully argued by Nagel (2021).[17] In this article we introduce a shrinkage facility, with the goal to make the strategies less risky and thereby improve their performance in empirical data. We provide empirical evidence for S&P 500 data.[18] For the feasible case where parameters are estimated, 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.

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, tuning of the shrinkage parameter, 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.

[17] 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].

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

$\gamma = 0.25$	$\mathbf{w}_t^{\flat}$	$\mathbf{w}_{t,LW}^{\flat}$	$\mathbf{w}_t^{\flat,\geq 0}$	$\mathbf{w}_{t,LW}^{\flat,\geq 0}$	$\mathbf{w}_t^{1/N}$	$\mathbf{w}_t^{\sharp}$	$\mathbf{w}_t^{\sharp,\geq 0}$
$\widehat{CI}$	0.7703	0.7563	1.0177	1.0127	1.0107	1.0582	1.0107
$\widehat{sd}\left(\widehat{CI}\right)$	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
$\gamma = 0.5$	$\mathbf{w}_t^{\flat}$	$\mathbf{w}_{t,LW}^{\flat}$	$\mathbf{w}_t^{\flat,\geq 0}$	$\mathbf{w}_{t,LW}^{\flat,\geq 0}$	$\mathbf{w}_t^{1/N}$	$\mathbf{w}_t^{\sharp}$	$\mathbf{w}_t^{\sharp,\geq 0}$
$\widehat{CI}$	0.3829	0.3223	1.0112	1.0165	1.0104	0.9783	1.0104
$\widehat{sd}\left(\widehat{CI}\right)$	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
$\gamma = 1$	$\mathbf{w}_t^{\flat}$	$\mathbf{w}_{t,LW}^{\flat}$	$\mathbf{w}_t^{\flat,\geq 0}$	$\mathbf{w}_{t,LW}^{\flat,\geq 0}$	$\mathbf{w}_t^{1/N}$	$\mathbf{w}_t^{\sharp}$	$\mathbf{w}_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.0365	1.0363	1.0068	1.0151	1.0099	1.0035	1.0099
$\widehat{sd}\left(\widehat{CI}\right)$	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
$\gamma = 2$	$\mathbf{w}_t^{\flat}$	$\mathbf{w}_{t,LW}^{\flat}$	$\mathbf{w}_t^{\flat,\geq 0}$	$\mathbf{w}_{t,LW}^{\flat,\geq 0}$	$\mathbf{w}_t^{1/N}$	$\mathbf{w}_t^{\sharp}$	$\mathbf{w}_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.0194	1.0192	1.0062	1.0122	1.0089	1.0033	1.0089
$\widehat{sd}\left(\widehat{CI}\right)$	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
$sd\,(w_{it} < -1)$	0.0000	0.0000	0.0000	0.0000	0.0073	0.0000	0.0000
$\gamma = 5$	$\mathbf{w}_t^{\flat}$	$\mathbf{w}_{t,LW}^{\flat}$	$\mathbf{w}_t^{\flat,\geq 0}$	$\mathbf{w}_{t,LW}^{\flat,\geq 0}$	$\mathbf{w}_t^{1/N}$	$\mathbf{w}_t^{\sharp}$	$\mathbf{w}_t^{\sharp,\geq 0}$
$\widehat{CI}$	1.0081	1.0080	1.0032	1.0068	1.0057	0.9921	1.0057
$\widehat{sd}\left(\widehat{CI}\right)$	0.0169	0.0171	0.0033	0.0103	0.0137	0.0292	0.0137
$mean(E_t)$	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
$mean(w_{it} < 0)$	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
$mean\,(w_{it} < -1)$	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

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 $c_p = 0.2$.

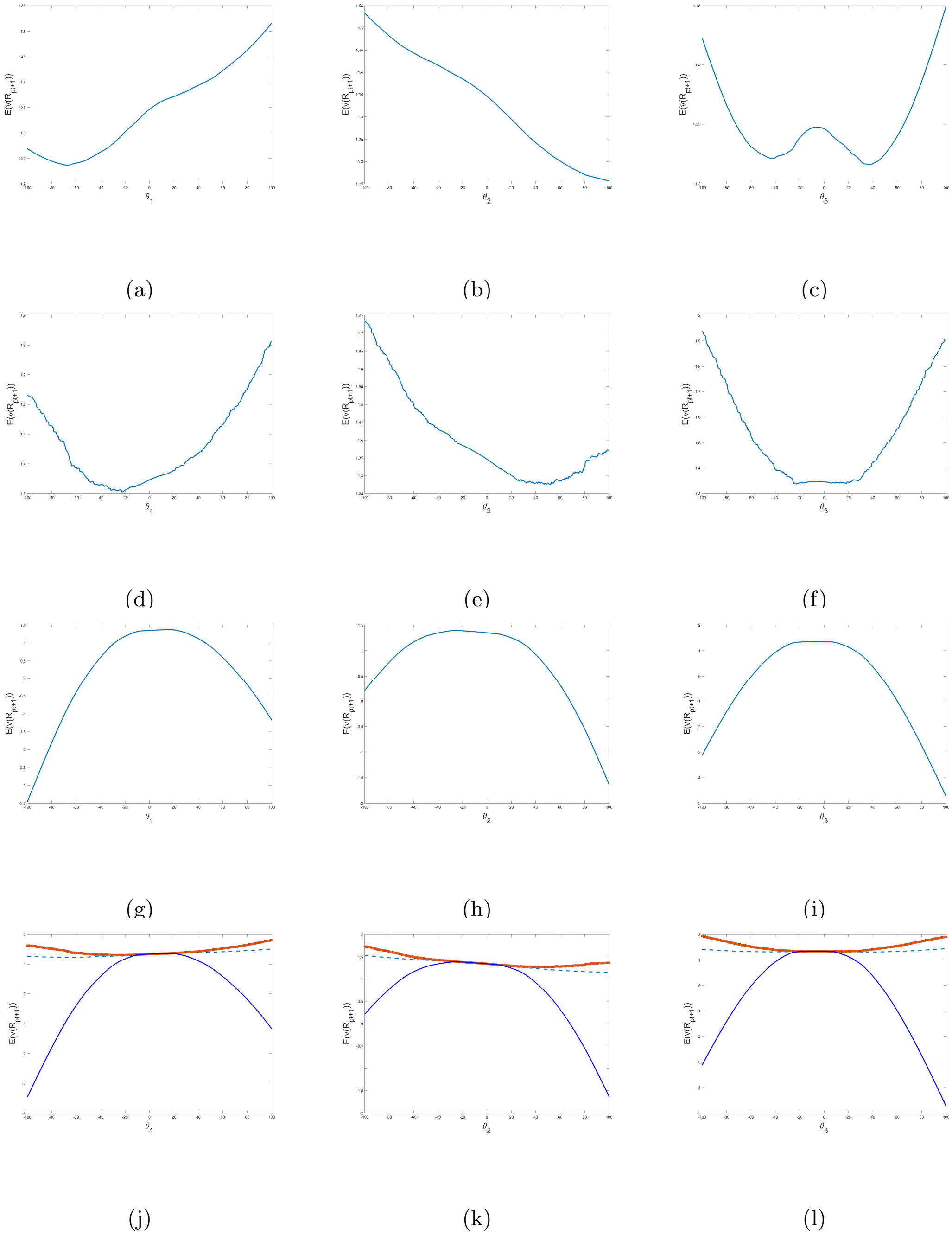
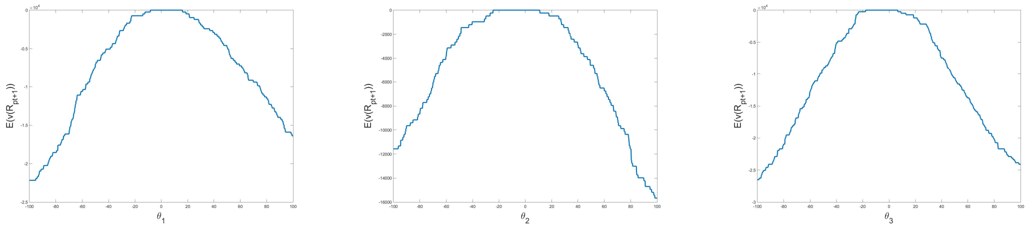
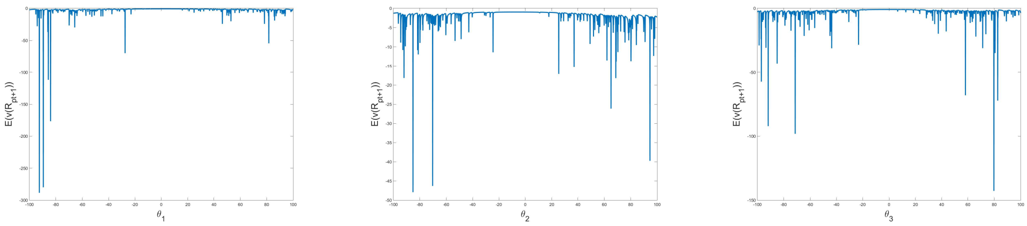


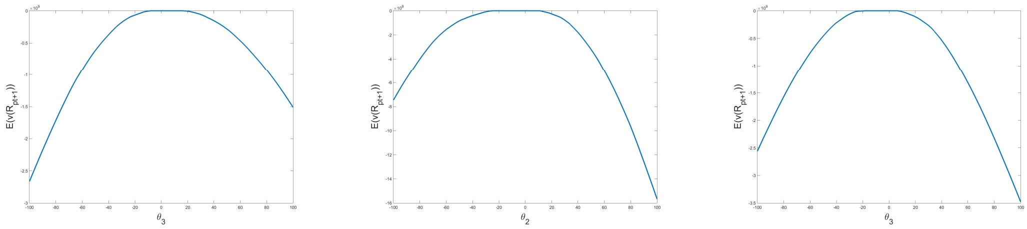
Figure 2: Expected utility $\mathbb{E}\left(v\left(R_{pt+1}\right)\right)$ against $\theta_1 \in [-500, 500]$, $\theta_2 = 0$ and $\theta_3 = 0$ in the first column; $\theta_2 \in [-500, 500]$, $\theta_1 = 0$ and $\theta_3 = 0$ in the second column; and $\theta_3 \in [-500, 500]$, $\theta_1 = 0$ and $\theta_2 = 0$ in the third column. CRRA type utility with $\gamma = 0.25$. Grid with step-width 0.25. $g(\boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}) = \boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}$, Subfigures (a-c) augmented CRRA utility $v_+(\cdot)$ with parameter γ, $\psi_R = 0.0001$ and $\underline{u} = v(\psi_R)$. Subfigures (d-f) augmented CRRA utility $v_-(\cdot)$, Subfigures (g-i) augmented CRRA utility $v_\flat(\cdot)$, with $\delta = 1$. Subfigures (j-l) summary, where the dotted lines denotes $v_+(\cdot)$, the solid red lines $v_-(\cdot)$ and the solid blue lines $v_\flat(\cdot)$. No trading cost. S&P 500 data; $T = 416$ and $N = 100$.



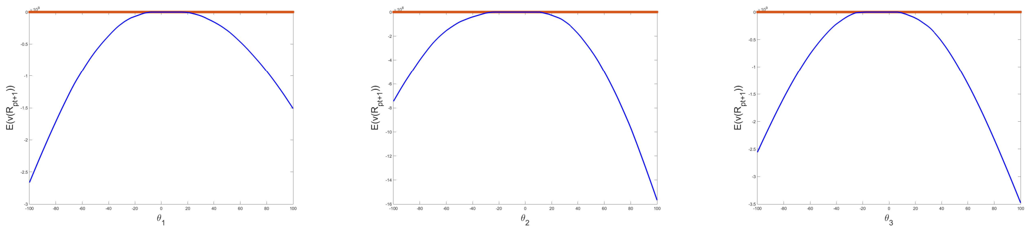
(a) (b) (c)



(d) (e) (f)



(g) (h) (i)



(j) (k) (l)

Figure 3: Expected utility $\mathbb{E}\left(v\left(R_{pt+1}\right)\right)$ against $\theta_1 \in [-500, 500]$, $\theta_2 = 0$ and $\theta_3 = 0$ in the first column; $\theta_2 \in [-500, 500]$, $\theta_1 = 0$ and $\theta_3 = 0$ in the second column; and $\theta_3 \in [-500, 500]$, $\theta_1 = 0$ and $\theta_2 = 0$ in the third column. CRRA type utility with $\gamma = 2$. Grid with step-width 0.25. $g(\boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}) = \boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}$, Subfigures (a-c) augmented CRRA utility $v_+(\cdot)$ with parameter γ, $\psi_R = 0.0001$ and $\underline{u} = v(\psi_R)$. Subfigures (d-f) augmented CRRA utility $v_-(\cdot)$, Subfigures (g-i) augmented CRRA utility $v_\flat(\cdot)$, with $\delta = 1$. Subfigures (j-l) summary, where the dotted lines denotes $v_+(\cdot)$, the solid red lines $v_-(\cdot)$ and the solid blue lines $v_\flat(\cdot)$. No trading cost. S&P-500 data; $T = 416$ and $N = 100$.

References

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

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

Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies*, 22(9):3411–3447.

Campbell, J. (2017). *Financial Decisions and Markets: A Course in Asset Pricing*. Princeton University Press.

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

Davidson, R. and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press, New York.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953.

Erlwein-Sayer, C., Grimm, S., Ruckdeschel, P., Sass, J., and Sayer, T. (2020). Filter-based portfolio strategies in an hmm setting with varying correlation parametrizations. *Applied Stochastic Models in Business and Industry*, 36(3):307–334.

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

Freyberger, J., Neuhierl, A., and Weber, M. (2020). Dissecting Characteristics Nonparametrically. *The Review of Financial Studies*, 33(5):2326–2377.

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

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

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

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

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

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer series in statistics. Springer.

He, Z. and Krishnamurthy, A. (2013). Intermediary asset pricing. *American Economic Review*, 103:732–757.

Hjalmarsson, E. and Manchev, P. (2012). Characteristic-based mean-variance portfolio choice. *Journal of Banking & Finance*, 36(5):1392–1401.

Hsiao, C. (2015). *Analysis of Panel Data.* Cambridge University Press, Econometric Society Monographs No. 54, Cambridge, 3 edition.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2017). *An Introduction to Statistical Learning – with Applications in R*, volume 103 of *Springer Texts in Statistics.* Springer, New York.

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

Klenke, A. (2008). *Probability Theory - A Comprehensive Course.* Springer.

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

Ledoit, O. and Wolf (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *Review of Financial Studies, forthcomming.*

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

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

Medeiros, M. C., Passos, A. M., and Vasconcelos, G. F. R. (2014). Parametric Portfolio Selection: Evaluating and Comparing to Markowitz Portfolios. *Brazilian Review of Finance*, 12(2):257–284.

Munkres, J. (2000). *Topology.* Prentice Hall, Upper Saddle River, NJ, 2nd edition.

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

Ruud, P. A. (2000). *An Introduction to Classical Econometric Theory.* Oxford University Press, New York.

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

Tanaka, T., Camerer, C. F., and Nguyen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from vietnam. *American Economic Review*, 100(1):557–71.

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

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

A A Micro-Foundation of Characteristics-Based Portfolio Choice

As already stated in the main text, we consider a myopic investor who maximizes conditionally expected utility with Bernoulli utility function $u : \mathbb{D} \to \mathbb{R}$. To simplify the notation $N_t = N$. We assume that $u(\cdot)$ is strictly increasing. To reduce the mathematical burden $u(\cdot)$ is at least twice continuously differentiable. This results in a constrained portfolio optimization problem

$$\begin{aligned} &\max_{\mathbf{w}_t, w_{ft} \in \mathbb{R}^{n_t}} \mathbb{E}_t \left(u \left(E_{t+1} \right) \right) \\ &s.t. \;\; E_{t+1} = e_t \mathbf{w}_t^\top \mathbf{R}_{t+1} + e_t w_{ft} R_{fT+1} \quad , \quad \sum_{i=(N-n)+1}^{N} w_{it} = 1 \; . \end{aligned} \tag{19}$$

Optimization problem (19) results in the optimal investment weights w_{it}^*, $i = (N-n)+1, \ldots, N$ (for w_{ft}^* the summation index 0 is used if a risk-free asset is considered).

Suppose that $\mathbb{E}_t \left(u \left(E_{t+1} \right) \right)$ exists (for all $\mathbf{w}_t \in \mathbb{R}^N$) and that differentiation and integration can be exchanged. The constraint optimization problem (19) results in the Lagrangian $\mathscr{L} \left(\mathbf{w}_t, \mu_t \right) = \mathbb{E}_t \left(u \left(E_{t+1} \right) \right) - \mu_t \left(\sum_{i=1}^{N} w_{it} - 1 \right)$. By taking partial derivatives with respect to $\mathbf{w}_t$ and μ_t we obtain the first order conditions

$$\begin{aligned} \frac{\partial}{\partial \mathbf{w}_t^\top} \mathscr{L} \left(\mathbf{w}_t, \mu_t \right) &= \mathbb{E}_t \left(u' \left(e_t \mathbf{w}_t^\top \mathbf{R}_{t+1} \right) \mathbf{R}_{t+1} \right) - \mu_t \mathbf{1}_{N \times 1} = \mathbf{0}_{N \times 1} \quad \text{and} \\ \frac{\partial}{\partial \mu_t} \mathscr{L} \left(\mathbf{w}_t, \mathbf{z}_t, \mu_t \right) &= 1 - \sum_{i=(N-n)+1}^{N} w_{it} = 0 \; . \end{aligned} \tag{20}$$

The second order condition for a constraint maximization problem is e.g. discussed in Simon and Blume (1994)[Chapter 19.3]. A negative definite Hessian $\frac{\partial}{\partial \mathbf{w}_t \partial \mathbf{w}_t^\top} \mathbb{E}_t \left(u'' \left(e_t R_{pt+1} \right) \right) = \mathbb{E}_t \left(u'' \left(e_t \mathbf{w}_t^\top \mathbf{R}_{t+1} \right) \mathbf{R}_{t+1} \mathbf{R}_{t+1}^\top \right)$, for any $\mathbf{w}_t$ is sufficient to satisfy the second order condition. If a global optimum exists, we abbreviate the optimal $N+1$-vector by $\left(\mathbf{w}_t^{*\top}, \mu_t^* \right)^\top$.

By contrast a characteristics-based policy, or parametric portfolio policy solves[19]

$$\max_{\boldsymbol{\theta}\in\mathbb{R}^k} \mathbb{E}\left(u\left(e_t\sum_{i=1}^{N}\left(\bar{w}_{it}+\boldsymbol{\theta}^\top\tilde{\mathbf{x}}_t\right)^\top \mathbf{R}_{t+1}\right)\right) , \tag{21}$$

where the first and the second order conditions are

$$\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\theta}^\top}\mathbb{E}_t\left(u\left(e_t\sum_{i=1}^{N}\left(\frac{1}{N}+\frac{1}{N}\boldsymbol{\theta}^\top\tilde{\mathbf{x}}_t\right)^\top \mathbf{R}_{t+1}\right)\right) &= \mathbb{E}_t\left(u'\left(E_{t+1}\right)\frac{1}{N}\sum_{i=1}^{N}R_{it+1}\tilde{\mathbf{x}}_{it}\right)=\mathbf{0}_{k\times 1} \text{ and}\\
\frac{\partial}{\partial\boldsymbol{\theta}^\top\partial\boldsymbol{\theta}}\mathbb{E}_t\left(u\left(e_t\sum_{i=1}^{N}\left(\frac{1}{N}+\boldsymbol{\theta}^\top\tilde{\mathbf{x}}_t\right)^\top \mathbf{R}_{t+1}\right)\right) &= \mathbb{E}_t\left(u''\left(E_{t+1}\right)\left(\frac{1}{N}\sum_{i=1}^{N}R_{it+1}\tilde{\mathbf{x}}_{it}\right)\left(\frac{1}{N}\sum_{i=1}^{N}R_{it+1}\tilde{\mathbf{x}}_{it}\right)^\top\right).
\end{aligned} \tag{22}$$

In the case an optimal $\boldsymbol{\theta}$, denoted $\boldsymbol{\theta}^*$, exists, the optimal parametric policy is provided by $w_{it}^{\sharp}(\tilde{\mathbf{x}}_{it},e_t) := \bar{w}_{it}+\left(\boldsymbol{\theta}^*\right)^\top\tilde{\mathbf{x}}_{it}$. [20]

By contrast for a characteristics-based policy, or parametric portfolio policy, the strategy is restricted to the affine rule $w_{it}=\bar{w}_{it}+\boldsymbol{\theta}^\top\mathbf{x}_t$. By including this constraint in the optimization problem (19) we obtain $w_{it}^{\sharp}(\tilde{\mathbf{x}}_{it},e_t) := \bar{w}_{it}+\left(\boldsymbol{\theta}^*\right)^\top\tilde{\mathbf{x}}_{it}$. Typically, the focus of a parametric portfolio policy is on risky assets only. Hence, $w_{ft}^*=0$ for all t, if a the risk-free asset is traded.

Now suppose that both w_{it}^* and $w_{it}^{\sharp}$ exist. How are they related to each other? One way of response is a regression type approach (see also Brandt et al., 2009). That is to say, consider the panel regression model

$$\underbrace{\begin{pmatrix} w_{1t}^* \\ \vdots \\ w_{Nt}^* \end{pmatrix}}_{\mathbf{w}_t^*} - \underbrace{\begin{pmatrix} \bar{w}_{1t} \\ \vdots \\ \bar{w}_{Nt} \end{pmatrix}}_{\bar{\mathbf{w}}_t} = \begin{pmatrix} \tilde{\mathbf{x}}_{1t}^\top \\ \vdots \\ \tilde{\mathbf{x}}_{Nt}^\top \end{pmatrix}\boldsymbol{\theta} + \begin{pmatrix} u_{1t} \\ \vdots \\ u_{Nt} \end{pmatrix} . \tag{23}$$

By using for example the least squares dummy variable estimator (see, e.g., Hsiao, 2015, Chapter 3), we get $\widehat{\boldsymbol{\theta}} = \left(\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}^\top\right)\right)^{-1}\left(\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\tilde{\mathbf{x}}_{it}\left(w_{it}^*-\bar{w}_{it}\right)\right)\right)$. By means of $\widehat{\boldsymbol{\theta}}$ the sum of squared approximation errors is minimized. Note that, for large T a law of large numbers yields $\lim_{T\to\infty}\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}=$

[19] An alternative to the approach presented here is to fix $\boldsymbol{\theta}$ a-priori and choose the weights $\bar{w}_{it}$, $i=1,\ldots,N$, optimally. For the maximization of the conditional expected see Appendix F.2.

[20] An alternative to the approach presented here is to fix $\boldsymbol{\theta}$ a-priori and choose the weights $\bar{w}_{it}$, $i=1,\ldots,N$, optimally. For the maximization of the conditional expected see Appendix F.2.

$\left(\mathbb{E}\left(\sum_{i=1}^{N}\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}^{\top}\right)\right)^{-1}\mathbb{E}\left(\sum_{i=1}^{N}\tilde{\mathbf{x}}_{it}u_{it}\right)$. The right hand side term need not be zero. To see this, consider e.g. a w_{it}^{*} depending on $\mathbf{x}_t \in \mathbb{R}^{kN}$ [see e.g. the CARA case in Section 4]. In this case u_{it} also contains $\mathbf{x}_{jt}$, $i \neq j$, where $\mathbf{x}_{jt}$ and $\mathbf{x}_{it}$ need not be uncorrelated (in terms of econometrics, we are confronted with an *omitted variables* problem). Since the main objective is to approximate w_{it}^{*} and not to perform inference about $\boldsymbol{\theta}$, this issue is of minor importance in the case considered here.

A.1 Some General Results on Parametric Portfolio Policies

Here we summarize our observations for the case of the affine rule:

Observation 6. Consider returns $r_{it+1} \in \mathbb{R}_{\geq -1}$ and the portfolio gross-return $R_{pt+1} = \sum_{i\in\mathbb{I}_t} w_{it}\left(1 + r_{it+1}\right)$. Let $u(\cdot)$ denote a strictly monotone increasing Bernoulli utility function defined on $\mathbb{D} = \mathbb{R}_{>0}$.

(a) Suppose that for at least one j the support of $\frac{1}{N_t}\sum_{i\in\mathbb{I}_t}\tilde{x}_{it,j}r_{it+1}$ is te real line. Then, for any fixed $\boldsymbol{\theta}$ with $\theta_j \neq 0$, the probability $\mathbb{P}\left(R_{pt+1} < 0\right)$ is strictly positive.

(b) Suppose that the joint probability

$$\mathbb{P}\left(\{|\frac{1}{N_t}\sum_{i\in\mathbb{I}_t}\tilde{x}_{it,j}r_{it+1}| > \varepsilon_{xr}\} \wedge \{r_{pt+1} \leq \varepsilon_r\}\right) = \delta_r > 0$$

for some $j \in \{1, \ldots, k\}$, where ε_{xr} and $\varepsilon_r > 0$. There exists $\boldsymbol{\theta} \in \mathbb{R}^k$ such that $\mathbb{P}\left(R_{pt+1} < 0\right) > 0$. Hence, for $\Theta = \mathbb{R}^k$ there exists $\boldsymbol{\theta} \in \Theta$ where expected utility $\mathbb{E}\left(u\left(R_{pt+1}\right)\right)$ is not well-defined.

(c) Bounded Support of $\tilde{\mathbf{x}}$ and r_{it+1}: Suppose that $0 < \underline{r} \leq r_{it+1} \leq \bar{r} < \infty$ and $\underline{x} \leq \tilde{\mathbf{x}}_{it,j} \leq \bar{x}$, for all i, t and j. The cost function satisfies $0 \leq c(\boldsymbol{\phi}, \zeta) \leq \bar{c} < \infty$. If $1 + \underline{r} - \bar{c} > \alpha$, there exists a set of parameters Θ_R' of positive Lebesgue measure containing $\mathbf{0}_{(k\times 1)}$ where $R_{pt+1} > 0$ for all $\boldsymbol{\theta} \in \Theta_R'$ (almost surely) and expected utility $\mathbb{E}\left(u\left(R_{pt+1}\right)\right)$ is well-defined.

(d) Non-negative investment weights: Suppose that $f_i(N_t, \tilde{\mathbf{x}}_{it}; \boldsymbol{\theta})$ results in $w_{it} \geq 0$ for all i and t . Let $r_{it+1} \in \mathbb{R}_{>-1}$. Then, at least one $w_{it} > 0$ by the constraint $\sum_{i\in\mathbb{I}_t} w_{it} = 1$ and the portfolio gross-return $R_{pt+1} = \sum_{i\in\mathbb{I}_t} w_{it}\left(1 + r_{it+1}\right)$ is strictly positive.

(e) Since a constraint is added to (19) in the case of parametric portfolio policies

$$\mathbb{E}_t\left(u\left(e_t\left(1+\sum_{i\in\mathbb{I}_t} w_{it}^{*} r_{it+1}\right)\right)\right) \geq \mathbb{E}_t\left(u\left(e_t\left(1+\sum_{i\in\mathbb{I}_t} w_{it}^{\sharp} r_{it+1}\right)\right)\right) .$$

This result also holds if (19) is augmented by trading cost and the investments in period $t-1$ are fixed at the same levels.[21] In the case of a strictly concave optimization problem, in general a strict inequality holds, while an equality can be maintained if the first order conditions arsing from (19) are of the form $-\mathbf{A}\mathbf{w}_t + \mathbf{B}\tilde{\mathbf{x}}_t + \mathbf{c}_t = \mathbf{0}_N$ for the N risk assets, where the following equalities hold: $\bar{\mathbf{w}}_t = \mathbf{A}^{-1}\mathbf{c}_t$ and $\boldsymbol{\theta}^{**\top} = \mathbf{A}^{-1}\mathbf{B}$, resulting in $\mathbf{w}_t^{*} = \mathbf{A}^{-1}\mathbf{c}_t + \mathbf{A}^{-1}\mathbf{B}\tilde{\mathbf{x}}_t = \bar{\mathbf{w}}_t + \boldsymbol{\theta}^{**\top}\tilde{\mathbf{x}}_t$ as well as $w_{ft}^{\sharp} = w_{ft}^{*}$.

Proof. See Appendix A.2. □

Note that if the returns $r_{it+1} \in \mathbb{R}_{\geq -1}$ and portfolio weights $\bar{w}_{it} \geq 0$ (for example $\bar{w}_{it} = 1/N_t$), then $\boldsymbol{\theta} = \mathbf{0}_{(k\times 1)}$ implies $R_{pt} \geq 0$ for $\boldsymbol{\theta} = \mathbf{0}_k$. With $r_{it+1} \in \mathbb{R}_{>-1}$ and weights $\bar{w}_{it} \geq 0$, $\sum_{i=1}^{N_t} \bar{w}_{it} = 1$, we obtain a strictly positive return at $\boldsymbol{\theta} = \mathbf{0}_{(k\times 1)}$. Part (b) of Proposition 6 is for example fulfilled if the support of the conditional distribution of $\tilde{\mathbf{x}}_{it}$ conditional on r_{it+1} is $\mathbb{R}^k$. Hence, strong assumptions on the stochastic properties of the returns and the variables $\mathbf{x}_{it}$ are necessary to obtain a well defined optimization problem when the domain of the Bernoulli utility function, $\mathbb{D}$, is a proper subset of the real line. In addition, typically optimal portfolio investment differs from the characteristics based approach. The main issue to be evaluated empirically below is the question of how good the approximation will be.

In our analysis, and in line with Brandt et al. (2009), the generalized method of moments is applied to maximize expected utility. [That is, we implicitely assume that the GMM assumptions are satisfied.] In particular, the GMM distance function is assumed to be strictly concave in the parameters. By using the first order condition to construct the GMM distance function, we observe that a concave distance function is associated with a concave expected utility in the parameters $\boldsymbol{\theta}$. To justify this assumption numerical checks have to be performed in addition to running a GMM estimation routine. Then, standard asymptotic

[21]We also get $\mathbb{E}_t\left(u\left(e_t\left(1+\sum_{i\in\mathbb{I}_t} w_{it}^{*} r_{it+1}\right)\right)\right) \geq \mathbb{E}_t\left(u\left(e_t\left(1+\sum_{i\in\mathbb{I}_t} w_{it}^{\flat} r_{it+1}\right)\right)\right) \geq \mathbb{E}_t\left(u\left(e_t\left(1+\sum_{i\in\mathbb{I}_t} w_{it}^{\sharp} r_{it+1}\right)\right)\right)$, where the terrm in the the middle will be obtained in next sections,

theory can be used in a In the case of short-selling constraints we numerically choose $\boldsymbol{\theta}$ by maximizing expected utility under the constraint $w_{it} \geq 0$ in the training sample. By applying this approach already for small $\boldsymbol{\theta}$, some w_{it} become negative for some t in the training sample. Hence, the weights obtained with this approach are very close to $\bar{w}_{it}$, which is the $1/N$-portfolio, with weights denoted $\mathbf{w}_t^{1/N}$, in our examples (typically $w_{ft}^{1/N} = 0$ and $w_{ft}^{\sharp} = 0$).[22]

A.2 Proof of Observation 6

Let $c(\boldsymbol{\theta}, \zeta) = c(\breve{\boldsymbol{\psi}}, \zeta)$, where $\breve{\Psi}_{it} = \bar{w}_{it} + \frac{1}{N_t}\boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}$.

Part (a): Considered some fixed $\boldsymbol{\theta} \in \Theta$, where the j-th coordinate is non-zero. Then $R_{pt+1} > 0$ demands for $\sum_{i\in\mathbb{I}_t} \bar{w}_{it} r_{it+1} + \boldsymbol{\theta}^\top \left(\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \tilde{x}_{it} r_{it+1}\right) > c(\boldsymbol{\theta}, \zeta) - 1$. Given $1 + r_{it+1} \geq 0$ and $\sum_{i\in\mathbb{I}_t} \bar{w}_{it} = 1$, the term $\sum_{i\in\mathbb{I}_t} 1 + \bar{w}_{it} r_{it+1} \geq 0$ (almost surely). [Hence for $\boldsymbol{\theta} = \mathbf{0}$ the returns are non-negative.] Since trading cost is non-negative $\sum_{i\in\mathbb{I}_t} \bar{w}_{it} r_{it+1} + \boldsymbol{\theta}^\top \left(\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \tilde{x}_{it} r_{it+1}\right) = \sum_{i\in\mathbb{I}_t} r_{it+1}\left(\bar{w}_{it} + \boldsymbol{\theta}^\top \frac{1}{N_t}\tilde{x}_{it}\right) < -1$ is sufficient for negative portfolio returns. Since, by assumption, the support of $\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \tilde{x}_{it,j} r_{it}$ is the real line for at least one j, $j = 1, \ldots, k$, there always exists an event of strictly positive probability, where $1 + \sum_{i\in\mathbb{I}_t} \bar{w}_{it} r_{it+1} + \boldsymbol{\theta}^\top \frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \tilde{\mathbf{x}}_{it} r_{it+1} < 0$. Since cost is non-negative this also implies that $R_{pt+1} < 0$ with strictly positive probability.

Part (b): Suppose that $r_{it} \in \mathbb{R}_{\geq -1}$ and that the joint probability

$$\mathbb{P}\left(\{|\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \tilde{x}_{it,j} r_{it+1}| > \varepsilon_{xr}\} \wedge \{\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} r_{it+1} \leq \varepsilon_r\}\right) = \delta_r > 0$$

for some $j \in \{1, \ldots, k\}$. ε_{xr} and $\varepsilon_r > 0$ are some (possibly small) numbers in $\mathbb{R}_{>0}$, while $\tilde{x}_{it,j}$ denotes the j-th coordinate of the vector $\tilde{\mathbf{x}}_{it}$.

This assumption implies that the probability of the joint event "$\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \tilde{x}_{it,j} r_{it+1} \neq 0$ and the return $\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} r_{it+1}$ is smaller than ε_r" is strictly positive. To get $R_{pt+1} \leq 0$ it is sufficient that $\sum_{i\in\mathbb{I}_t}\left[\frac{1}{N_t} + \frac{1}{N_t}\boldsymbol{\theta}^\top \mathbf{x}_{it}\right] r_{it+1} - c(\boldsymbol{\theta}, \zeta) = \frac{1}{N_t}\sum_{i\in\mathbb{I}_t} r_{it+1} + \frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \theta_1 \tilde{x}_{it,1} r_{it+1} + \cdots + \frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \theta_j \tilde{x}_{it,j} r_{it+1} + \cdots + \frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \theta_k \tilde{x}_{it,k} r_{it+1} - c(\boldsymbol{\theta}, \zeta) < -1$. With $\theta_l = 0$ for $l \neq j$, we obtain $R_{pt+1} = \frac{1}{N_t}\sum_{i\in\mathbb{I}_t} r_{it+1} + \frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \theta_j \tilde{x}_{it,j} r_{it+1} - c(\boldsymbol{\theta}, \zeta)$.

[22] Another approach to implement short-selling constraints in a pragmatic way, is to set $w_{it} = 0$ if $\bar{w}_{it} + \boldsymbol{\theta}^{*\top} \tilde{\mathbf{x}}_{it} < 0$.

Since $-c(\boldsymbol{\theta}, \zeta) \leq 0$, $\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} r_{it+1} + \left(\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \tilde{x}_{it,j} r_{it+1}\right)\theta_j < -1$ is sufficient to result in $R_{pt+1} < -1$. Since the joint probability that $\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} r_{it+1} \leq \varepsilon_r$ and $|\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \mathbf{x}_{it,j} r_{it+1}| > \varepsilon_{xr}$ is larger than zero, there exists $\theta_j \in \mathbb{R}$ where $\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} r_{it+1} + \left(\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \tilde{x}_{it,j} r_{it+1}\right)\theta_j < -1$. Therefore, there exsits $\boldsymbol{\theta} \in \Theta$, where $R_{pt+1} < 0$ with strictly positive probability.

Part (c): We assumed $0 < \underline{r} \leq r_{it+1} \leq \bar{r} < \infty$, $\underline{x} \leq \tilde{\mathbf{x}}_{it,j} \leq \bar{x}$, for all i, t and $j = 1, \ldots, k$, $0 \leq c(\boldsymbol{\theta}, \zeta) \leq \bar{c}$ as well as $1 + \underline{r} - \bar{c} > \alpha > 0$. These assumptions and the definition of the portfolio return R_{pt+1} results in $R_{pt+1} = 1 + \frac{1}{N_t}\sum_{i\in\mathbb{I}_t} r_{it+1} + \sum_{j=1^k} \theta_j \frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \tilde{x}_{it,j} r_{it+1} - c(\boldsymbol{\theta}, \zeta) > 1 + \frac{N_t}{N_t}\underline{r} + \sum_{j=1}^{k} \theta_j \frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \tilde{x}_{it,j} r_{it+1}\bar{r} - \bar{c} > \alpha + \sum_{j=1}^{k} \theta_j \frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \tilde{x}_{it,j} r_{it+1}$. Note that $|\frac{1}{N_t}\sum_{i\in\mathbb{I}_t} \tilde{x}_{it,j} r_{it+1}| \leq \max\{\bar{x}\bar{r}, |\underline{x}\bar{r}|\}$. To obtain $R_{pt+1} > 0$ it is sufficient that $\alpha > \sum_{j=1}^{k} |\theta_j|\, k \max\{\bar{x}\bar{r}, |\underline{x}\bar{r}|\} > 0$. This yields $|\theta_j| < \frac{\alpha}{k \max\{\bar{x}\bar{r}, |\underline{x}\bar{r}|\}}$. By the properties of real numbers there exist a set of θ_j fulfilling this inequality, the volume of the set where this inequality is fulfilled is $\left(\frac{\alpha}{k \max\{\bar{x}\bar{r}, |\underline{x}\bar{r}|\}}\right)^k > 0$. Hence an open set Θ'_R of positive Lebesgue measure exists where $R_{pt+1} > 0$. By construction $\mathbf{0}_{(k\times 1)}$ is contained in this set.

From the second order condition obtained in (22) we observe that the optimization problem is concave. From optimization theory it is well known that (i) if $\mathbb{E}\left(u\left(R_{pt+1}\right)\right)$ is strictly concave in $\boldsymbol{\theta}$ and a $\boldsymbol{\theta}$ satisfying the first order condition exists, we get a unique global maximum, (ii) if $\mathbb{E}\left(u\left(R_{pt+1}\right)\right)$ is concave in $\boldsymbol{\theta}$ and a $\boldsymbol{\theta}$ satisfying the first order condition exists, we get a global maximum but this maximum need not be unique, while (iii) if $\mathbb{E}\left(u\left(R_{pt+1}\right)\right)$ is concave in $\boldsymbol{\theta}$ and a $\boldsymbol{\theta}$ satisfying the first order condition does not exist, then a maximum does not exist. In the case of an open parameter set Θ, the supremum of $\mathbb{E}\left(u\left(R_{pt+1}\right)\right)$ is attained for some $\boldsymbol{\theta}$ in the closure of Θ. In this article we follow Brandt et al. (2009) and apply GMM. The first order condition (22) is used to construct a GMM distance function. In addition, we investigate $\mathbb{E}\left(u\left(R_{pt+1}\right)\right)$ by means of or graphical tools.

It turns out that the number of parameters to be estimated and the of number observations are obvious characteristics affecting the stability of the $\boldsymbol{\theta}$ estimation. For example, Brandt et al. (2009) show that conditioning the $\boldsymbol{\theta}$s on the slope of the yield curve strongly improves portfolio performance. However, it also doubles the effective number of parameters to be estimated (one set of $\boldsymbol{\theta}$s for each macro-economic condition). Moreover, by definition, the yield curve is more often “normal” than inverted. Hence, the number of observations for estimating the set of $\boldsymbol{\theta}$s applied in times of an inverted yield curve is much lower.

A similar situation arises, when parametric portfolio policies are applied to assets or with characteristics that do not have long time series available. In these cases, portfolio returns for a given $\boldsymbol{\theta}$ are less diversified than in our sample, and hence non-convergence can be more problematic *even* for higher levels of risk aversion.

A.3 Trading Cost

Following Brandt et al. (2009) we allow for the possibility of trading cost modelled by a trading cost function $c(\boldsymbol{\phi}, \zeta)$ with parameter $\zeta \geq 0$. If $\zeta = 0$, then $c(\boldsymbol{\phi}, 0) = 0$ for all strategies, while $c(\boldsymbol{\phi}, 0) \geq 0$ for all strategies and $\zeta \geq 0$. Note that trading costs also depend on prior and actual, $\boldsymbol{\phi}_s$, $1 \leq s \leq t$ and the parameter ζ. We apply the short hand $c(\boldsymbol{\phi}, \zeta)$. In particular, in the case of linear or quadratic cost functions, we get $c_1(\boldsymbol{\phi}, \zeta) := \zeta \sum_{i=1}^{N} |\phi_{it} - \phi_{it-1}|$ and $c_2(\boldsymbol{\phi}, \zeta) := \zeta \sum_{i=1}^{N} (\phi_{it} - \phi_{it-1})^2$ for a constant $N_t = N$. Then, $E_{t+1} = e_t \left(w_{ft} R_{ft+1} + \sum_{i=1}^{N} w_{it} R_{it+1} \right) - c_j(\boldsymbol{\phi}, \zeta) = \boldsymbol{\phi}_t^\top \mathbf{R}_{t+1} + \left(e_t - \sum_{i=1}^{N} \boldsymbol{\phi}_t^\top \mathbf{1}_N \right) R_{ft+1} - c_j(\boldsymbol{\phi}, \zeta)$, $j = 1, 2$. Finally, we assume that the risk-free asset can be traded at zero cost.

For general N_t we get: We consider two examples for the cost function: $c_1(\boldsymbol{\psi}, \zeta) := \zeta \sum_{i \in \mathbb{I}_t} |w_{it} - w_{+it-1}|$ (see, e.g., Brandt et al., 2009) or $c_2(\boldsymbol{\psi}, \zeta) := \zeta \sum_{i \in \mathbb{I}_t} (w_{it} - w_{+it-1})^2$, where $w_{it} - w_{+it-1}$ measures non-zero trades in period t. Hence, in these expressions $w_{+it-1} = w_{it-1}$ if asset $i \in \mathbb{I}_t$ was also traded in $t-1$, $w_{+it-1} = w_{is}$ if $0 \leq s < t-1$ is the last point of time before t where asset i was traded, and $w_{+it-1} = 0$ if asset i was not traded in some period $0 \leq s < t$. If $N = N_t$ for all t, we get $w_{+it-1} = w_{it-1}$ and the cost functions defined in Section A.3 of the main text. To highlight that cost is related to a parametric portfolio policy, we also use the notation $c_j(\boldsymbol{\theta}, \zeta) = c_j(\boldsymbol{\psi}^\sharp, \zeta)$, where $\Psi_{it}^\sharp = \bar{w}_{it} + \frac{1}{N_t} \boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}$. Note that $c_j(\boldsymbol{\theta}, \zeta)$ are convex in $\boldsymbol{\theta}$. We also observe that $c_1(\boldsymbol{\theta}, \zeta)$ and $c_2(\boldsymbol{\theta}, \zeta)$ are convex in $\boldsymbol{\theta}$. The Online-Appendix in Gehrig et al. (2018) provides conditions for a strictly convex $c_2(\boldsymbol{\theta}, \zeta)$.

B Determining Conditional Certainty Equivalents

In order to evaluate investment strategies we compare certainty equivalents. Consider an investment strategy $\boldsymbol{\phi}_t$ [either for the constrained or the unconstrained problem], CARA utility with parameter ρ, wealth/or endowment e_t, variables driving returns $\mathbf{x}_t$ and a cost function c_i with parameter ζ. For any strategy $\boldsymbol{\phi}_t$, by means of (5) the (conditional) certainty equivalent, that is the value c where $\mathbb{E}_t(-\exp(-\rho E_{t+1})) = -exp(-\rho c)$, is provided by

$$\mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t, c_i, \zeta) = e_t R_{ft+1} + \boldsymbol{\phi}_t^\top \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) - \frac{\rho}{2}\boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1})\boldsymbol{\phi}_t - \rho\zeta c_j\left(\boldsymbol{\phi}, \zeta\right) \ . \tag{24}$$

If trading costs are zero, $\mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}^*, 0, 0) \geq e_t R_{ft+1}$. To see this, for $\boldsymbol{\phi}_t = \mathbf{0}$, we get $\phi_0 = e_t$ and the certainty equivalent is $e_t R_{ft+1}$. Since $\boldsymbol{\phi}_t^*$ is chosen optimally the inequality $\mathscr{C}_t(\rho, e_t, \mathbf{x}, \boldsymbol{\phi}^*, 0, 0) \geq e_t R_{ft+1}$ has to hold. If $\boldsymbol{\phi}^* \neq \mathbf{0}$, by the strict concavity of the problem, we obtain a strict inequality. If no risk-free asset is available and all wealth e_t has to be spent, $\mathscr{C}_t(\rho, e_t, \mathbf{x}, \boldsymbol{\phi}^+, 0, 0) > e_t R_{ft+1}$ need not hold in general. In addition, since $\mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^*, c_i, \zeta)$ obtained in (24) arises from utility maximization and expected utility is strictly concave in $\boldsymbol{\phi}$ [see equation (49) in Appendix F], we get for any strategy $\boldsymbol{\phi}$,

$$0 \ \leq \ \mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^*, c_1, \zeta) - \mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t, c_1, \zeta) \tag{25}$$

with equality if and only of $\boldsymbol{\phi}_t^* = \boldsymbol{\phi}_t$.

The (unconditional) certainty equivalent $\mathscr{C}(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t, c_i, \zeta)$ is the value c where $\mathbb{E}(-\exp(-\rho E_{t+1})) = -exp(-\rho c)$. Given some data $(\mathbf{x}_t, \mathbf{r}_t: \ t = 1, \ldots, T)$, it is estimated by means of $\widehat{\mathscr{C}}(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t, c_j, \zeta) = -\frac{1}{\rho}\widehat{\mathbb{E}}\left(e^{-\rho E_t}\right)$, where $\widehat{\mathbb{E}}\left(e^{-\rho E_t}\right) = \frac{1}{T}\sum_{\ell=1}^T e^{-\rho E_t}$. By the law of iterated expectations $\mathbb{E}(-\exp(-\rho E_{t+1})) = \mathbb{E}\left(\mathbb{E}_t(-\exp(-\rho E_{t+1}))\right)$. Hence, by the law of iterated expectations and (25), also $\mathscr{C}(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t, c_i, \zeta) \leq \mathscr{C}(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^*, c_i, \zeta)$. This yields, the following large sample properties of the certainty equivalent.

Observation 7. *(i)* $\mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t, c_i, \zeta) \leq \mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^*, c_i, \zeta)$ *for any strategy* $\boldsymbol{\phi}_t$*, while* $\mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t, c_i, \zeta) \leq \mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^+, c_i, \zeta)$ *if no risk-free asset is traded and the investor has to invest* e_t*.*
(ii) $\mathscr{C}(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t, c_i, \zeta) \leq \mathscr{C}(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^*, c_i, \zeta)$ *for any strategy* $\boldsymbol{\phi}_t$*, while* $\mathscr{C}(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t, c_i, \zeta) \leq \mathscr{C}(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^+, c_i, \zeta)$ *if no risk-free asset is traded and the investor has to invest* e_t*.*

	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

Table 7: Pearson correlation coefficients for S&P 500 data.

(iii) The $1/N$-strategy, $\boldsymbol{\phi}_t^N = e_t/N \cdot \mathbf{1}_N$, corresponds to a parametric strategy with $\boldsymbol{\theta} = \mathbf{0}$ and $\bar{w}_{it} = 1/N$. Hence, $\mathscr{C}(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^{1/N}, c_i, \zeta) \leq \mathscr{C}(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^{\sharp}, c_i, \zeta)$, where $\bar{w}_{it} = 1/N$ for the parametric strategy $\boldsymbol{\phi}_t^{\sharp}$.

For the CARA the (unconditional) certainty equivalent $\mathscr{C}(\gamma, e_t, \mathbf{x}_t, \mathbf{w}_t, c_i, \zeta)$ is the value c where $\mathbb{E}(u(E_{t+1})) = u(c)$. Given that $\mathbb{E}(u(E_{t+1})) \geq u(\Psi)$, we get $c = u^{-1}\left(\mathbb{E}(u(E_{t+1}))\right) = u^{-1}\left(\mathbb{E}(u(E_{t+1}))\right)$ as already described in the main text.

C 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 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 $\mathbf{x}_{it} = (me_{it}, btm_{it}, mom_{it})^\top$. To be included in the estimation, a firm must fulfill three conditions at the portfolio formation. It must be a continous 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_t is always equal to 100. All characteristics $\mathbf{x}_{it}$ are cross-sectionally standardized according to equation (3) resulting in $\tilde{\mathbf{x}}_{it}$.

Table C 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

$$r_{it} = a + \mathbf{b}^\top \tilde{\mathbf{x}}_{it} + u_{it} \ , \tag{26}$$

where the noise terms are – in a first step – assumed to be exogenous. The ordinary least squares estimates are $\widehat{a} = 0.1471$, $\widehat{\mathbf{b}} = (0.0104, -0.0021, -0.1178)^\top$, 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 $\tilde{\mathbf{x}}_{it}$ is at least partially jointly determined with the returns, the assumption of exogenous regressors is a strong one. Therefore, we estimated the panel regression model (26) 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 $\widehat{a}_{IV} = 0.1495$ and $\widehat{\mathbf{b}}_{IV} = (0.0111, -0.0020, -0.1214)^\top$, all p-values are < 0.001.

D Simulated Data

It is insightful to compare both, simulated and empirical data. For this purpose we exploit an empirical data set that comprises $N = 100$ assets contained in S&P for the time span April 1979 to December 2013. Since $N = 100$ and $T = 416$ for the empirical data set, we simulate data with $T = 420$ and $N = 100$ with $k = 3$ characteristics $\mathbf{x}_{it}$ and gross returns $R_{it} = 1 + r_{it}$, for $i = 1, \ldots, N = 100$ assets. In every period we assume that $e_t = 1$ and zero trading cost. The observations from $t = 1, \ldots, T_{est} = 200$ are used to estimate the model parameters. Then $\boldsymbol{\phi}_t^+$ is applied for $t = T_{est} + 1, \ldots, T$. Estimates of the certainty equivalent and further descriptive statistics are obtained from this out-of-sample analysis. For simulated data, we replicate this experiment for $M = 50$ times resulting in $50 \times 220 = 11,000$ samples. The empirical data set of returns and $\mathbf{x}_{it}$ of $N = 100$ assets contained in the S&P is used to obtain the simulated data with properties close to the empirical data. The simulated data has stochastic properties close to the empirical data set. Further details on the exact simulation design are provided in Appendix D.1.

D.1 Simulation Designs

To obtain simulated data we proceed as follows: We use monthly data of the $N = 100$ companies contained in the S&P index – used in the main text and described in Appendix C – to simulate the variables $\mathbf{x}_{it}$ and asset returns R_{it}. The time span is April 1979 to December 2013. First we use the empirical data set to obtain the following estimates:

1. We obtain sample means of the returns and the variables. That is, $\bar{R}_i = \frac{1}{T}\sum_{t=1}^{T} R_{it} \in \mathbb{R}$ and $\bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it} \in \mathbb{R}^k$ for $i = 1, \ldots, N$. Let $\bar{\mathbf{R}} := \left(\bar{R}_1, \ldots, \bar{R}_N\right)^\top \in \mathbb{R}^N$, $\bar{\mathbf{x}} := (\bar{x}_{11}, \ldots, \bar{x}_{1N}, \ldots, \bar{x}_{k1}, \ldots, \bar{x}_{kN})^\top \in \mathbb{R}^{Nk}$, $\grave{R}_{it} := R_{it} - \bar{R}_i$ and $\grave{\mathbf{x}}_{it} := \mathbf{x}_{it} - \bar{\mathbf{x}}_i$.

2. The first order sample auto-covariance of the factors is $\widehat{Cov}(\mathbf{x}_{it-1}, \mathbf{x}_{it}) \in \mathbb{R}^{k\times k}$. The means of the diagonal elements of $\widehat{Cov}(\mathbf{x}_{it-1}, \mathbf{x}_{it})$ are abbreviated by $\widehat{\mathbf{c}}_x \in \mathbb{R}^k$.

3. We consider the observations of $\mathbf{x}_t \in \mathbb{R}^{Nk}$ to estimate their $Nk \times Nk$ covariance matrix $\widehat{\boldsymbol{\Sigma}}_{xx}$ by applying the Ledoit and Wolf (2004)-covariance estimator.[23]

4. We consider the panel regression model

$$\begin{aligned} \grave{R}_{it} &= \grave{\mathbf{x}}_{it}^\top \boldsymbol{\beta} + \grave{\varepsilon}_{it} \quad \text{or } \grave{\mathbf{R}} = \grave{\mathbf{X}}\beta + \grave{\boldsymbol{\varepsilon}} \ , \quad \text{where} \\ \grave{\mathbf{R}} &= \begin{pmatrix} \grave{R}_{11} \\ \vdots \\ \grave{R}_{1T} \\ \vdots \\ \grave{R}_{N1} \\ \vdots \\ \grave{R}_{NT} \end{pmatrix} \in \mathbb{R}^{NT} \ , \ \grave{\mathbf{X}} = \begin{pmatrix} \grave{\mathbf{x}}_{11}^\top \\ \vdots \\ \grave{\mathbf{x}}_{1T}^\top \\ \vdots \\ \grave{\mathbf{x}}_{N1}^\top \\ \vdots \\ \grave{\mathbf{x}}_{NT}^\top \end{pmatrix} \in \mathbb{R}^{NT\times k} \ , \text{ and } \ \grave{\boldsymbol{\varepsilon}} = \begin{pmatrix} \grave{\varepsilon}_{11} \\ \vdots \\ \grave{\varepsilon}_{1T} \\ \vdots \\ \grave{\varepsilon}_{N1} \\ \vdots \\ \grave{\varepsilon}_{NT} \end{pmatrix} \in \mathbb{R}^{NT} \ . \end{aligned} \tag{27}$$

Then $\boldsymbol{\beta}$ can be estimated by pooled ordinary least. That is $\widehat{\beta} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \grave{\mathbf{x}}_{it}\grave{\mathbf{x}}_{it}^\top\right)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} \grave{\mathbf{x}}_{it}\grave{R}_{it} = \left(\grave{\mathbf{X}}^\top\grave{\mathbf{X}}\right)^{-1} \grave{\mathbf{X}}^\top\grave{\mathbf{R}}$. Due to possible endogenity we estimate $\boldsymbol{\beta}$ by means of an instrumental variable

[23] We applied the Matlab-code available at `https://www.econ.uzh.ch/en/people/faculty/wolf/publications.html9` with default values.

estimator, where the lagged factors are used as instruments. Hence,

$$
\begin{aligned}
\widehat{\beta}_{IV} &= \left(\sum_{i=1}^{N}\sum_{t=2}^{T} \grave{\mathbf{x}}_{it-1}\grave{\mathbf{x}}_{it}^{\top}\right)^{-1} \sum_{i=1}^{N}\sum_{t=2}^{T} \grave{\mathbf{x}}_{it-1}\grave{R}_{i}t = \left(\grave{\mathbf{X}}_{IV-1}^{\top}\grave{\mathbf{F}}_{IV}\right)^{-1}\grave{\mathbf{X}}_{IV-1}^{\top}\grave{\mathbf{R}}_{IV}\ , \text{ where} \\
\grave{\mathbf{R}}_{IV} &= \grave{\mathbf{X}}_{IV}\beta + \grave{\varepsilon}_{IV}\ , \\
\grave{\mathbf{R}}_{IV} &= \begin{pmatrix} \grave{R}_{12} \\ \vdots \\ \grave{R}_{1T} \\ \vdots \\ \grave{R}_{N2} \\ \vdots \\ \grave{R}_{NT} \end{pmatrix} \in \mathbb{R}^{N(T-1)}\ ,\ \grave{\mathbf{X}}_{IV} = \begin{pmatrix} \grave{\mathbf{x}}_{12}^{\top} \\ \vdots \\ \grave{\mathbf{x}}_{1T}^{\top} \\ \vdots \\ \grave{\mathbf{x}}_{N2}^{\top} \\ \vdots \\ \grave{\mathbf{x}}_{NT}^{\top} \end{pmatrix} \in \mathbb{R}^{N(T-1)\times k}\ ,\ \grave{\mathbf{X}}_{IV-1} = \begin{pmatrix} \grave{\mathbf{x}}_{11}^{\top} \\ \vdots \\ \grave{\mathbf{x}}_{1T-1}^{\top} \\ \vdots \\ \grave{\mathbf{x}}_{N1}^{\top} \\ \vdots \\ \grave{\mathbf{x}}_{NT-1}^{\top} \end{pmatrix} \in \mathbb{R}^{N(T-1)\times k}\ ,\ \grave{\varepsilon}_{IV} = \begin{pmatrix} \grave{\varepsilon}_{12} \\ \vdots \\ \grave{\varepsilon}_{1T} \\ \vdots \\ \grave{\varepsilon}_{N2} \\ \vdots \\ \tilde{\varepsilon}_{NT} \end{pmatrix} \in \mathbb{R}^{N(T-1)}\ .
\end{aligned}
\tag{28}
$$

5. By using $\widehat{\beta}_{IV}$ we obtain the residuals $\widehat{\tilde{\varepsilon}}_{it} = \tilde{R}_{it} - \grave{\mathbf{x}}_{it}^{\top}\widehat{\boldsymbol{\beta}}_{IV}$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$. By applying the Ledoit and Wolf (2004)-covariance estimator, we obtain an estimate of the $N \times N$ covariance matrix of the noise terms $\widehat{\mathbf{\Sigma}}_{\grave{\varepsilon}\grave{\varepsilon}}$. The elements on the main diagonal are abbreviated by $\widehat{\sigma}^2_{\tilde{\varepsilon}_i}$, $i = 1, \ldots, N$.

Then, to simulate factors and returns we use the following parameters:

1. The matrix $\mathbf{\Sigma}_{xx} = \widehat{\mathbf{\Sigma}}_{xx}$ is used to obtain the covariance of the simulated factors.

2. The factors $\mathbf{x}_t$ are assumed to follow a first order autoregressive process. In particular, we use rounded values of the mean first order autocorrelation coefficients of the factors $\mathbf{x}_{it}$ denoted $\widehat{\mathbf{c}}_x$ in the above paragraphs. For $k = 3$ we get $\mathbf{c}_x = (0.96, 0.97, 0.86)^{\top} \in \mathbb{R}^k$, $\mathbf{C}_x = diag(\mathbf{c}_x) \in \mathbb{R}^{k\times k}$, $diag(\cdot)$ transforms a vector to a diagonal matrix. Let $\mathbf{C}_{xx} = \mathbf{I}_N \otimes \mathbf{C}_x \in \mathbb{R}^{Nk\times Nk}$ and $\mathbf{D}_{xx} = \mathbf{I}_N \otimes diag(\mathbf{1}_{k\times 1} - \mathbf{c}_x^2) \in \mathbb{R}^{Nk\times Nk}$, where $\otimes$ denotes the Kronecker product. Then,

$$\mathbf{x}_t = \mathbf{C}_{xx}\mathbf{x}_{t-1} + \mathbf{D}_{xx}^{1/2}\mathbf{\Sigma}_{xx}^{1/2}\mathbf{v}_t\ , \tag{29}$$

where $\mathbf{v}_t$ follows a standard normal distribution with mean vector zero and covariance matrix $\mathbf{I}_{Nk}$. Hence, the covariance matrix of $\mathbf{x}_t$ is provided by $\mathbf{\Sigma}_{xx}$, given that $\mathbf{x}_t$ is started from its stationary distribution, which is a normal distribution with mean zero and covariance $\mathbf{\Sigma}_{xx}$. We also observe that the sampling variation of first and second moments of the autoregressive process $\mathbf{x}_t$ become high due to the high serial correlation.

By using (29), the conditional expectation of the characteristics is equal to

$$\mathbb{E}_t\left(\mathbf{x}_{t+1}\right) \quad = \quad \mathbf{C}_{xx}\mathbf{x}_t \ . \tag{30}$$

3. The parameter $\mathbf{a}^0 = \bar{\mathbf{R}}$ for all t. The matrices $\mathbf{A}^j$, $j = 1, \ldots, k$, follow from the vector $\boldsymbol{\beta}_{IV} = (0.0111, -0.0020, -0.1214)^\top$, which is based on a (rounded version) of $\widehat{\mathbf{b}}_{IV}$ estimated in Section C. In particular, we proceed with $\mathbf{A}^1 = 0.0111 \cdot \mathbf{I}_{Nk}$, $\mathbf{A}^1 = -0.0020 \cdot \mathbf{I}_{Nk}$ and $\mathbf{A}^3 = -0.1214 \cdot \mathbf{I}_{Nk}$, where $\mathbf{I}_{Nk}$ denotes the kN-dimensional identity matrix. The risk-free rate $r_{ft} = 0.001$ such that $R_{ft} = 1.001$. Then by using (29) and (30) we observe that

$$\begin{aligned}
\underbrace{\mathbf{x}_t}_{[Nk\times 1]} &= \underbrace{\mathbf{C}_{xx}}_{[Nk\times Nk]} \underbrace{\mathbf{x}_{t-1}}_{[Nk\times 1]} + \underbrace{\mathbf{D}_{xx}^{1/2}}_{[Nk\times Nk]} \underbrace{\boldsymbol{\Sigma}_{xx}^{1/2}}_{[Nk\times Nk]} \underbrace{\mathbf{v}_t}_{[Nk\times 1]} \ , \\
\mathbb{E}_t\left(\mathbf{x}_{t+1}\right) &= \mathbb{E}_t\left(\mathbf{C}_{xx}\mathbf{x}_t + \mathbf{D}_{ff}^{1/2}\boldsymbol{\Sigma}_{xx}^{1/2}\mathbf{v}_{t+1}\right) = \mathbf{C}_{xx}\mathbf{x}_t \ , \\
\underbrace{\mathbf{R}_{t+1}}_{[N\times 1]} &= \underbrace{\mathbf{a}^0}_{[N\times 1]} + \underbrace{\left(\mathbf{A}^1, \ldots, \mathbf{A}^q\right)}_{[N\times Nq]} \underbrace{\mathbf{x}_{t+1}}_{[Nk\times 1]} + \underbrace{\mathbf{u}_{t+1}}_{[N\times 1]} = \bar{\mathbf{R}} + \left(\mathbf{I}_N \otimes \boldsymbol{\beta}_{IV}^\top\right)\mathbf{x}_{t+1} + \mathbf{u}_{t+1} \ , \\
\underbrace{\mathbb{E}_t\left(\mathbf{R}_{t+1}\right)}_{[N\times 1]} &= \mathbf{a}_t^0 + \left(\mathbf{A}^1, \ldots, \mathbf{A}^q\right)\mathbf{C}_{xx}\mathbf{x}_t = \underbrace{\bar{\mathbf{R}}}_{[N\times 1]} + \underbrace{\left(\mathbf{I}_N \otimes \boldsymbol{\beta}_{IV}^\top\right)}_{[N\times Nk]} \underbrace{\mathbf{C}_{xx}}_{[Nk\times Nk]} \underbrace{\mathbf{x}_t}_{[Nk\times 1]} \ , \text{ and} \\
\underbrace{\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)}_{[N\times N]} &= \underbrace{\mathbf{A}}_{[N\times Nk]} \underbrace{\mathbf{D}_{xx}^{1/2}}_{[Nk\times Nk]} \underbrace{\boldsymbol{\Sigma}_{xx}}_{[Nk\times Nk]} \underbrace{\mathbf{D}_{xx}^{1/2\top}}_{[Nk\times Nk]} \underbrace{\mathbf{A}^\top}_{[Nk\times N]} + \underbrace{\boldsymbol{\Sigma}_{xx}}_{[N\times N]} \ .
\end{aligned} \tag{31}$$

The noise term $\mathbf{u}_t$ has diagonal covariance matrix $\boldsymbol{\Sigma}_{uu}$ for all t. The elements on the main diagonal of matrix $\boldsymbol{\Sigma}_{uu}$ follow from $\widehat{\sigma}_{\varepsilon_i}^2$, $i = 1, \ldots, N$.

D.2 Some Results

On the basis of the simulated and the empirical data we run the investment strategies $\boldsymbol{\phi}_t^+$, $\boldsymbol{\phi}_t^\sharp$, etc. to obtain samples $E_{t+\ell}$, $\ell = 1, \ldots, M(T - T_{est} + 1)$, of wealth in the subsequent periods. For empirical data these samples are obtained for the time span $T_{est} + 1$ to T, while the observations from $1, \ldots, T_{est}$ are used to estimate model parameters. We work with $T_{est} = 200$ while $T = 415$. For simulated data we set

$T = 420$ and proceed with $T_{est} = 200$. The number of replications is $M = 50$; for empirical data we have $M = 1$.

In both cases an ex-ante estimate of the certainty equivalent follows from $\widehat{\mathscr{C}}(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t, c_j, \zeta) = -\frac{1}{\rho} \ln \widehat{\mathbb{E}}\left(e^{-\rho E_{t+1}}\right)$, where $\widehat{\mathbb{E}}\left(e^{-\varrho E_{t+1}}\right) = \frac{1}{(T-Test+1)} \frac{1}{M} \sum_{\ell=1}^{(T-T_{est})M} e^{-\varrho E_{\ell,t+1}}$. The standard error of the estimator $\widehat{\mathscr{C}}(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t, c_j, \zeta)$ is obtained by means of the Delta-method (see, e.g., Ruud, 2000, Lemma 16.1). On the basis of our simulations we compare the following hypothetical scenarios:

$\boldsymbol{\phi}_t^{**}$ Full information: All parameters are known such that the conditional expectation $\mathbb{E}_t\left(\mathbf{R}_{t+1}\right)$ and the conditional variance $\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)$ are known. [For simulated data only.]

$\boldsymbol{\phi}_t^{*}$ Unknown parameters: The conditional expectation and $\mathbb{E}_t\left(\mathbf{R}_{t+1}\right)$ variance $\mathbb{V}_t\left(\mathbf{R}_{t+1}\right) = \boldsymbol{\Sigma}$ are not known and have to be estimated. To do this we consider the predictive regression

$$R_{it+1} = \gamma_0 + \boldsymbol{\gamma}^\top \mathbf{x}_{it} + u_{it}$$

and obtain $\widehat{\gamma}_0$ and $\widehat{\boldsymbol{\gamma}}$ for each asset $i = 1, \ldots, N$ by means of the least-squares estimator. The observations $t = 1, \ldots, T_{est}$ are used to estimate these parameters. Then, $\widehat{\mathbb{E}}_t\left(R_{it+1}\right) = \widehat{\gamma}_0 + \widehat{\boldsymbol{\gamma}}^\top \mathbf{x}_{it}$ and $\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)$ follows from the residuals $\widehat{u}_{it}$, $i = 1, \ldots, N$.

$\boldsymbol{\phi}_{t,LW}^{*}$ Unknown parameters and reduction technique for estimating second moments: This strategy corresponds to S.2 with the difference that the conditional covariance is estimated by applying the Ledoit and Wolf (2004)-covariance estimator to $\widehat{u}_{it}$, $i = 1, \ldots, N$.

$\boldsymbol{\phi}_t^{1/N}$ Rule of the insufficient reason: Consider the $1/N$-portfolio, where the amount invested into the risky assets is $e_t = 1$.

$\boldsymbol{\phi}_t^{\sharp}$ Characteristics-based portfolio choice: Consider a parametric policy $\boldsymbol{\phi}_t^{\sharp}$, where $\bar{w}_{it} = 1/N$ and $\boldsymbol{\theta}$ is estimated by means of GMM using the observations $t = 1, \ldots, T_{est}$. The amount invested into the risky assets is $e_t = 1$.

We run simulations for three different parameters of constant absolute risk aversion, $\rho = .5$, $\rho = 1$ and $\rho = 2$. In simulated data the known parameters and the optimal strategy (i.e. $\boldsymbol{\phi}_t^{*}$), not surprisingly, dominate all other strategies. However, if the parameters have to be estimated the certainty equivalents for ϕ_t^* and $\phi_{t,LW}^*$ are below the values obtained for the 1/N-rule and parametric policies. Reduction based (Ledoit-Wolf) estimation strategy $\phi_{t,LW}^*$ dominates ϕ_t^*. Higher Sharpe ratios can be obtained with ϕ_t^* and $\phi_{t,LW}^*$ compared to $\phi_t^\sharp$ and $\phi_t^{1/N}$ but at an expense of much higher risk. This higher risk reduces the certainty equivalent of ϕ_t^* and $\phi_{t,LW}^*$ below those of $\phi_t^\sharp$ and $\phi_t^{1/N}$. Regarding this increase in risk Erlwein-Sayer et al. (2020)[Proposion 3] show how in a mean-variance setting the risk aversion parameter can be increased to account for this increase in risk due to estimation. Interestingly, in the simulated data the parametric portfolio policy generates more short-selling than the optimal investment policies.

In addition, to the results we obtained out-of-sample, we also performed an in-sample analysis where $t = 201, \ldots, T$ is used for parameter estimation and to obtain estimates of the certainty equivalent and further statistics. Here we observe that the optimal strategy also with estimated parameters is superior to the 1/N-strategy and $\phi_t^\sharp$, the best results are obtained when the Ledoit and Wolf (2004)-covariance estimator is applied.

Summing up, we observe that the feasible optimal strategies, i.e. those where model parameters are estimated, result in quite risky strategies. By comparing estimates of the certainty equivalent we observe that in this case simple rules such as parametric portfolio policies or the $1/N$-rule outperform ϕ_t^* and $\phi_{t,LW}^*$. To reduce this negative effect on performance, a shrinkage device was proposed in the main text.

$\rho = 0.5$	ϕ_t^{**}	ϕ_t^{*}	$\phi_{t,LW}^{*}$	$\phi_t^{1/N}$	$\phi_t^{\sharp}$
$\widehat{\mathscr{C}_t}$	1.2199	-4.6420	0.8242	1.0071	0.9372
$\widehat{sd}\left(\widehat{CI}\right)$	0.0025	0.4395	0.0198	0.0005	0.0182
$mean\left(E_{t+1}\right)$	1.3067	2.5221	1.7472	1.0085	1.1811
$sd\left(E_{t+1}\right)$	0.5948	5.4372	1.9351	0.0688	0.9287
e_t	1.0000	1.0000	1.0000	1.0000	1.0000
$mean(Return)=\frac{mean(E_{t+1})-e_t}{e_t}$	0.3067	1.5221	0.7472	0.0085	0.1811
Sharpe ratio $= \frac{mean(Return)-mean(r_{ft+1})}{sd(Return)}$	0.5139	0.2798	0.3856	0.1096	0.1939
$mean(w_{it} < 0)$	0.5037	0.4934	0.4942	0.0000	0.0794
$sd\left(w_{it} < 0\right)$	0.0409	0.0327	0.0367	0.0000	0.0746
$mean\left(w_{it} < -1\right)$	0.0475	0.4338	0.3346	0.0000	0.0110
$sd\left(w_{it} < -1\right)$	0.0293	0.0351	0.0357	0.0000	0.0215
$\rho = 1$	ϕ_t^{**}	ϕ_t^{*}	$\phi_{t,LW}^{*}$	$\phi_t^{1/N}$	$\phi_t^{\sharp}$
$\widehat{\mathscr{C}_t}$	1.1197	-4.3849	0.9097	1.0052	0.9149
$\widehat{sd}\left(\widehat{CI}\right)$	0.0008	15.8949	0.0056	0.0004	0.0065
$mean\left(E_{t+1}\right)$	1.1611	1.7647	1.3869	1.0086	1.0584
$sd\left(E_{t+1}\right)$	0.2908	2.8270	0.9790	0.0689	0.5025
e_t	1.0000	1.0000	1.0000	1.0000	1.0000
$mean(Return)=\frac{mean(E_{t+1})-e_t}{e_t}$	0.1611	0.7647	0.3869	0.0086	0.0584
Sharpe ratio $= \frac{mean(Return)-mean(r_{ft+1})}{sd(Return)}$	0.5507	0.2701	0.3942	0.1108	0.1143
$mean(w_{it} < 0)$	0.4967	0.4923	0.4878	0.0000	0.0565
$sd\left(w_{it} < 0\right)$	0.0409	0.0325	0.0351	0.0000	0.0550
$mean\left(w_{it} < -1\right)$	0.0029	0.3780	0.2050	0.0000	0.0000
$sd\left(w_{it} < -1\right)$	0.0062	0.0367	0.0324	0.0000	0.0000
$\rho = 2$	ϕ_t^{**}	ϕ_t^{*}	$\phi_{t,LW}^{*}$	$\phi_t^{1/N}$	$\phi_t^{\sharp}$
$\widehat{CI}$	1.0617	-0.6256	0.9673	0.9980	0.9804
$\widehat{sd}\left(\widehat{CI}\right)$	0.0002	0.8378	0.0008	0.0003	0.0008
$mean(E_t)$	1.0828	1.3707	1.1830	1.0082	1.0117
$sd(E_t)$	0.1461	1.3367	0.4704	0.0688	0.1545
e_t	1.0000	1.0000	1.0000	1.0000	1.0000
$mean(Return)=\frac{mean(E_{t+1})-e_t}{e_t}$	0.0828	0.3707	0.1830	0.0082	0.0117
Sharpe ratio $= \frac{mean(Return)-mean(r_{ft+1})}{sd(Return)}$	0.5595	0.2766	0.3869	0.1045	0.0690
$mean(w_{it} < 0)$	0.4797	0.4939	0.4899	0.0000	0.2808
$sd\left(w_{it} < 0\right)$	0.0406	0.0352	0.0351	0.0000	0.0902
$mean\left(w_{it} < -1\right)$	0.0000	0.2704	0.0546	0.0000	0.0000
$sd\left(w_{it} < -1\right)$	0.0003	0.0495	0.0237	0.0000	0.0000

Table 8: Comparison of investment strategies in simulated data. $mean(Return) = mean\left(E_{t+1}\right) - e_t$. Sample means (mean) and sample standard deviations (sd) are obtained form $50 \cdot (T - T_{est}) = 11,000$ samples. $T = 420$, $T_{est} = 200$. $mean(w_{it} < \nu)$ denotes the average proportion of weights smaller than ν, while $sd(w_{it} < \nu)$ is the corresponding standard deviation.

E Extending CRRA Uility

In the following we propose some simple ways, we to extend CRRA utility to the real line. Let $N_{R_{it+1}>0}$ the number of (risky) assets where $R_{it+1} > 0$, for all $t = 1, \dots, T$, for some chosen $\boldsymbol{\theta}$. For $\theta = \mathbf{0}$ we have $N_{R_{it+1}>0} = N_t$. From the above analysis we already know that some R_{pt+1} become negative if $\boldsymbol{\theta}$ becomes large in absolute terms. For $\theta \in \mathbb{R}$ the number of assets with positive returns $N_{R_{it+1}>0}$ is monotone decreasing in $|\theta|$. In particular, we observe that $N_{R_{it+1}>0}$ is a step function.

In order to obtain a well defined model in mathematical terms, we can either augment the CRRA Bernoulli utility function $v(\cdot)$ defined on the domain $\mathbb{D} = \mathbb{R}$ or work directly with a different utility function, which is defined on the whole real line (CARA utility or mean variance preferences were applied e.g. in Ammann et al., 2016). This section follows the first suggestion and starts with the augmentation approach by specifying simple and easily implementable rules for dealing with large negative returns (i.e., close to or below zero). In particular, we consider the following three alternatives to augment the CRRA Bernoulli utility function: extending utility horizontally to the left, ignoring negative returns, or extending utility in a linear way for low returns:

Augmentation I (extending utility horizontally to the left), $v_+ (E_{t+1})$: Choose some $\psi_R > 0$ and define

$$v_+ (E_{t+1}) \quad := \quad \begin{cases} v (e_t R_{pt+1}) & \text{for } R_{pt+1} \geq \Psi_R, \\ \underline{u} \leq v (e_t \psi_R) & \text{for } R_{pt+1} < \Psi_R. \end{cases} \tag{32}$$

Recall that $E_{pt+1} = e_t R_{pt+1}$.

Augmentation II (ignoring negative returns), $v_- (E_{pt+1})$: Consider $N_{R_{it+1}>0}$ defined in the above paragraphs [i.e., $N_{R_{it+1}>0}$ is the number of assets where $R_{it+1} > 0$]. The parametric portfolio weight function (4) is now applied to those assets where $R_{it+1} > 0$. That is, N_t is replaced by $N_{R_{it+1}>0}$ and $\bar{w}_{it} = \frac{1}{N_{t,R_{it+1}\geq 0}}$ in (4). This results in the portfolio return $R_{pt+1,-}$. Then,

$$v_- (E_{pt+1}) \quad := \quad v (e_t R_{pt+1,-}) \ . \tag{33}$$

Since $N_{R_{it+1}>0}$ is known after R_{it+1}, $i =, 1 \dots, N$, realizes, the calculation of $R_{pt+1,-}$ only works in-sample.

However, in-sample $v_-(R_{pt+1})$ can be used to find an optimal $\boldsymbol{\theta}$.

Augmentation III (linear continuation), $v_\flat(E_{pt+1})$: Another extension of $v(x)$, similar to (32), but less radical is a smooth linear extension for low returns:

$$v_\flat(e_t R_{pt+1}) := \begin{cases} v(e_t R_{pt+1}) & \text{for } R_{pt+1} \geq \psi_R, \\ (v(e_t\psi_R) - \delta v'(e_t\psi_R)) + \delta v'(e_t\psi_R)\, e_t R_{pt+1} & \text{for } R_{pt+1} < \psi_R, \end{cases} \quad (34)$$

where $\psi_R > 0$. As already stated in the main text, with (12) we apply $v(e_t R_{pt+1})$ for all $R_{pt+1} \geq \psi_R$. At $R_{pt+1} = \psi_R$ we get $v_\flat(e_t\psi_R) = v(e_t\psi_R) = (v(e_t\psi_R) - \delta v'(e_t\psi_R)) + \delta v'(e_t\psi_R)\, e_t\psi_R$. For $\delta = 1$, we observe that $v'_\flat(e_t\psi_R) = v'(e_t\psi_R)$ is equal to the slope of the line described by $(e_t v(\psi_R) - 1 \cdot v'(e_t\psi_R)) + 1 \cdot v'(e_t\psi_R)\, R_{pt+1}$.

The domain of $v_+(e_t R_{pt+1})$ and $v_\flat(e_t R_{pt+1})$ is the real line, while for $v_-(e_t R_{pt+1})$ non-positive returns were excluded. These simple remedies avoid strong a-priory restrictions on the support of $\mathbf{x}_{it}$ and r_{it+1}. (32) can be seen as a special case of (12) when we set $v' = 0$. Note that by the flat segment, the function $v_+(R_{pt+1})$ is neither concave nor strictly concave on the domain $\mathbb{R}$, while $v_\flat(e_t R_{pt+1})$ is at least concave. The function $v_-(e_t R_{pt+1})$ is locally strictly convex on those segments where $N_{t,R_{it+1}\geq 0}$ stays constant. Since $N_{t,R_{it+1}\geq 0}$ does not stay constant in our data sets, this generates non-well behaved expected utility as will be demonstrated in the following Figures.

In simulation runs not reported here and in the empirical data, for $\boldsymbol{\theta}$ in the neighborhood of $\mathbf{0}$ we observe essentially no differences between the expected utilities obtained with $v(\cdot)$ and $v_+(\cdot)$, $v_-(\cdot)$ or $v_\flat(\cdot)$. For different values of relative risk aversion, Figures 2 and 3 plot (numerical estimates of) expected utility $\mathbb{E}(v(e_t R_{pt+1}))$, where $e_t = 1$ and $v_+(\cdot)$, $v_-(\cdot)$ or $v_\flat(\cdot)$ are used, against θ_1, θ_2 or θ_3 ($\theta_i \in [-100, 100]$), the other coordinates are set to zero (note that θ_1 is the parameter associated with ln of book-to-market ratio, θ_2 with the ln of the firm's market equity and θ_3 with the momentum variable). So for illustration purposes we only consider projections on one policy variable.[24] Figures 2 and 3 are organized as follows:

[24]We proceeded in this way to highlight the result observed with (32) and illustrate it in graphical terms. Another practical reason is that the performance of the numerical optimization routine is easier to investigate with a single parameter.

In addition, since $\mathbb{E}(u(R_{pt+1}))$ cannot be derived analytically in our examples, numerical estimates of $\mathbb{E}(u(R_{pt+1}))$ are obtained by means of $\frac{1}{T}\sum_{t=1}^{T} u(R_{pt+1}) =: \widehat{\mathbb{E}}(u(R_{pt+1}))$. To improve the readability $\mathbb{E}(u(R_{pt+1}))$ will also be used also for numerical estimates of expected utility.

Subfigures (a,b,c) consider expected utilities with $v_+(\cdot)$, Subfigures (d,e,f) consider expected utilities with $v_-(\cdot)$, Subfigures (g,h,i) consider expected utilities with $v_\flat(\cdot)$, while Subfigures (j,k,l) jointly plot expected utilities with $v_+(\cdot)$ [dotted line], $v_-(\cdot)$ [solid red line] and $v_\flat(\cdot)$ [solid blue line]. The first, the second and the third column presents expected utility levels when θ_1, θ_2 and θ_3 is varied, respectively. Proceeding this way we find a surprising result: expected utility may not attain a global maximum. Even a local maximum may not be globally optimal.

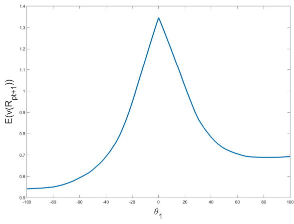
For small values of relative risk aversion, e.g. $\gamma = 0.25$ and in the absence of trading costs (i.e. $\zeta = 0$), we do not observe an interior maximum (see Figure 2) for $v_+(\cdot)$ and $v_-(\cdot)$ while for $v_\flat(\cdot)$ we derive an interior maximum for our empirical data set. Figure 3 shows the results with a higher risk aversion parameter of $\gamma = 2$. Here we observe interior maxima with $v_+(\cdot)$ and $v_\flat(\cdot)$. The lines for $v_+(\cdot)$ look less smooth due to the fact that a constant utility level is used for returns $\leq \psi_R$. For $v_-(\cdot)$ we observe some downward spikes, which arise at those θ where $N_{R_{it}>0}$ changes.[25] In Appendix E.1 we show that adding linear cost (see Figures 4 and 5) or quadratic cost (see Figures 6 and 7), the effects discussed without cost become less pronounced. Hence, higher degrees of risk aversion and higher costs render the optimization problems more concave, in which case unique maxima of expected utility result more often.[26] The above analysis suggests that the augmentation $v_\flat(R_{pt+1})$ seems to work.

E.1 CRRA Utility with Transaction Costs

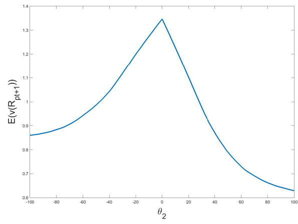
Figures 4 to 7 complement the analysis performed for Figures 2 and 3 with linear (Figures 4 and 5) and quadratic (Figures 6 and 7) transaction costs. The augmentation strategies are described in the above Appendix E.

[25] In some of the aggregated Subfigures (j,k,l) these effects are very hard to observe due to the scaling used on the vertical axis. This scaling follows from the scaling arsing with $v_\flat(\cdot)$ and is the same as in the subfigures (g,h,i).

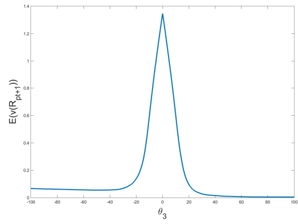
[26] Further results with S&P 500 as well as CRSP-data are provided in the main text and in the Online-Appendix in our prior version of this paper (Gehrig et al., 2018).



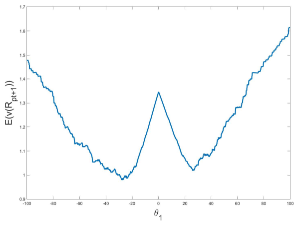
(a)



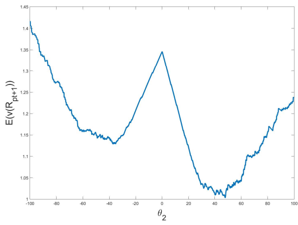
(b)



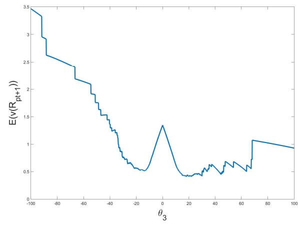
(c)



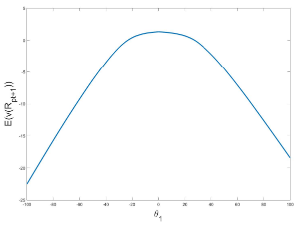
(d)



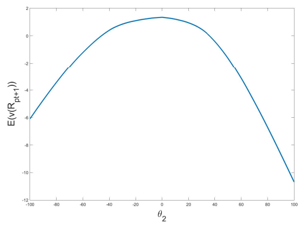
(e)



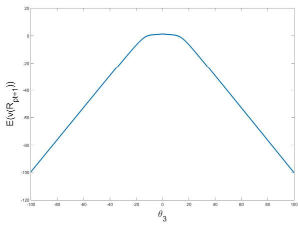
(f)



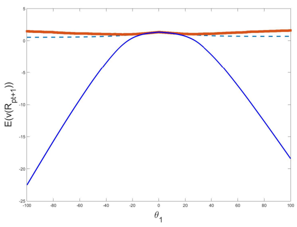
(g)



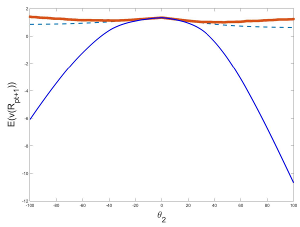
(h)



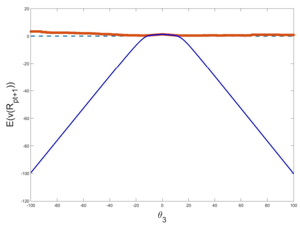
(i)



(j)

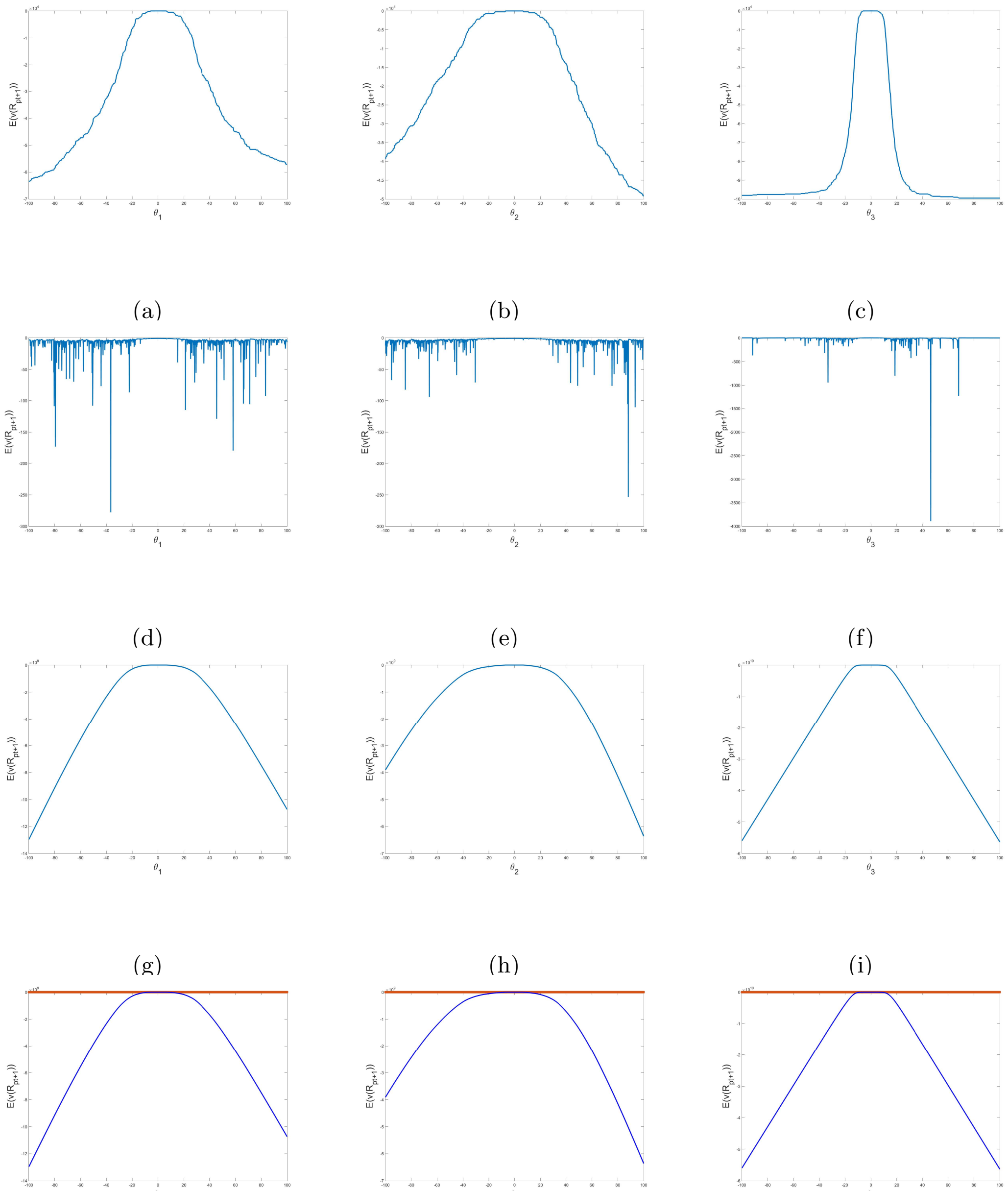


(k)



(l)

Figure 4: Expected utility $\mathbb{E}\left(v\left(R_{pt+1}\right)\right)$ against $\theta_1 \in [-500, 500]$, $\theta_2 = 0$ and $\theta_3 = 0$ in the first column; $\theta_2 \in [-500, 500]$, $\theta_1 = 0$ and $\theta_3 = 0$ in the second column; and $\theta_3 \in [-500, 500]$, $\theta_1 = 0$ and $\theta_2 = 0$ in the third column. CRRA type utility with $\gamma = 0.25$. Grid with step-width 0.25. $g(\boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}) = \boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}$, Subfigures (a-c) augmented CRRA utility $v_+(\cdot)$ with parameter γ, $\psi_R = 0.0001$ and $\underline{u} = v(\psi_R)$. Subfigures (d-f) augmented CRRA utility $v_-(\cdot)$, Subfigures (g-i) augmented CRRA utility $v_\flat(\cdot)$, with $\delta = 1$. Subfigures (j-l) summary, where the dotted lines denotes $v_+(\cdot)$, the solid red lines $v_-(\cdot)$ and the solid blue lines $v_\flat(\cdot)$. S&P 500 data; $T = 416$ and $N = 100$.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure 5: Expected utility $\mathbb{E}\left(v\left(R_{pt+1}\right)\right)$ against $\theta_1 \in [-500, 500]$, $\theta_2 = 0$ and $\theta_3 = 0$ in the first column; $\theta_2 \in [-500, 500]$, $\theta_1 = 0$ and $\theta_3 = 0$ in the second column; and $\theta_3 \in [-500, 500]$, $\theta_1 = 0$ and $\theta_2 = 0$ in the third column. CRRA type utility with $\gamma = 2$. Grid with step-width 0.25. $g(\boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}) = \boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}$, Subfigures (a-c) augmented CRRA utility $v_+(\cdot)$ with parameter γ, $\psi_R = 0.0001$ and $\underline{u} = v(\psi_R)$. Subfigures (d-f) augmented CRRA utility $v_-(\cdot)$, Subfigures (g-i) augmented CRRA utility $v_\flat(\cdot)$, with $\delta = 1$. Subfigures (j-l) summary, where the dotted lines denotes $v_+(\cdot)$, the solid red lines $v_-(\cdot)$ and the solid blue lines $v_\flat(\cdot)$. Cost function $c_1(\cdot)$. S&P-500 data; $T = 416$ and $N = 100$.

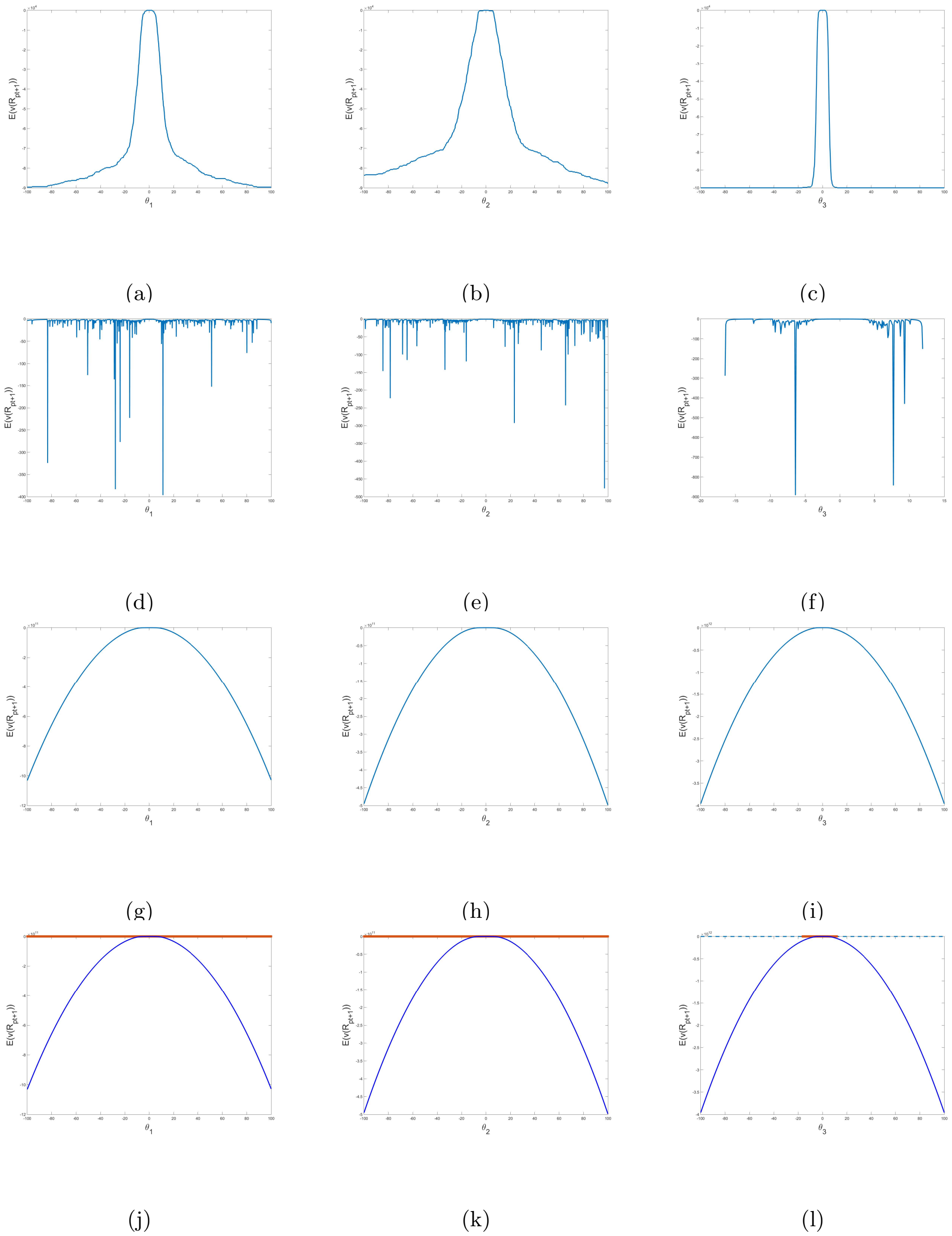


Figure 6: Expected utility $\mathbb{E}\left(v\left(R_{pt+1}\right)\right)$ against $\theta_1 \in [-500, 500]$, $\theta_2 = 0$ and $\theta_3 = 0$ in the first column; $\theta_2 \in [-500, 500]$, $\theta_1 = 0$ and $\theta_3 = 0$ in the second column; and $\theta_3 \in [-500, 500]$, $\theta_1 = 0$ and $\theta_2 = 0$ in the third column. CRRA type utility with $\gamma = 0.25$. Grid with step-width 0.25. $g(\boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}) = \boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}$, Subfigures (a-c) augmented CRRA utility $v_+(\cdot)$ with parameter γ, $\psi_R = 0.0001$ and $\underline{u} = v(\psi_R)$. Subfigures (d-f) augmented CRRA utility $v_-(\cdot)$, Subfigures (g-i) augmented CRRA utility $v_\flat(\cdot)$, with $\delta = 1$. Subfigures (j-l) summary, where the dotted lines denotes $v_+(\cdot)$, the solid red lines $v_-(\cdot)$ and the solid blue lines $v_\flat(\cdot)$. Cost function $c_2(\cdot)$. S&P-500 data; $T = 416$ and $N = 100$.

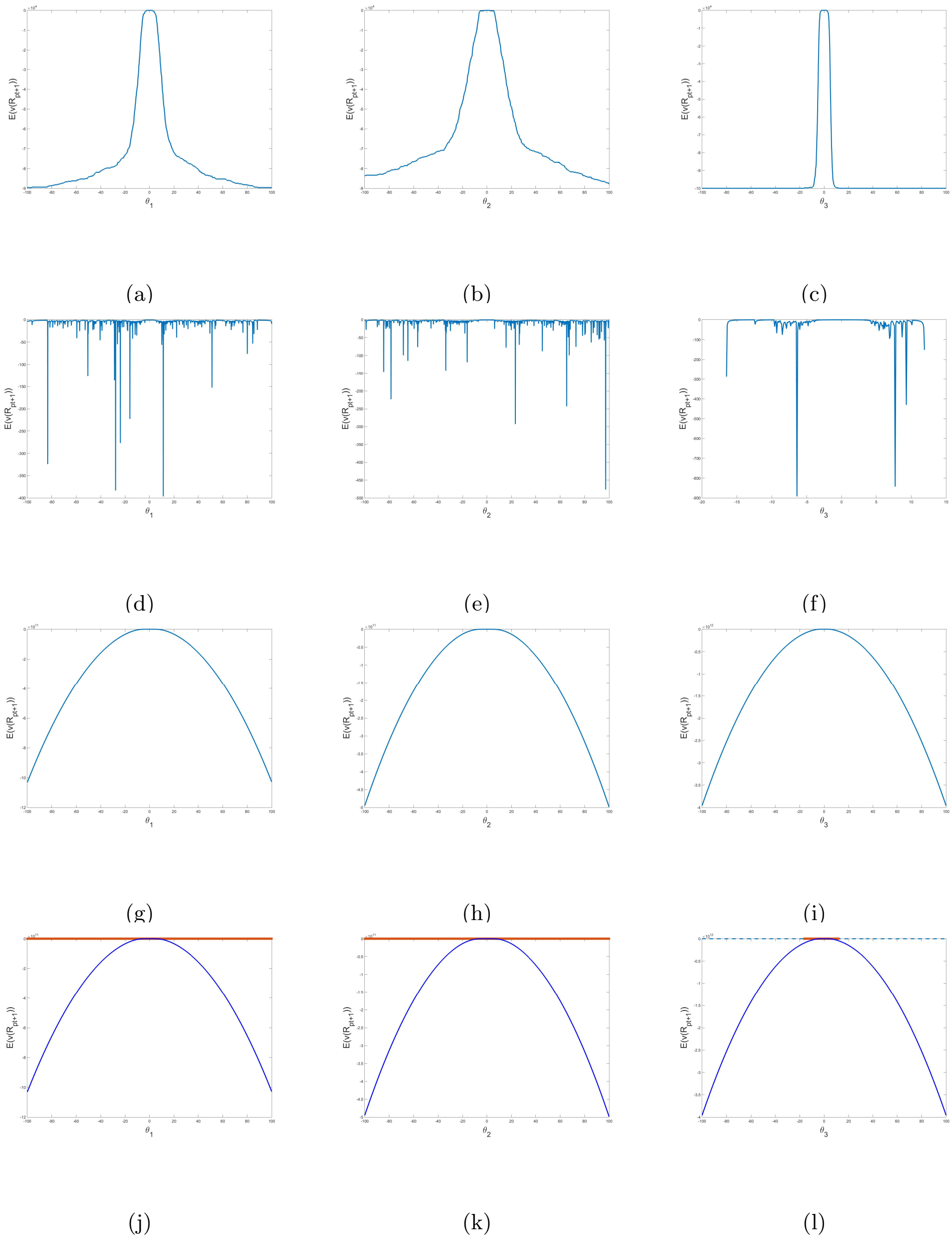


Figure 7: Expected utility $\mathbb{E}\left(v\left(R_{pt+1}\right)\right)$ against $\theta_1 \in [-500, 500]$, $\theta_2 = 0$ and $\theta_3 = 0$ in the first column; $\theta_2 \in [-500, 500]$, $\theta_1 = 0$ and $\theta_3 = 0$ in the second column; and $\theta_3 \in [-500, 500]$, $\theta_1 = 0$ and $\theta_2 = 0$ in the third column. CRRA type utility with $\gamma = 2$. Grid with step-width 0.25. $g(\boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}) = \boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}$, Subfigures (a-c) augmented CRRA utility $v_+(\cdot)$ with parameter γ, $\psi_R = 0.0001$ and $\underline{u} = v(\psi_R)$. Subfigures (d-f) augmented CRRA utility $v_-(\cdot)$, Subfigures (g-i) augmented CRRA utility $v_\flat(\cdot)$, with $\delta = 1$. Subfigures (j-l) summary, where the dotted lines denotes $v_+(\cdot)$, the solid red lines $v_-(\cdot)$ and the solid blue lines $v_\flat(\cdot)$. Cost function $c_2(\cdot)$. S&P-500 data; $T = 416$ and $N = 100$.

F Parametric Portfolio Policies with Constant Absolute Risk Aversion

Consider constant absolute risk aversion, where the Bernoulli utility function is given by $u(x) = -\exp(-\rho x)$ and the parameter $\rho > 0$ expresses constant relative risk aversion (note that $\frac{u''(x)}{u'(x)} = \rho$). In addition to N risky assets we also consider the case where a risk-free asset is traded [the risk-free asset has cross-sectional index $i = f$]. The portfolio vector of risky-assets is $\boldsymbol{\phi}_t = (\phi_{1t}, \ldots, \phi_{Nt})^\top \in \mathbb{R}^N$, where ϕ_{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. We analyse optimal investments strategies first in the absence of any transactions costs and then by including linear or quadratic transactions costs.

F.1 Optimal Strategy in the Absence of Transaction Costs

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 $\boldsymbol{\phi}_t \in \Theta = \mathbb{R}^N$. We follow the notation from the main text, and write $R_{it+1} \in \mathbb{R}$ and $\mathbf{R}_{t+1} = (R_{1t+1}, \ldots, R_{Nt+1})^\top \in \mathbb{R}^N$; $\mathbf{x}_{it} \in \mathbb{R}^q$ and $\mathbf{x}_t \in \mathbb{R}^{Nq}$ denote a random vectors of characteristics [as well as – we some abuse of notation – also their realizations]. Following Sections 3 and 4 we obtain

$$
\begin{aligned}
R_{it} &= a_{0it} + \underbrace{\left(A_i^1, \ldots, A_i^k \right)}_{[1 \times q]} \underbrace{\mathbf{x}_{it}}_{[q \times 1]} + \varepsilon_{it} \text{ and} \\
\underbrace{\mathbf{R}_t}_{[N \times 1]} &= \underbrace{\mathbf{a}_{0t}}_{[N \times 1]} + \underbrace{\left(\mathbf{A}^1, \ldots, \mathbf{A}^k \right)}_{[N \times Nk]} \underbrace{\mathbf{x}_t}_{[Nk \times 1]} + \underbrace{\boldsymbol{\varepsilon}_t}_{[N \times 1]}, \qquad (35)
\end{aligned}
$$

where $\mathbf{A}^j$ are diagonal $N \times N$ matrices. $\mathbf{A}^j$, $j = 1, \ldots, k$, follows from $\mathbf{A}_l$, $l = 1, \ldots, N$. Consider some strategy $\boldsymbol{\phi}_t$ and suppose that a risk-free asset is traded. The total amount invested into risky assets is $e_{rt} := \mathbf{1}_N^\top \boldsymbol{\phi}_t$. Let $\phi_{it} = [\boldsymbol{\phi}_t]_i$. Suppose that a risk-free asset is traded. Then strategy $\boldsymbol{\phi}_t$ translates into investment weights as follows:

$$\tilde{w}_{it} = \frac{\phi_{it}}{e_t} \text{ and } \tilde{w}_{0t} = w_{ft} = \frac{e_t - \mathbf{1}_N^\top \boldsymbol{\phi}_t}{e_t} \text{ for weights such that } \sum_{i=0}^{N} \tilde{w}_{it} = 1 \text{ or}$$

$$w_{it} = \frac{\phi_{it}}{\mathbf{1}_N^\top \boldsymbol{\phi}_t} , \text{ for weights in terms of the amount invested into risky assets } \mathbf{1}_N^\top \boldsymbol{\phi}_t \text{ such that } \sum_{i=1}^{N} w_{it} = 1 . \quad (36)$$

Note that the weights $\tilde{w}_{it}$ depend on e_t, while the weights w_{it} do not depend on e_t but on the total amount invested into risk assets $\mathbf{1}_N^\top \boldsymbol{\phi}_t$. That is, even for a strategy ϕ_{it} which is linear in $\mathbf{x}_{it}$, the weights w_{it} need not be linear in $\mathbf{x}_{it}$. By contrast if no risk-free asset is traded and the preferences of the investor / or the economic constraints are such that all wealth is invested into the risky assets we get

$$e_t = \mathbf{1}_N^\top \boldsymbol{\phi}_t \text{ and } \tilde{w}_{it} = w_{it} = \frac{\phi_{it}}{e_t} = \frac{\phi_{it}}{\mathbf{1}_N^\top \boldsymbol{\phi}_t} . \quad (37)$$

Suppose that ϕ_{it} can be written as $\phi_{it} = \gamma_{0i} + \boldsymbol{\gamma}_1^\top \tilde{\mathbf{x}}_{it}$, where $\tilde{\mathbf{x}}_{it}$ is a standardized characteristic, $\gamma_{0it} = \gamma_{0i}$ for all t and $\boldsymbol{\gamma}_{1it} = \boldsymbol{\gamma}_1$ for all t, i. Then, $\underline{e}_{rt} := \mathbf{1}_N^\top \boldsymbol{\phi}_t = \sum_{i=1}^N \left(\gamma_{0i} + \boldsymbol{\gamma}_1^\top \tilde{\mathbf{x}}_{it}\right) = \sum_{i=1}^N \gamma_{0i} + \boldsymbol{\gamma}_1^\top \sum_{i=1}^N \tilde{\mathbf{x}}_{it}$ $= \sum_{i=1}^N \gamma_{0i}$ and the total amount invested into the risky assets does not depend on $\tilde{\mathbf{x}}_{it}$. Next, suppose that $\mathbb{E}_t(-\exp(-\rho E_{t+1}))$ exists, then conditional expect utility is given by

$$\mathbb{E}_t(-\exp(-\rho E_{t+1})) = \mathbb{E}_t\left(-\exp\left(-\rho\left[(e_t - \boldsymbol{\phi}_t^\top \mathbf{1}_N) R_{ft+1} + \rho \boldsymbol{\phi}_t^\top \mathbf{R}_{t+1}\right]\right)\right) . \quad (38)$$

To obtain the optimal ϕ_{it} we take first partial derivatives in (38), resulting in:

$$\begin{aligned}
0 &= \frac{\partial}{\partial \phi_{it}} \mathbb{E}_t(-\exp(-\rho E_{t+1})) \\
&= \mathbb{E}_t\left(\exp\left(-\rho\left[(e_t - \boldsymbol{\phi}_t^\top \mathbf{1}_N) R_{ft+1} + \boldsymbol{\phi}_t^\top \mathbf{R}_{t+1}\right]\right) \rho\left[R_{it+1} - R_{ft+1}\right]\right) \\
&= -\mathbb{E}_t\left(\exp(-\rho e_t) \exp\left(-\rho\left[-\boldsymbol{\phi}_t^\top \mathbf{1}_N R_{ft+1} + \boldsymbol{\phi}_t^\top \mathbf{R}_{t+1}\right]\right) \rho\left[R_{it+1} - R_{ft+1}\right]\right) \\
&= -\exp(-\rho e_t)\, \mathbb{E}_t\left(\exp\left(-\rho\left[-\boldsymbol{\phi}_t^\top \mathbf{1}_N R_{ft+1} + \boldsymbol{\phi}_t^\top \mathbf{R}_{t+1}\right]\right) \rho\left[R_{it+1} - R_{ft+1}\right]\right) \\
&= -\exp(-\rho e_t)\, \mathbb{E}_t\left(\exp\left(-\rho\left[\boldsymbol{\phi}_t^\top (\mathbf{R}_{t+1} - R_{ft+1} \mathbf{1}_N)\right]\right) \rho\left[R_{it+1} - R_{ft+1}\right]\right) , \quad (39)
\end{aligned}$$

for $i = 1, \ldots, N$. From (39) we cannot obtain a closed from solution. Hence, we require that the noise terms are conditionally normal. Using Assumption 2 of normally distributed innovations we derive conditional expected utility

$$\begin{aligned}
\mathbb{E}_t(-\exp(-\rho E_{t+1})) &= \mathbb{E}_t\left(-\exp\left(-\rho\left[(e_t-\boldsymbol{\phi}_t^\top \mathbf{1}_N)R_{ft+1}+\boldsymbol{\phi}_t^\top \mathbf{R}_{t+1}\right]\right)\right)\\
&= -\exp\left[-\rho(e_t-\boldsymbol{\phi}_t^\top \mathbf{1}_N)R_{ft+1}-\rho\boldsymbol{\phi}_t^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right)+\frac{\rho^2}{2}\boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1})\boldsymbol{\phi}_t\right]\\
&= -\exp\left[-\rho e_t-\rho\boldsymbol{\phi}_t^\top\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right)-\mathbf{1}_N R_{ft+1}\right)+\frac{\rho^2}{2}\boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1})\boldsymbol{\phi}_t\right]\ . \qquad (40)
\end{aligned}$$

Maximizing (40) yields the vector of optimal amounts invested into the risky assets (see also, Ferson and Siegel, 2001):

$$\boldsymbol{\phi}_t^*\left(\mathbf{x}_t\right) = \frac{1}{\rho}\left(\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\right)^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right)-R_{ft}\mathbf{1}_N\right)\ . \qquad (41)$$

The remaining wealth $\phi_{ft}=e_t-\mathbf{1}_N^\top\boldsymbol{\phi}_t^*\in\mathbb{R}$ is invested into the risk-free asset.

To account for the case where no risk-free asset is traded we include the constraint $e_t\geq\mathbf{1}_N^\top\boldsymbol{\phi}_t$ and – to reduce the computational burden – we take the log of 'minus (40)' and divide by $\rho>0$. This results in the Lagrangian (for a similar problem see also Campbell, 2017, Section 2.2.3)

$$\mathscr{L}_t\left(\boldsymbol{\phi}_t,\lambda_t\right) = e_t+\boldsymbol{\phi}_t^\top\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right)-\mathbf{1}_N R_{ft+1}\right)-\frac{\rho}{2}\boldsymbol{\phi}_t^\top\mathbb{V}_t(\mathbf{R}_{t+1})\boldsymbol{\phi}_t+\lambda_t\left(e_t-\mathbf{1}_N^\top\boldsymbol{\phi}_t\right)\ . \qquad (42)$$

By taking first derivatives with respect to ϕ_{it}, $i=1,\ldots,N$, and λ_t we obtain the first order conditions

$$\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\phi}_t^\top}\mathscr{L}_t\left(\boldsymbol{\phi}_t,\lambda_t\right) &= \left[\mathbf{I}_N\mathbb{E}_t\left(\mathbf{R}_{t+1}\right)-\mathbf{I}_N\mathbf{1}_N R_{ft+1}-\frac{2}{2}\rho\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\boldsymbol{\phi}_t\right]-\lambda_t\mathbf{1}_N=\mathbf{0}_{N\times 1}\ ,\\
\frac{\partial}{\partial\lambda_t}\mathscr{L}_t\left(\boldsymbol{\phi}_t,\lambda\right) &= e_t-\mathbf{1}_N^\top\boldsymbol{\phi}_t\geq 0\ ,\ \lambda_t\frac{\partial}{\partial\lambda_t}\mathscr{L}_t\left(\boldsymbol{\phi}_t,\lambda\right)=0\ . \qquad (43)
\end{aligned}$$

If the constraint $e_t-\mathbf{1}_N^\top\boldsymbol{\phi}_t$ is not binding, then $\lambda_t=0$ and the optimal strategy is given by (41). By contrast, suppose that $e_t - \mathbf{1}_N^\top \boldsymbol{\phi}_t^* < 0$, then the constraint becomes binding and

$$\boldsymbol{\phi}_t^+(\mathbf{x}_t) = \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right)^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - R_{ft}\mathbf{1}_N\right) - \lambda_t^+ \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right)^{-1}\mathbf{1}_N \, .$$

By plugging in $\boldsymbol{\phi}_t^+(\mathbf{x}_t)$ into $e_t = \mathbf{1}_N^\top \boldsymbol{\phi}_t^+$, we get

$$e_t = \mathbf{1}_N^\top \left[\frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right)^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - R_{ft}\mathbf{1}_N\right) - \lambda_t^+ \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right)^{-1}\mathbf{1}_N\right] ,$$

such that

$$\begin{aligned}
\lambda_t^+ &= \frac{\rho}{\mathbf{1}_N^\top \mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\left[\frac{1}{\rho}\mathbf{1}_N^\top \mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - R_{ft}\mathbf{1}_N\right) - e_t\right] \\
&= \frac{\rho}{\mathbf{1}_N^\top \mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\left[\mathbf{1}_N^\top \boldsymbol{\phi}_t^*(\mathbf{x}_t) - e_t\right] \geq 0 \text{ and} \\
\boldsymbol{\phi}_t^+(\mathbf{x}_t) &= \boldsymbol{\phi}_t^*(\mathbf{x}_t) - \lambda_t^+ \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right)^{-1}\mathbf{1}_N \\
&= \boldsymbol{\phi}_t^*(\mathbf{x}_t) - \frac{\rho}{\mathbf{1}_N^\top \mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\left[\mathbf{1}_N^\top \boldsymbol{\phi}_t^*(\mathbf{x}_t) - e_t\right]\frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right)^{-1}\mathbf{1}_N \\
&= \frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - R_{ft}\mathbf{1}_N\right) \\
&\quad -\frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N\left(\frac{\rho}{\mathbf{1}_N^\top \mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\left(\mathbf{1}_N^\top \frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbb{E}_t(\mathbf{R}_{t+1}) - e_t\right)\right) \\
&\quad +\frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\frac{\rho}{\mathbf{1}_N^\top \mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\mathbf{1}_N^\top \frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N R_{ft}\right) \\
&= \frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - \frac{\rho\left(\frac{1}{\rho}\mathbf{1}_N^\top \mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbb{E}_t(\mathbf{R}_{t+1}) - e_t\right)}{\mathbf{1}_N^\top \mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\mathbf{1}_N\right) \, .
\end{aligned} \tag{44}$$

Note that the Lagrange multiplier λ_t is proportional to the difference between the optimal investments into the risky assets if a risk-free asset would be traded, $\mathbf{1}_N^\top \boldsymbol{\phi}_t^*(\mathbf{x}_t)$, and the initial wealth e_t.[27]

Note that we have to distinguish between an investment strategy linear in $\mathbf{x}_t$, that is to say $\tilde{\phi}_{it} = \tilde{\phi}_{0it} + \tilde{\boldsymbol{\phi}}_{1i}^\top \mathbf{x}_t$, an investment strategy linear in $\mathbf{x}_{it}$, that is to say $\tilde{\phi}_{it}^* = \tilde{\phi}_{0it} + \tilde{\boldsymbol{\phi}}_{1i}^\top \mathbf{x}_{it}$ and investment weights $\mathbf{w}_t$ which are linear in $\mathbf{x}_t$ or $\tilde{\mathbf{x}}_{it}$. By means of (36) we observe that even for an investment strategy affine linear in $\mathbf{x}_t$ or $\tilde{\mathbf{x}}_{it}$, the weights need not be affine linear. From (41) and (44) we observe:

- The optimal $\boldsymbol{\phi}_t^*$ does not depend on the initial wealth e_t. The total amount invested into risky assets $\boldsymbol{\phi}_t(\mathbf{x}_t)^{*\top}\mathbf{1}_N$ depends on $\mathbf{x}_t$. The amount invested into the risk-free asset follows from $\phi_{ft} = e_t -$

[27]By means of the Hessian obtained later in (49), we observe that also the second order condition is satisfied. To consider the case without trading cost set $\zeta = 0$ in (49). For the constraint problem the strictly concave value function is also sufficient to obtain a constraint maximum (see, e.g., Simon and Blume, 1994).

$\boldsymbol{\phi}_t^*(\mathbf{x}_t)^\top \mathbf{1}_N$. Given that $\boldsymbol{\phi}_t(\mathbf{x}_t)^{*\top} \mathbf{1}_N \geq e_t$, for the constraint problem $\phi_{ft} = 0$ and $\boldsymbol{\phi}_t^+(\mathbf{x}_t)^\top \mathbf{1}_N = e_t$.

- Suppose that $\mathbb{V}_t(\mathbf{R}_{t+1}) = \boldsymbol{\Sigma}$ is diagonal and does not depend on t, while $\mathbb{E}_t(\mathbf{R}_{t+1})$ is affine-linear. Then, the conditional expectations are $\mathbb{E}_t(\mathbf{x}_{t+1})$ are affine linear in $\mathbf{x}_t$, such that

$$\begin{aligned}
\underbrace{\mathbb{E}_t(\mathbf{x}_{t+1})}_{[Nq\times 1]} &= \underbrace{\boldsymbol{\gamma}_t^0}_{[Nq\times 1]} + \underbrace{\boldsymbol{\Gamma}}_{[Nq\times Nq]} \underbrace{\mathbf{x}_t}_{[Nq\times 1]}, \quad \text{and} \\
\underbrace{\mathbb{E}_t(\mathbf{R}_{t+1})}_{[N\times 1]} &= \underbrace{\mathbf{a}_t^0}_{[N\times 1]} + \underbrace{\left(\mathbf{A}^1, \ldots, \mathbf{A}^q\right)}_{[N\times Nq]} \underbrace{\left(\boldsymbol{\gamma}_t^0 + \boldsymbol{\Gamma}\mathbf{x}_t\right)}_{[Nq\times 1]}.
\end{aligned} \tag{45}$$

  Then, by (41) and (45) the optimal strategy is affine linear in the full vector of characters $\mathbf{x}_t$ [but, in general, not in $\mathbf{x}_{it}$].

  If the conditional variance $\mathbb{V}_t(\mathbf{R}_{t+1})$ depends on $\mathbf{x}_t$ or the conditional mean $\mathbb{E}_t(\mathbf{R}_{t+1})$ depends on $\mathbf{x}_t$ in a non-linear way, then the optimal strategy is neither linear in $\mathbf{x}_t$ nor in $\tilde{\mathbf{x}}_{it}$.

- Suppose that (45) holds. Let $R_{it+1} = \gamma_{0it} + \gamma_{1i}^\top \mathbf{x}_{it} + u_{it}$, where $\mathbf{x}_{it}$ and u_{it} are independent for all i and t. Then,

$$\mathbb{E}_t(\mathbf{x}_{t+1}) = \underbrace{\begin{pmatrix} \gamma_{01t} \\ \vdots \\ \gamma_{0Nt} \end{pmatrix}}_{\boldsymbol{\gamma}_t^0 \in \mathbb{R}^{Nq}} + \underbrace{\begin{pmatrix} diag(\boldsymbol{\gamma}_{11}) & 0 & \ldots \\ & \ddots & \\ \ldots & 0 & diag(\boldsymbol{\gamma}_{1N}) \end{pmatrix}}_{=:\boldsymbol{\Gamma}_1 \in \mathbb{R}^{Nq\times Nq}} \mathbf{x}_t ,$$

  $[\mathbb{E}_t(\mathbf{R}_{t+1})]_i = \gamma_{0it} + \gamma_{1i}^\top \mathbf{x}_{it}$ and the optimal strategy $\boldsymbol{\phi}_t^*$ is of the form $\phi_{it}^* = \phi_{0it} + \boldsymbol{\phi}_{1i}^\top \mathbf{x}_{it}$, for $i = 1, \ldots, N$; $\boldsymbol{\phi}_{1i} = [\boldsymbol{\phi}_1]_{i,1:Nq}$, where $\boldsymbol{\phi}_1$ is obtained by the following equation (46). That is,

$$\boldsymbol{\phi}_t^* = \begin{pmatrix} \phi_{1t}^* \\ \vdots \\ \phi_{Nt}^* \end{pmatrix} = \underbrace{\frac{1}{\rho}\boldsymbol{\Sigma}^{-1}\left(\mathbf{a}_t^0 + \boldsymbol{\gamma}_{0t} - R_{ft+1}\mathbf{1}_N\right)}_{\boldsymbol{\phi}_{0t} \in \mathbb{R}^N} + \underbrace{\frac{1}{\rho}\boldsymbol{\Sigma}^{-1}\left(\mathbf{A}^1, \ldots, \mathbf{A}^q\right)\boldsymbol{\Gamma}_1}_{\boldsymbol{\phi}_1 \in \mathbb{R}^{N\times Nq}} \mathbf{x}_t . \tag{46}$$

- If $\boldsymbol{\Sigma}$ is diagonal [and does not depend on t] and $[\mathbb{E}_t(\mathbf{R}_{t+1})]_i$ is affine linear in $\mathbf{x}_{it}$, i.e. $[\mathbb{E}_t(\mathbf{R}_{t+1})]_i = \gamma_{0it} + \gamma_{1i}^\top \mathbf{x}_{it}$, then the optimal strategy $\boldsymbol{\phi}_t^*$ is of the form $[\phi^*]_{it} = \phi_{it}^* = \phi_{0it} + \boldsymbol{\phi}_{1i}^\top \mathbf{x}_{it}$. Note that if the asset returns and the variables $\mathbf{x}_t$ are jointly normally distributed, by the properties of the normal distribution $\mathbf{x}_{it}$ and $\mathbf{x}_{jt}$ have to be independent for all i, j, $i \neq j$ (see, e.g., Ruud, 2000, Lemma 10.4). If, in addition, $\mathbf{x}_{it}$ is standardized we still get $\phi^*_{it} = \phi_{0it} + \boldsymbol{\phi}_{1i}^\top \mathbf{x}_{it} = \phi_{0it} + \tilde{\boldsymbol{\phi}}_{1i}^\top \left(\widehat{\boldsymbol{\Sigma}}_{xi}^{0.5} \tilde{\mathbf{x}}_{it} - \widehat{\boldsymbol{\mu}}_{xi}\right) = \tilde{\phi}_{0it} + \tilde{\boldsymbol{\phi}}_{1i}^\top \tilde{\mathbf{x}}_{it}$, where according to (3) $\widehat{\boldsymbol{\mu}}_{xi} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_{it}$, $\widehat{\boldsymbol{\Sigma}}_{xx} = \frac{1}{N-1} \sum_{i=1}^N \left(\mathbf{x}_{it} - \widehat{\boldsymbol{\mu}}_{xi}\right) \left(\mathbf{x}_{it} - \widehat{\boldsymbol{\mu}}_{xi}\right)^\top$ and $\widehat{\boldsymbol{\Sigma}}_{xi}$ is a diagonal matrix with the same main diagonal as $\widehat{\boldsymbol{\Sigma}}_{xx}$. An optimal strategy where $\tilde{\boldsymbol{\phi}}_{1i} = \boldsymbol{\theta} \in \mathbb{R}^k$ requires $\mathbb{V}_t\left(\mathbf{R}_{t+1}\right) \boldsymbol{\phi}_{1i} \widehat{\boldsymbol{\Sigma}}_{xi} = \boldsymbol{\theta} \in \mathbb{R}^k$, for all $i = 1, \ldots, N$ and t.

- For the constrained problem we observe that the Lagrange multiplier λ_t^+ in general depends on all $\mathbf{x}_{it}$, $i = 1, \ldots, N$. Hence, even if $\left[\mathbb{E}_t\left(\mathbf{R}_{t+1}\right)\right]_i$ is an affine linear function in $\mathbf{x}_{it}$, the optimal ϕ_i^+ still depends on all $\mathbf{x}_{it}$, $i = 1, \ldots, N$. By considering the last term in (44) we observe that the term $\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)^{-1} \mathbb{E}_t\left(\mathbf{R}_{t+1}\right)$ has to become independent of $\mathbf{x}_t$, which is by the way a very strong assumption.

F.2 Optimal Policies with Transaction Costs

Next we allow for trading cost. Suppose that at all t, exactly the same N assets are observed. In this case the linear and quadratic cost functions introduced in Section A.3 result in $c_1(\boldsymbol{\phi}, \zeta) := \zeta \sum_{i=1}^N |\phi_{it} - \phi_{it-1}|$ and $c_2(\boldsymbol{\phi}, \zeta) := \zeta \sum_{i=1}^N (\phi_{it} - \phi_{it-1})^2$, such that $E_{t+1} = e_t \left(w_{ft} R_{ft+1} + \sum_{i=1}^N w_{it} R_{it+1}\right) - c_j(\boldsymbol{\theta}, \zeta) = \boldsymbol{\phi}_t^\top \mathbf{R}_{t+1} + \left(e_t - \sum_{i=1}^N \boldsymbol{\phi}_t^\top \mathbf{1}_N\right) R_{ft+1} - c_j(\boldsymbol{\phi}, \zeta)$, $j = 1, 2$. By augmenting expected utility (43) by these cost terms we obtain

$$\begin{aligned} \mathbb{E}_t(-\exp(-\rho E_{t+1})) &= \mathbb{E}_t\left(-\exp\left(-\rho\left[(e_t - \boldsymbol{\phi}_t^\top \mathbf{1}_N) R_{ft+1} + \boldsymbol{\phi}_t^\top \mathbf{R}_{t+1} - c(\boldsymbol{\phi}, \zeta)\right]\right)\right) \\ &= -\exp\left[-\rho(e_t - \boldsymbol{\phi}_t^\top \mathbf{1}_N) R_{ft+1} - \rho \boldsymbol{\phi}_t^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{\rho^2}{2} \boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1}) \boldsymbol{\phi}_t + \rho c(\boldsymbol{\phi}, \zeta)\right] \\ &= -\exp\left[-\rho e_t R_{ft} - \rho \boldsymbol{\phi}_t^\top \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) + \frac{\rho^2}{2} \boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1}) \boldsymbol{\phi}_t + \rho c(\boldsymbol{\phi}, \zeta)\right] . \end{aligned} \tag{47}$$

Note that the cost function $c_1(\boldsymbol{\phi}, \zeta) := \zeta \sum_{i=1}^N |\phi_{it} - \phi_{it-1}|$ is continuous but only differentiable on the (sub-)set $\underline{\Theta}_t := \Theta \setminus \boldsymbol{\phi}_{t-1}$. Hence, all the following expressions including first or second partial derivatives only hold for $\boldsymbol{\phi}_t \in \underline{\Theta}_t$ and not for all $\boldsymbol{\phi}_t \in \Theta = \mathbb{R}^q$ for the linear cost case. By contrast for $c_2(\boldsymbol{\phi}, \zeta)$ first and second partial derivatives exist for all $\boldsymbol{\phi}_t \in \Theta$.

Let us start with the unconstrained case in the following. Taking first derivatives with respect to ϕ_t,

$i = 1, \ldots, N$, we obtain the first order conditions

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\phi}_t^\top}\mathbb{E}_t(-\exp(-\rho E_{t+1})) &= -\exp\left[-\rho(e_t - \boldsymbol{\phi}_t^\top \mathbf{1}_N)R_{ft+1} - \rho\boldsymbol{\phi}_t^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{\rho^2}{2}\boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x}))\boldsymbol{\phi}_t + \rho\zeta\|\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\|_1\right] \\
&\quad \cdot \left[\rho\mathbf{I}_n\mathbf{1}_N R_{ft+1} - \rho\mathbf{I}_N\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \rho^2\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\boldsymbol{\phi}_t + \rho\zeta sgn\left(\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\right)\right] = \mathbf{0}_{N\times 1} \ , \\
&\quad \text{for linear cost, } \boldsymbol{\phi}_t \in \underline{\Theta}_t, \text{ and} \\
\frac{\partial}{\partial \boldsymbol{\phi}_t^\top}\mathbb{E}_t(-\exp(-\rho E_{t+1})) &= -\exp\left[+\rho(e_t - \boldsymbol{\phi}_t^\top \mathbf{1}_N)R_{ft+1} - \rho\boldsymbol{\phi}_t^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{\rho^2}{2}\boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x}))\boldsymbol{\phi}_t + \rho\zeta\|\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\|_2^2\right] \\
&\quad \cdot \left[\rho\mathbf{I}_N\mathbf{1}_N R_{ft+1} - \rho\mathbf{I}_N\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{2}{2}\rho^2\mathbb{V}_t\left(\mathbf{R}_{rt+1}\left(\mathbf{x}_t\right)\right)\boldsymbol{\phi}_t + 2\rho\zeta\left(\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\right)\right] = \mathbf{0}_{N\times 1} \ , \\
&\quad \text{for quadratic cost .} \qquad (48)
\end{aligned}
$$

$\|\cdot\|_1$ and $\|\cdot\|_2$ denote the l_1 and the Euclidean norm. That is, for any $\mathbf{v} \in \mathbb{R}^N$, $\|\mathbf{v}\|_1 = \sum_{i=1}^N |v_i|$ and $\|\mathbf{v}\|_2 = \sqrt{\sum_{i=1}^N v_i^2}$. $sgn\left(\mathbf{v}\right)$ applies the signum function to each coordinate of $\mathbf{v}$. The $N \times N$ Hessian (matrix of second order partial derivatives) is provided by

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\phi}_t^\top \partial \boldsymbol{\phi}_t}\mathbb{E}_t(-\exp(-\rho E_{t+1})) &= -\exp\left[-\rho(e_t - \boldsymbol{\phi}_t^\top \mathbf{1}_N)R_{ft+1} - \rho\boldsymbol{\phi}_t^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{\rho^2}{2}\boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x}))\boldsymbol{\phi}_t + \rho\zeta\|\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\|_1\right] \\
&\quad \cdot\rho^2\mathbb{V}_t\left(\mathbf{R}_{rt+1}\left(\mathbf{x}_t\right)\right) \\
&\quad -\exp\left[-\rho(e_t - \boldsymbol{\phi}_t^\top \mathbf{1}_N)R_{ft+1} - \rho\boldsymbol{\phi}_t^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{\rho^2}{2}\boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x}))\boldsymbol{\phi}_t + \rho\zeta\|\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\|_1\right] \\
&\quad \cdot\left[\rho\mathbf{1}_N R_{ft+1} - \rho\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \rho\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\boldsymbol{\phi}_t + \rho\zeta sgn\left(\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\right)\right] \\
&\quad \cdot\left[\rho\mathbf{1}_N R_{ft+1} - \rho\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \rho\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\boldsymbol{\phi}_t + \rho\zeta sgn\left(\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\right)\right]^\top \ , \\
&\quad \text{for linear cost } c_1,\ \boldsymbol{\phi}_t \in \underline{\Theta}_t, \text{ and} \\
\frac{\partial}{\partial \boldsymbol{\phi}_t^\top \partial \boldsymbol{\phi}_t}\mathbb{E}_t(-\exp(-\rho E_{t+1})) &= -\exp\left[-\rho(e_t - \boldsymbol{\phi}_t^\top \mathbf{1}_N)R_{ft+1} - \rho\boldsymbol{\phi}_t^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{\rho^2}{2}\boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x}))\boldsymbol{\phi}_t + \rho\zeta\|\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\|_2^2\right] \\
&\quad \cdot\left(\rho^2\mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x})) + 2\rho\zeta\mathbf{I}_N\right) \\
&\quad -\exp\left[-\rho(e_t - \boldsymbol{\phi}_t^\top \mathbf{1}_N)R_{ft+1} - \rho\boldsymbol{\phi}_t^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{\rho^2}{2}\boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x}))\boldsymbol{\phi}_t + \rho\zeta\|\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\|_2^2\right] \\
&\quad \cdot\left[\rho\mathbf{1}_N R_{ft+1} - \rho\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \rho\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\boldsymbol{\phi}_t + 2\rho\zeta\left(\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\right)\right] \\
&\quad \cdot\left[\rho\mathbf{1}_N R_{ft+1} - \rho\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \rho\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\boldsymbol{\phi}_t + 2\rho\zeta\left(\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\right)\right]^\top \ , \\
&\quad \text{for quadratic cost } c_2 \ . \qquad (49)
\end{aligned}
$$

The Hessian matrix for the quadratic cost case in (49) are negative definite. To see this, $\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)$ is positive definite by Assumption 2, which implies that $\frac{\partial}{\partial\boldsymbol{\phi}_t\partial\boldsymbol{\phi}_t^\top}\mathbb{E}_t(-\exp(-\rho E_{t+1}))$, which the sum of a negative and a negative semidefinite matrix, is negative definite. With quadratic cost also the assumption that $\zeta > 0$ is sufficient for a negative definite Hessian also if $\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)$ is only positive semidefinite. Hence, for the quadratic cost case $\mathbb{E}_t(-\exp(-\rho E_{t+1}))$ is positive definite and $\mathbb{E}_t(-\exp(-\rho E_{t+1}))$ is strictly concave in $\boldsymbol{\phi}_t$. For the linear cost case the above statement only holds for $\boldsymbol{\phi}_t \in \underline{\Theta}_t$. However, by considering $\mathbb{E}_t(-\exp(-\rho E_{t+1}))$ in we observe that

$$\left[-\rho e_t R_{ft+1} - \rho \boldsymbol{\phi}_t^\top \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_f\right) + \frac{\rho^2}{2} \boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1}) \boldsymbol{\phi}_t + \rho c_1(\boldsymbol{\phi}, \zeta)\right] \tag{50}$$

is strictly convex. To see this, the first terms are strictly convex and the cost function $c_1(\boldsymbol{\phi}, \zeta)$ is convex. Since the sum of a strictly convex function and a convex function is strictly convex, the term $\left[-\rho e_t R_{ft+1} - \rho \boldsymbol{\phi}_t^\top \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_f\right) + \frac{\rho^2}{2} \boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1}) \boldsymbol{\phi}_t + \rho c_1(\boldsymbol{\phi}, \zeta)\right]$ is strictly convex. Next, the exponential function is strictly monotone and strictly convex, such that

$$\exp\left[-\rho e_t R_{ft+1} - \rho \boldsymbol{\phi}_t^\top \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) + \frac{\rho^2}{2} \boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1}) \boldsymbol{\phi}_t + \rho c_1(\boldsymbol{\phi}, \zeta)\right] \tag{51}$$

is strictly convex and

$$-\exp\left[-\rho e_t R_{ft+1} - \rho \boldsymbol{\phi}_t^\top \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) + \frac{\rho^2}{2} \boldsymbol{\phi}_t^\top \mathbb{V}_t(\mathbf{R}_{t+1}) \boldsymbol{\phi}_t + \rho c_1(\boldsymbol{\phi}, \zeta)\right] \tag{52}$$

is strictly concave. Therefore, also for the linear cost case we obtain a strictly concave $\mathbb{E}_t(-\exp(-\rho E_{t+1}))$ in $\boldsymbol{\phi}_t$.

Note that for linear cost in the case of a maximum $sgn\left(\phi_{it} - \phi_{it-1}\right) > 0$ if the new optimal coordinate $\phi_{it} > \phi_{it-1}$ and vice versa. In this case, to obtain an (almost) closed form solution for the optimal strategy let $\mathbf{h}_t^* := sgn\left(\boldsymbol{\phi}_t^* - \boldsymbol{\phi}_{t-1}^*\right)$. By means of (48) we obtain the at 2^N candidates [where $\boldsymbol{\phi}_t \in \underline{\Theta}$] for a maximum

$$\frac{1}{\rho} \mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x})) \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1} \mathbf{1}_N - \left(\begin{array}{c} \pm\zeta \cdot \mathbb{I}_{\theta_{1t} \neq \theta_{1t-1}} \\ \vdots \\ \pm\zeta \cdot \mathbb{I}_{\theta_{Nt} \neq \theta_{Nt-1}} \end{array} \right) \right). \tag{53}$$

Suppose that at $\boldsymbol{\phi}_t^*$ the first order condition for linear or the quadratic cost case described in (48) holds [for the linear case $\boldsymbol{\phi}_t \in \underline{\Theta}_t$ and $sgn\left(\phi_{it}^* - \phi_{it-1}^*\right) \geq 0$ if $\phi_{it}^* \geq \phi_{it-1}^*$ and vice versa has to hold]. Since also for linear cost CARA expected utility is strictly concave in $\boldsymbol{\phi}_t$, we obtain a global interior maximum.

Then,

$$\boldsymbol{\phi}_t^* = \begin{cases} \frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x}_t))^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - \zeta\mathbf{h}_t^*\right) & , \text{ for linear cost when } \boldsymbol{\phi}_t \in \underline{\Theta}_t \text{ and} \\ \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x}_t)) + 2\zeta\mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - 2\zeta\boldsymbol{\phi}_{t-1}^*\right) & , \text{ for quadratic cost .} \end{cases} \quad (54)$$

Also if $\boldsymbol{\phi}_t \notin \underline{\Theta}_t$ then a maximum exists. To see this consider e.g. $\boldsymbol{\phi}_t = \boldsymbol{\phi}_{t-1}$. By the strict concavity of $\mathbb{E}_t(-\exp(-\rho E_{t+1}))$ in $\boldsymbol{\phi}_t$ we can find a compact and convex upper contour set where $\mathbb{E}_t(-\exp(-\rho E_{t+1}))$ evaluated at elements of this set is larger or equal to $\mathbb{E}_t(-\exp(-\rho E_{t+1}))$ evaluated at $\boldsymbol{\phi}_t = \boldsymbol{\phi}_{t-1}$. This set is non-empty since $\boldsymbol{\phi}_t = \boldsymbol{\phi}_{t-1}$ is always an element of this set. Since $\mathbb{E}_t(-\exp(-\rho E_{t+1}))$ is continuous the extreme value theorem (see, e.g., Munkres (2000), Theorem 27.4) guarantees the existence of a maximum in this compact upper contour set. By strict concavity this maximum is unique. Also in this case the optimal $\boldsymbol{\phi}_t$ is also abbreviated $\boldsymbol{\phi}_t^*$.[28] In any case for the linear and the quadratic cost case we get an interior global maximum with $\boldsymbol{\phi}^* \in \Theta$. For the regular cases [no cost, linear cost and $\boldsymbol{\phi}_t \in \underline{\Theta}_t$ or quadratic cost], by Assumption 2 the conditional moments in (54) exist and $\mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x}_t))$ is invertible. In addition, we assumed no trading cost for the risk-free asset shows up. To account for trading cost assume e.g. that R_{ft+1} is trading cost adjusted. We observe that the optimal strategy provided in (54) is path dependent in the presence of cost. That is to say, it depends on the histories $\boldsymbol{\phi}_0^*, \ldots, \boldsymbol{\phi}_{t-1}^*$. If no trading cost shows up we get $\boldsymbol{\phi}_t^* = \frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x}_t))^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N\right)$ already obtained in (41), the optimal $\boldsymbol{\phi}_t^*$ does not depend on any $\boldsymbol{\phi}_s^*$, $s \neq t$. Due to the history dependence of the optimal strategy for the cases with cost we obtain:

Lemma 1. Consider an economy with trading cost, i.e. $\zeta > 0$. Then a unique optimal strategy $\boldsymbol{\phi}_t^*$ exists. In general, $\boldsymbol{\phi}_t^*$ depends on $\boldsymbol{\phi}_s^*$, $s < t$.

For the constrained case we get the Lagrangian

$$\mathscr{L}_t\left(\boldsymbol{\phi}_t, \lambda_t\right) = e_t + \boldsymbol{\phi}_t^\top\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) - \frac{\rho}{2}\boldsymbol{\phi}_t^\top\mathbb{V}_t(\mathbf{R}_{t+1})\boldsymbol{\phi}_t - c\left(\boldsymbol{\phi}, \zeta\right) + \lambda_t\left(e_t - \mathbf{1}_N^\top\boldsymbol{\phi}_t\right) . \quad (55)$$

Consider the linear cost case: By taking first derivatives with respect to ϕ_{it}, $i = 1, \ldots, N$, and λ_t we

[28] For a maximum where $\boldsymbol{\phi}_t \notin \underline{\Theta}_t$, the function $\mathbb{E}_t(-\exp(-\rho E_{t+1}))$ can have a kink at the point of a maximum. Such a case is relatively easy to construct with only one risk asset and a maximum at $\phi_t = \phi_{t-1}$.

obtain the first order conditions

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\phi}_t^\top}\mathscr{L}_t(\boldsymbol{\phi}_t,\lambda_t) &= \left[\mathbf{I}_N\mathbb{E}_t(\mathbf{R}_{t+1}) - \mathbf{I}_N\mathbf{1}_N R_{ft+1} - \frac{2}{2}\rho\mathbb{V}_t(\mathbf{R}_{t+1})\boldsymbol{\phi}_t\right] - \zeta sgn\left(\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\right) - \lambda_t\mathbf{1}_N = \mathbf{0}_{N\times 1}\ ,\\ \frac{\partial}{\partial \lambda_t}\mathscr{L}_t(\boldsymbol{\phi}_t,\lambda) &= e_t - \mathbf{1}_N^\top\boldsymbol{\phi}_t \geq 0\ ,\ \lambda_t\frac{\partial}{\partial\lambda_t}\mathscr{L}_t(\boldsymbol{\phi}_t,\lambda) = 0\ . \end{aligned} \tag{56}$$

Again, if the constraint $e_t - \mathbf{1}_N^\top\boldsymbol{\phi}_t$ is not binding, then $\lambda_t = 0$ and (54) provides us with the optimal solution. By contrast, suppose that $e_t - \mathbf{1}_N^\top\boldsymbol{\phi}_t* \geq 0$ hold, that the constraint becomes binding and

$$\boldsymbol{\phi}_t^+(\mathbf{x}_t) = \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right)^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - R_{ft}\mathbf{1}_N - \zeta sgn\left(\boldsymbol{\phi}_t^+ - \boldsymbol{\phi}_{t-1}^+\right)\right) - \lambda_t^+\frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right)^{-1}\mathbf{1}_N\ .$$

By plugging in $\boldsymbol{\phi}_t^+(\mathbf{x}_t)$ into $e_t = \mathbf{1}_N^\top\boldsymbol{\phi}_t^+$, we get

$$e_t = \mathbf{1}_N^\top\left[\frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right)^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - R_{ft}\mathbf{1}_N - \zeta sgn\left(\boldsymbol{\phi}_t^+ - \boldsymbol{\phi}_{t-1}^+\right)\right) - \lambda_t^+\frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right)^{-1}\mathbf{1}_N\right]\ ,$$

such that

$$\begin{aligned}\lambda_t^+ &= \frac{\rho}{\mathbf{1}_N^\top\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\left[\frac{1}{\rho}\mathbf{1}_N^\top\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - R_{ft}\mathbf{1}_N - \zeta sgn\left(\boldsymbol{\phi}_t^+ - \boldsymbol{\phi}_{t-1}^+\right)\right) - e_t\right]\\ &= \frac{\rho}{\mathbf{1}_N^\top\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\left[\mathbf{1}_N^\top\boldsymbol{\phi}_t^*(\mathbf{x}_t) - e_t\right] \geq 0 \text{ and}\\ \boldsymbol{\phi}_t^+(\mathbf{x}_t) &= \boldsymbol{\phi}_t^{**}(\mathbf{x}_t) - \lambda_t^+\frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right)^{-1}\mathbf{1}_N\\ &= \boldsymbol{\phi}_t^{**}(\mathbf{x}_t) - \frac{\rho}{\mathbf{1}_N^\top\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\left[\mathbf{1}_N^\top\boldsymbol{\phi}_t^*(\mathbf{x}_t) - e_t\right]\frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right)^{-1}\mathbf{1}_N\\ &= \frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - R_{ft}\mathbf{1}_N - \zeta sgn\left(\boldsymbol{\phi}_t^+ - \boldsymbol{\phi}_{t-1}^+\right)\right)\\ &\quad -\frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N\left(\frac{\rho}{\mathbf{1}_N^\top\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\left(\mathbf{1}_N^\top\frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - \zeta sgn\left(\boldsymbol{\phi}_t^+ - \boldsymbol{\phi}_{t-1}^+\right)\right) - e_t\right)\right)\\ &\quad +\frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\frac{\rho}{\mathbf{1}_N^\top\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\mathbf{1}_N^\top\frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N R_{ft}\right)\\ &= \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}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - \zeta sgn\left(\boldsymbol{\phi}_t^+ - \boldsymbol{\phi}_{t-1}^+\right)\right) - e_t\right)}{\mathbf{1}_N^\top\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N}\mathbf{1}_N\right)\ . \end{aligned} \tag{57}$$

$\boldsymbol{\phi}_t^{**}(\mathbf{x}_t)$ need not be equal to $\boldsymbol{\phi}_t^*(\mathbf{x}_t)$ since the former contains $sgn\left(\boldsymbol{\phi}_t^+ - \boldsymbol{\phi}_{t-1}^+\right)$ while the latter contains the term $sgn\left(\boldsymbol{\phi}_t^* - \boldsymbol{\phi}_{t-1}^*\right)$. For quadratic cost we get

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\phi}_t^\top}\mathscr{L}_t(\boldsymbol{\phi}_t,\lambda_t) &= \left[\mathbf{I}_N\mathbb{E}_t(\mathbf{R}_{t+1}) - \mathbf{I}_N\mathbf{1}_N R_{ft+1} - \frac{2}{2}\rho\mathbb{V}_t(\mathbf{R}_{t+1})\boldsymbol{\phi}_t\right] - 2\zeta\left(\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}\right) - \lambda_t\mathbf{1}_N = \mathbf{0}_{N\times 1}\ ,\\ \frac{\partial}{\partial \lambda_t}\mathscr{L}_t(\boldsymbol{\phi}_t,\lambda) &= e_t - \mathbf{1}_N^\top\boldsymbol{\phi}_t \geq 0\ ,\ \lambda_t\frac{\partial}{\partial\lambda_t}\mathscr{L}_t(\boldsymbol{\phi}_t,\lambda) = 0\ . \end{aligned} \tag{58}$$

Once again, if the constraint $e_t - \mathbf{1}_N^\top \boldsymbol{\phi}_t^*$ is not binding, then $\lambda_t = 0$ and (54) provide us with the optimal solution. For $e_t - \mathbf{1}_N^\top \boldsymbol{\phi}_t^* \leq 0$ we get

$$\begin{aligned}
\boldsymbol{\phi}_t^+(\mathbf{x}_t) &= \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - R_{ft}\mathbf{1}_N + 2\zeta\boldsymbol{\phi}_{t-1}^+\right) - \lambda_t^+ \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\mathbf{1}_N \, . \\
& \text{By plugging in } \boldsymbol{\phi}_t^+(\mathbf{x}_t) \text{ into } e_t = \mathbf{1}_N^\top \boldsymbol{\phi}_t^+ \text{, we get} \\
e_t &= \mathbf{1}_N^\top \left[\frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - R_{ft}\mathbf{1}_N + \zeta\boldsymbol{\phi}_{t-1}^+\right) - \lambda_t^+ \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\mathbf{1}_N\right] \, , \\
& \text{such that} \\
\lambda_t^+ &= \frac{\rho}{\mathbf{1}_N^\top \mathbb{V}_t\left(\mathbf{R}_{t+1} + 2\zeta \mathbf{I}_N\right)^{-1}\mathbf{1}_N}\left[\frac{1}{\rho}\mathbf{1}_N^\top\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - R_{ft}\mathbf{1}_N + \zeta\boldsymbol{\phi}_{t-1}^+\right) - e_t\right] \\
&= \frac{\rho}{\mathbf{1}_N^\top\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\mathbf{1}_N}\left[\mathbf{1}_N^\top \boldsymbol{\phi}_t^*(\mathbf{x}_t) - e_t\right] \geq 0 \text{ and} \\
\boldsymbol{\phi}_t^+(\mathbf{x}_t) &= \boldsymbol{\phi}_t^*(\mathbf{x}_t) - \lambda_t^+ \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\mathbf{1}_N \\
&= \boldsymbol{\phi}_t^*(\mathbf{x}_t) - \frac{\rho}{\mathbf{1}_N^\top\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\mathbf{1}_N}\left[\mathbf{1}_N^\top \boldsymbol{\phi}_t^*(\mathbf{x}_t) - e_t\right]\frac{1}{\rho}\left(\left(\mathbb{V}_t(\mathbf{R}_{t+1})\right) + 2\zeta \mathbf{I}_N\right)^{-1}\mathbf{1}_N \\
&= \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - R_{ft}\mathbf{1}_N + \zeta\boldsymbol{\phi}_{t-1}^+\right) \\
&\quad - \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\mathbf{1}_N\left(\frac{\rho}{\mathbf{1}_N^\top\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\mathbf{1}_N}\left(\mathbf{1}_N^\top\frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) + \zeta\boldsymbol{\phi}_{t-1}^+\right) - e_t\right)\right) \\
&\quad + \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\left(\frac{\rho}{\mathbf{1}_N^\top\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\mathbf{1}_N}\mathbf{1}_N^\top\frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\mathbf{1}_N R_{ft}\right) \\
&= \frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) - \frac{\rho\left(\frac{1}{\rho}\mathbf{1}_N^\top\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t(\mathbf{R}_{t+1}) + \zeta\boldsymbol{\phi}_{t-1}^+\right) - e_t\right)}{\mathbf{1}_N^\top\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta \mathbf{I}_N\right)^{-1}\mathbf{1}_N}\mathbf{1}_N\right) \, . \qquad (59)
\end{aligned}$$

Lemma 2. Consider an economy with trading cost, i.e. $\zeta > 0$. Then for the constraint problem a unique optimal strategy $\boldsymbol{\phi}_t^+$ exists. In general, $\boldsymbol{\phi}_t^+$ is history dependent, that is, it depends on $\boldsymbol{\phi}_s^+$, $s < t$.

For any $\boldsymbol{\phi}_t^*$ where $\boldsymbol{\phi}_t^* \neq \boldsymbol{\phi}_{t-1}^*$, we observe that the optimal strategy cannot be of the form $\phi_{it}^* = \phi_{0it} + \boldsymbol{\phi}_{1i}^\top \mathbf{x}_t$ or $\phi_{it}^* = \phi_{0it} + \boldsymbol{\phi}_{1i}^\top \mathbf{x}_{it}$, for $i = 1, \ldots, N$, where ϕ_{0it} and $\boldsymbol{\phi}_{1i}$ only follow from the conditional moments of the returns, the risk-free rate and the degree of risk aversion (as observed in (46) for case where $\zeta = 0$). The same argument also holds for $\boldsymbol{\phi}_t^+$.

F.3 Deriving Certainty Equivalents

By using the optimal strategy $\boldsymbol{\phi}_t^*$ we obtain by means of (40) the value function (conditional on $\mathbf{x}_t$)

$$\begin{aligned} &\mathscr{V}_t(\rho, e_t, \mathbf{x}_t, \zeta, c_1) := \mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}=\boldsymbol{\phi}^*} \\ = \;& \mathbb{E}\left(-\exp\left(-\rho\left[(e_t - \boldsymbol{\phi}_t^{*\top}\mathbf{1}_N)R_{ft+1} + \boldsymbol{\phi}_t^{*\top}\mathbf{R}_{t+1} - \zeta\|\boldsymbol{\phi}_t^* - \boldsymbol{\phi}_{t-1}\|_1\right]\right)\right) \;, \end{aligned} \tag{60}$$

for the linear cost case. If in addition, $\boldsymbol{\phi}_t^* \in \underline{\Theta}_t$, then

$$\begin{aligned} &\mathscr{V}_t(\rho, e_t, \mathbf{x}_t, \zeta, c_1) := \mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}=\boldsymbol{\phi}^*} \\ = \;& -\exp\left[-\rho e_t R_{ft+1} + \rho\boldsymbol{\phi}_t^{*\top}\mathbf{1}_N R_{ft+1} - \rho\boldsymbol{\phi}_t^{*\top}\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{\rho^2}{2}\boldsymbol{\phi}_t^{*\top}\mathbb{V}_t(\mathbf{R}_{t+1})\boldsymbol{\phi}_t^* + \rho\zeta\|\boldsymbol{\phi}_t^* - \boldsymbol{\phi}_{t-1}^*\|_1\right] \\ = \;& -\exp\Bigg[-\rho e_t R_{ft+1} + \rho\left(\frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - \zeta\mathbf{h}_t^*\right)\right)^\top \mathbf{1}_N R_{ft+1} \\ & -\rho\left(\frac{1}{\rho}\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - \zeta\mathbf{h}_t^*\right)\right)^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) \\ & +\frac{\rho^2}{2\rho^2}\left(\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - \zeta\mathbf{h}_t^*\right)\right)^\top \mathbb{V}_t(\mathbf{R}_{t+1})\mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x}_t))^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - \zeta\mathbf{h}_t^*\right) \\ & +\rho\zeta\|\boldsymbol{\phi}_t^* - \boldsymbol{\phi}_{t-1}^*\|_1\Bigg] \\ = \;& -\exp\Bigg[-\rho e_t R_{ft+1} - \left(\mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - \zeta\mathbf{h}^*\right)\right)^\top \\ & \cdot\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - \frac{1}{2}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - \zeta\mathbf{h}_t^*\right)\right) \\ & +\rho\zeta\|\boldsymbol{\phi}_t^* - \boldsymbol{\phi}_{t-1}^*\|_1\Bigg] \\ = \;& -\exp\left[-\rho e_t R_{ft+1} - \frac{1}{2}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - \zeta\mathbf{h}_t^*\right)^\top \mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N + \zeta\mathbf{h}_t^*\right)\right] \\ & \cdot\exp\left[\rho\zeta\|\boldsymbol{\phi}_t^* - \boldsymbol{\phi}_{t-1}^*\|_1\right] \;, \end{aligned} \tag{61}$$

for the linear cost case, while

$$
\begin{aligned}
&\mathscr{V}_t(\rho, e_t, \mathbf{x}_t, \zeta, c_2) := \mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}=\boldsymbol{\phi}^*} \\
= \quad & \mathbb{E}_t\left(-\exp\left(-\rho\left[(e_t - \boldsymbol{\phi}_t^{*\top}\mathbf{1}_N)R_{ft+1} + \boldsymbol{\phi}_t^{*\top}\mathbf{R}_{t+1} - \zeta\|\boldsymbol{\phi}_t^* - \boldsymbol{\phi}_{t-1}\|_2^2\right]\right)\right) \\
= \quad & -\exp\left[-\rho e_t R_{ft+1} + \rho\boldsymbol{\phi}_t^{*\top}\mathbf{1}_N R_{ft+1} - \rho\boldsymbol{\phi}_t^{*\top}\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{\rho^2}{2}\boldsymbol{\phi}_t^{*\top}\mathbb{V}_t(\mathbf{R}_{t+1})\boldsymbol{\phi}_t^* + \rho\zeta\|\boldsymbol{\phi}_t^* - \boldsymbol{\phi}_{t-1}^*\|_2^2\right] \\
= \quad & -\exp\left[-\rho e_t R_{ft+1} + \rho\left(\frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta\mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - 2\zeta\boldsymbol{\phi}_{t-1}^*\right)\right)^\top \mathbf{1}_N R_{ft+1}\right. \\
& \quad -\rho\left(\frac{1}{\rho}\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta\mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - 2\zeta\boldsymbol{\phi}_{t-1}^*\right)\right)^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) \\
& \quad +\frac{\rho^2}{2\rho^2}\left(\left(\mathbb{V}_t(\mathbf{R}_{t+1}) + 2\zeta\mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - 2\zeta\boldsymbol{\phi}_{t-1}^*\right)\right)^\top \mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x})) \\
& \quad \left.\cdot\left(\mathbb{V}_t(\mathbf{R}_{t+1}) - \zeta\mathbf{I}_N\right)^{-1}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - 2\zeta\boldsymbol{\phi}_{t-1}^*\right) + \rho\zeta\|\boldsymbol{\phi}_t^* - \boldsymbol{\phi}_{t-1}^*\|_2^2\right], \qquad (62)
\end{aligned}
$$

for the quadratic cost case. For zero trading cost the above expressions yield

$$
\mathscr{V}_t(\rho, e_t, \mathbf{x}_t, 0, 0) = -\exp\left[-\rho e_t R_{ft+1} + \frac{1}{2}\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N\right)^\top \mathbb{V}_t(\mathbf{R}_{t+1})^{-1}\left(R_{ft+1}\mathbf{1}_N - \mathbb{E}_t\left(\mathbf{R}_{t+1}\right)\right)\right]. \qquad (63)
$$

In addition, if the investor does not invest into risky assets, such that $\boldsymbol{\phi}_t = \mathbf{0}_N$, for all t, and $\boldsymbol{\phi}_0 = \mathbf{0}_N$, expected utility is $\mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}=\mathbf{0}_N} = -\exp(-\rho e_t R_{ft+1})$. Also the constrained case can be investigated in a similar way by considering the Lagrangian (42) augmented by cost.

Next consider an investment strategy $\tilde{\boldsymbol{\phi}}_t$ [either for the constrained or the unconstrained problem], CARA utility with parameter ρ, wealth/or endowment e_t, variables driving returns $\mathbf{x}_t$ and a cost function c_i with parameter ζ. The conditional certainty equivalent, denoted $\mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \tilde{\boldsymbol{\phi}}_t, c_j, \zeta)$ in the following, is provided by

$$
\mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \tilde{\boldsymbol{\phi}}_t, c_i, \zeta) = e_t R_{ft+1} + \tilde{\boldsymbol{\phi}}_t^\top\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) - \frac{\rho}{2}\tilde{\boldsymbol{\phi}}_t^\top\mathbb{V}_t(\mathbf{R}_{t+1})\tilde{\boldsymbol{\phi}}_t - \rho\zeta c_j\left(\tilde{\boldsymbol{\phi}}, \zeta\right). \qquad (64)
$$

For the optimal strategy $\boldsymbol{\phi}_t^*$ this yields

$$\begin{aligned}
\mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^*, c_1, \zeta) &= e_t R_{ft+1} + \boldsymbol{\phi}_t^{*\top} \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) - \frac{\rho}{2} \boldsymbol{\phi}_t^{*\top} \mathbb{V}_t(\mathbf{R}_{t+1}) \boldsymbol{\phi}_t^* - \zeta \|\boldsymbol{\phi}_t^* - \boldsymbol{\phi}_{t-1}^*\|_1 \\
& \quad \text{for the linear cost case, and} \\
\mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^*, c_2, \zeta) &= e_t R_{ft+1} + \boldsymbol{\phi}_t^{*\top} \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) - \frac{\rho}{2} \boldsymbol{\phi}_t^{*\top} \mathbb{V}_t(\mathbf{R}_{t+1}) \boldsymbol{\phi}_t^* - \zeta \|\boldsymbol{\phi}_t^* - \boldsymbol{\phi}_{t-1}^*\|_2^2 \\
& \quad \text{for the quadratic cost case .}
\end{aligned} \tag{65}$$

If trading cost is zero, $\mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}^*, 0, 0) \geq e_t R_{ft+1}$. To see this, for $\boldsymbol{\phi}_t = \mathbf{0}$, we get $\phi_0 = e_t$ and the conditional certainty equivalent is $e_t R_{ft+1}$. Since $\boldsymbol{\phi}_t^*$ is chosen optimally the inequality has to hold $\mathscr{C}_t(\rho, e_t, \mathbf{x}, \boldsymbol{\phi}^*, 0, 0) \geq e_t R_{ft+1}$. If $\boldsymbol{\phi}^* \neq \mathbf{0}$ by the strict concavity of the problem we obtain a strict inequality. If no risk-free asset is available, $\mathscr{C}_t(\rho, e_t, \mathbf{x}, \boldsymbol{\phi}^+, 0, 0) > e_t R_{ft+1}$ need not hold in general. The unconditional certainty equivalent is provided in Section B of the main text.

F.4 Optimal Parametric Portfolio Policies

Next we consider parametric portfolio policies proposed in (4). Hence, consider investment weights $w_{it} = \bar{w}_{it} + \frac{1}{N}\boldsymbol{\theta}^\top \tilde{\mathbf{x}}_{it}$ for $i = 1, \ldots, N$; here $\boldsymbol{\theta} \in \Theta = \mathbb{R}^k$ and N is fixed. By (36), the amounts invested are $\phi_{it}^\sharp = e_t w_{it} \in \mathbb{R}$. With parametric portfolio policies usually the weights are obtained by an optimization problem based on the risky assets only. Hence, to obtain the weights invested into risky assets only the case where the investor invests some fixed amount e_t has to be considered [therefore, we are closer to the constrained case considered above.]. Often $e_t = 1$. Since some of the following calculations are more straightforward when working with amounts invested, we proceed to work with $\boldsymbol{\phi}_t^\sharp$. Then, $\mathbf{w}_t = (w_{1t}, \ldots, w_{Nt})^\top$ is an $N \times 1$ column vector collecting the investment weights following from (36). In addition,

$$\tilde{\mathscr{X}}_t := \frac{1}{N} \cdot \begin{pmatrix} \tilde{\mathbf{x}}_{1t}^\top \\ \tilde{\mathbf{x}}_{2t}^\top \\ \vdots \\ \tilde{\mathbf{x}}_{Nt}^\top \end{pmatrix}$$

is an $N \times k$ matrix [a reordering the elements of $\tilde{}_t$ provides us with $\frac{1}{N}\tilde{\mathbf{X}}_t$, the same works with non-standardized values] and $\bar{\mathbf{w}}_t$ is an $N \times 1$ column vector collecting $\bar{w}_{it}$. Suppose that $\bar{\mathbf{w}}_t \in \mathbb{R}^N$ can be chosen optimally every period t and any asset i, following (48) yields $e_t\bar{\mathbf{w}}_t = \boldsymbol{\phi}_t^*$ and $\boldsymbol{\theta} = \mathbf{0}_{k\times 1}$. To obtain a non-trivial solution with parametric portfolio policies we use (54) and proceed with some $\bar{\mathbf{w}}_t = \frac{1}{e_t}\boldsymbol{\phi}_{0t}^\sharp$. For example $\boldsymbol{\phi}_{0t}^\sharp = e_t\frac{1}{N}\mathbf{1}_N \in \mathbb{R}^N$ ("$1/N$-portfolio") or

$$\boldsymbol{\phi}_{0t}^\sharp = \begin{cases} \frac{1}{\rho}\mathbb{V}(\mathbf{R}_{t+1})^{-1}\left(\mathbb{E}\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - \zeta sgn\left(\boldsymbol{\phi}_t^\sharp - \boldsymbol{\phi}_{t-1}^\sharp\right)\right) & , \text{ for linear cost and} \\ \frac{1}{\rho}\left(\mathbb{V}(\mathbf{R}_{t+1}) - \zeta\mathbf{I}_N\right)^{-1}\left(\mathbb{E}\left(\mathbf{R}_{t+1}\right) - R_{ft+1}\mathbf{1}_N - 2\zeta\boldsymbol{\phi}_{t-1}^\sharp\right) & , \text{ for quadratic cost ,} \end{cases} \tag{66}$$

where $\mathbb{E}_t\left(\mathbf{R}_{t+1}\right)$ and $\mathbb{V}(\mathbf{R}_{t+1})$ denote the (unconditional) expectation and variance, respectively.

By plugging in $\boldsymbol{\phi}_{0t}^\sharp$ we obtain the the optimal $\boldsymbol{\theta}$. This yields

$$\begin{aligned}
& \mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}_t^\sharp=\boldsymbol{\phi}_{0t}^\sharp+e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}} \\
= \;& \mathbb{E}\left(-\exp\left(-\rho\left[\left(e_t - \left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right)^\top\mathbf{1}_N\right)R_{ft+1} + \left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right)^\top\mathbf{R}_{t+1}\right.\right.\right. \\
& \quad \left.\left.\left. -\zeta\|\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1}\tilde{\mathscr{X}}_{t-1}\boldsymbol{\theta}\right)\|_1\right]\right)\right) \\
= \;& -\exp\left(-\rho\left[\left(e_t - \left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right)^\top\mathbf{1}_N\right)R_{ft+1} + \left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right)^\top\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \frac{\rho^2}{2}\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right)^\top\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right)\right.\right. \\
& \quad \left.\left. -\zeta\|\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1}\tilde{\mathscr{X}}_{t-1}\boldsymbol{\theta}\right)\|_1\right]\right) \\
& \text{for the linear cost case, and} \\
& \mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}_t^\sharp=\boldsymbol{\phi}_{0t}^\sharp+e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}} \\
= \;& \mathbb{E}\left(-\exp\left(-\rho\left[\left(e_t - \left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right)^\top\mathbf{1}_N\right)R_{ft+1} + \left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right)^\top\mathbf{R}_{t+1}\right.\right.\right. \\
& \quad \left.\left.\left. -\zeta\|\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1}\tilde{\mathscr{X}}_{t-1}\boldsymbol{\theta}\right)\|_2^2\right]\right)\right) \\
= \;& -\exp\left(-\rho\left[\left(e_t - \left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right)^\top\mathbf{1}_N\right)R_{ft+1} + \left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right)^\top\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \frac{\rho^2}{2}\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right)^\top\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right)\right.\right. \\
& \quad \left.\left. -\zeta\|\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1}\tilde{\mathscr{X}}_{t-1}\boldsymbol{\theta}\right)\|_2^2\right]\right) \\
& \text{for the quadratic cost case .}
\end{aligned} \tag{67}$$

First, suppose that $\boldsymbol{\theta}$ can be chosen every period t, yielding $\boldsymbol{\theta}_t$. Taking partial derivatives in (67) with respect to $\boldsymbol{\theta}_t$ yields the first order condition

$$
\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\theta}_t^\top} \mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}_t^\sharp = \boldsymbol{\phi}_{0t}^\sharp + e_t \tilde{\mathscr{X}}_t \boldsymbol{\theta}_t} \\
= \quad & \mathbb{E}_t\left(-\exp(-\rho E_{t+1})\right) \\
& \cdot \Big[\rho e_t \tilde{\mathscr{X}_t}^\top \mathbf{1}_N R_{ft+1} - \rho e_t \tilde{\mathscr{X}_t}^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{2\rho^2}{2} e_t \tilde{\mathscr{X}_t}^\top \mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\left(\boldsymbol{\phi}_{0t}^\sharp + e_t \tilde{\mathscr{X}_t} \boldsymbol{\theta}_t\right) \\
& \quad + \rho\zeta \left(e_t \tilde{\mathscr{X}_t} - e_{t-1} \tilde{\mathscr{X}}_{t-1} \mathbb{I}_{(\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1})}\right)^\top sgn\left(\left(\boldsymbol{\phi}_{0t}^\sharp + e_t \tilde{\mathscr{X}_t} \boldsymbol{\theta}_t\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1} \tilde{\mathscr{X}}_{t-1} \boldsymbol{\theta}_{t-1}\right)\right)\Big] = \mathbf{0}_{k\times 1} \\
& \text{for the linear cost case and } \boldsymbol{\theta} \text{ such } \boldsymbol{\phi}_t \in \underline{\Theta}_t \text{, and} \\
& \frac{\partial}{\partial \boldsymbol{\theta}_t^\top} \mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}_t^\sharp = \boldsymbol{\phi}_{0t}^\sharp + e_t \tilde{\mathscr{X}}_t \boldsymbol{\theta}_t} \\
= \quad & \mathbb{E}_t\left(-\exp(-\rho E_{t+1})\right) \\
& \cdot \Big[\rho e_t \tilde{\mathscr{X}_t}^\top \mathbf{1}_N R_{ft+1} - \rho e_t \tilde{\mathscr{X}_t}^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{2\rho^2}{2} e_t \tilde{\mathscr{X}_t}^\top \mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\left(\boldsymbol{\phi}_{0t}^\sharp + e_t \tilde{\mathscr{X}_t} \boldsymbol{\theta}_t\right) \\
& \quad + 2\rho\zeta \left(e_t \tilde{\mathscr{X}_t} - e_{t-1} \tilde{\mathscr{X}}_{t-1} \mathbb{I}_{(\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1})}\right)^\top \left(\left(\boldsymbol{\phi}_{0t}^\sharp + e_t \tilde{\mathscr{X}_t} \boldsymbol{\theta}_t\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1} \tilde{\mathscr{X}}_{t-1} \boldsymbol{\theta}_{t-1}\right)\right)\Big] = \mathbf{0}_{k\times 1} \\
& \text{for the quadratic cost case .}
\end{aligned} \tag{68}
$$

Here all the terms in the last $[\cdot]$ of (68) are $\mathcal{F}_t$-measurable; the indicator function $\mathbb{I}_{(\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1})}$ was added to consider the case $\boldsymbol{\theta} = \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}$ later.

By taking second partial derivatives we observe that $\frac{\partial}{\partial \boldsymbol{\theta}_t^\top \partial \boldsymbol{\theta}_t} \mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}_t^\sharp = \boldsymbol{\phi}_{0t}^\sharp + e_t \tilde{\mathscr{X}}_t \boldsymbol{\theta}_t} < 0$ for zero or quadratic cost. That is, we consider a strictly concave function. To show that

$\mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}_t^\sharp} = \boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t$ is strictly concave in $\boldsymbol{\theta}_t$, from (68) we obtain

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\theta}_t^\top \partial \boldsymbol{\theta}_t}\mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}_t^\sharp=\boldsymbol{\phi}_{0t}^\sharp+e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t} \\
= \;& \mathbb{E}_t\left(-\exp(-\rho E_{t+1})\right)\left[\rho^2 e_t^2 \tilde{\mathscr{X}}_t^\top \mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\tilde{\mathscr{X}}_t\right] \\
&+\mathbb{E}_t\left(-\exp(-\rho E_{t+1})\right) \\
&\cdot\Big[\rho e_t\tilde{\mathscr{X}}_t^\top \mathbf{1}_N R_{ft+1} - \rho e_t\tilde{\mathscr{X}}_t^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \rho^2 e_t\tilde{\mathscr{X}}_t^\top \mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t\right) \\
&\quad +\zeta\rho\left(e_t\tilde{\mathscr{X}}_t - e_{t-1}\tilde{\mathscr{X}}_{t-1}\mathbb{I}_{(\boldsymbol{\theta}_t=\boldsymbol{\theta}_{t-1})}\right)^\top sgn\left(\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1}\tilde{\mathscr{X}}_{t-1}\boldsymbol{\theta}_{t-1}\right)\right)\Big] \\
&\cdot\Big[\rho e_t\tilde{\mathscr{X}}_t^\top \mathbf{1}_N R_{ft+1} - \rho e_t\tilde{\mathscr{X}}_t^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \rho^2 e_t\tilde{\mathscr{X}}_t^\top \mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t\right) \\
&\quad +\zeta\rho\left(e_t\tilde{\mathscr{X}}_t - e_{t-1}\tilde{\mathscr{X}}_{t-1}\mathbb{I}_{(\boldsymbol{\theta}_t=\boldsymbol{\theta}_{t-1})}\right)^\top sgn\left(\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1}\tilde{\mathscr{X}}_{t-1}\boldsymbol{\theta}_{t-1}\right)\right)\Big]^\top \\
&\text{for the linear cost case and } \boldsymbol{\theta} \text{ such that } \boldsymbol{\phi}_t \in \underline{\Theta}_t\text{, and} \\
&\frac{\partial}{\partial \boldsymbol{\theta}_t^\top \partial \boldsymbol{\theta}_t}\mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}_t^\sharp=\boldsymbol{\phi}_{0t}^\sharp+e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t} \\
= \;& \mathbb{E}_t\left(-\exp(-\rho E_{t+1})\right)\left[\rho^2 e_t^2 \tilde{\mathscr{X}}_t^\top \mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\tilde{\mathscr{X}}_t\right] \\
&+\mathbb{E}_t\left(-\exp(-\rho E_{t+1})\right)\left[\rho\zeta\left(e_t\tilde{\mathscr{X}}_t - e_{t-1}\tilde{\mathscr{X}}_{t-1}\mathbb{I}_{(\boldsymbol{\theta}_t=\boldsymbol{\theta}_{t-1})}\right)^\top\left(e_t\tilde{\mathscr{X}}_t - e_{t-1}\tilde{\mathscr{X}}_{t-1}\mathbb{I}_{(\boldsymbol{\theta}_t=\boldsymbol{\theta}_{t-1})}\right)\right] \\
&+\mathbb{E}_t\left(-\exp(-\rho E_{t+1})\right) \\
&\cdot\Big[\rho e_t\tilde{\mathscr{X}}_t^\top \mathbf{1}_N R_{ft+1} - \rho e_t\tilde{\mathscr{X}}_t^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \rho^2 e_t\tilde{\mathscr{X}}_t^\top \mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t\right) \\
&\quad +2\zeta\rho\left(e_t\tilde{\mathscr{X}}_t - e_{t-1}\tilde{\mathscr{X}}_{t-1}\mathbb{I}_{(\boldsymbol{\theta}_t=\boldsymbol{\theta}_{t-1})}\right)^\top\left(\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1}\tilde{\mathscr{X}}_{t-1}\boldsymbol{\theta}_{t-1}\right)\right)\Big] \\
&\cdot\Big[\rho e_t\tilde{\mathscr{X}}_t^\top \mathbf{1}_N R_{ft+1} - \rho e_t\tilde{\mathscr{X}}_t^\top \mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \rho^2 e_t\tilde{\mathscr{X}}_t^\top \mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t\right) \\
&\quad +2\zeta\rho\left(e_t\tilde{\mathscr{X}}_t - e_{t-1}\tilde{\mathscr{X}}_{t-1}\mathbb{I}_{(\boldsymbol{\theta}_t=\boldsymbol{\theta}_{t-1})}\right)^\top\left(\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1}\tilde{\mathscr{X}}_{t-1}\boldsymbol{\theta}_{t-1}\right)\right)\Big]^\top \\
&\text{for the quadratic cost case .} \qquad (69)
\end{aligned}
$$

By Assumption 2 the Hessian matrices obtained in (69) are negative definite. The case with linear cost and $\boldsymbol{\theta}$ such that $\boldsymbol{\phi}_t \notin \underline{\Theta}_t$ can be considered in the same way as $\boldsymbol{\phi}_t \notin \underline{\Theta}_t$ case in the derivation of the optimal strategy described above (see (50) to (52) where $\boldsymbol{\phi}_t$ is replaced by $\boldsymbol{\phi}_t^\sharp$). This implies that $\mathbb{E}_t\left(-\exp\left(-\rho E_{t+1}\right)\right)$ is strictly concave in $\boldsymbol{\theta}_t$, for linear and quadratic cost respectively. In the case of zero trading cost, we observe by setting $\zeta = 0$ in (69) that also without trading cost the Hessian is negative definite. Then the optimal $\boldsymbol{\theta}_t$ follows from (68). To obtain $\boldsymbol{\theta} = \boldsymbol{\theta}_t$, $\forall t$, we impose

Assumption 3. The stochastic properties of $\mathbf{R}_t$ and $\mathbf{x}_t$ are such that $\mathbb{E}\left(-\exp(-\rho E_{t+1})\right)$ exists and that taking partial derivatives and expectations can be interchanged.

Hence, $\mathbb{E}\left(-\exp(-\rho E_{t+1})\right) = \mathbb{E}\left(\mathbb{E}_t(-\exp(-\rho E_{t+1}))\right)$ (see, e.g., Klenke, 2008, "tower rule" (Theorem 6.28)) and taking partial derivatives and expectations can be interchanged (see, e.g., Klenke, 2008, "Differentiation Lemma" (Theorem 6.28)). This yields

$$\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\theta}^\top}\mathbb{E}\left(\mathbb{E}_t(-\exp(-\rho E_{t+1}))\right)|_{\boldsymbol{\phi}_t^\sharp=\boldsymbol{\phi}_{0t}^\sharp+e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t} \\
= \;& \mathbb{E}\Bigg(-\mathbb{E}_t\left(\exp(-\rho E_{t+1})\right)\cdot\Big[-\rho e_t\tilde{\mathscr{X}}_t^\top\mathbf{1}_N R_{ft+1} + \rho e_t\tilde{\mathscr{X}}_t^\top\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \frac{2\rho^2}{2}e_t\tilde{\mathscr{X}}_t^\top\mathbb{V}_t\left(\mathbf{R}_{t+1}\left(\mathbf{x}_t\right)\right)\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right)\Big] \\
& \quad -\zeta\left(-\exp(-\rho E_{t+1})sgn\left(\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1}\tilde{\mathscr{X}}_{t-1}\boldsymbol{\theta}\right)\right)\left(e_t\tilde{\mathscr{X}}_t - e_{t-1}\tilde{\mathscr{X}}_{t-1}\right)\right)\Bigg) = \mathbf{0}_{k\times 1}\;, \\
& \text{for the linear cost case } \boldsymbol{\theta} \text{ such that } \boldsymbol{\phi}_t \in \underline{\Theta}_t, \text{ and} \\
& \frac{\partial}{\partial \boldsymbol{\theta}^\top}\mathbb{E}\left(\mathbb{E}_t(-\exp(-\rho E_{t+1}))\right)|_{\boldsymbol{\phi}_t^\sharp=\boldsymbol{\phi}_{0t}^\sharp+e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t} \\
= \;& \mathbb{E}\Big(\mathbb{E}_t\left(-\exp(-\rho E_{t+1})\right) \\
& \cdot\Big[\rho e_t\tilde{\mathscr{X}}_t^\top\mathbf{1}_N R_{ft+1} - \rho e_t\tilde{\mathscr{X}}_t^\top\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \rho^2 e_t\tilde{\mathscr{X}}_t^\top\mathbb{V}_t\left(\mathbf{R}_{t+1}\right)\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t\right) \\
& \quad +2\rho\zeta\left(e_t\tilde{\mathscr{X}}_t - e_{t-1}\tilde{\mathscr{X}}_{t-1}\right)^\top\left(\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}}_t\boldsymbol{\theta}_t\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1}\tilde{\mathscr{X}}_{t-1}\boldsymbol{\theta}_{t-1}\right)\right)\Big]\Big) = \mathbf{0}_{k\times 1} \\
& \text{for the quadratic cost case .}
\end{aligned} \tag{70}$$

In contrast to (68), where we were able to solve for $\boldsymbol{\theta}_t$, now the stochastic matrix $\tilde{\mathscr{X}}_t$ remains within the expectation operator. Although the moment generating function can be computed for some terms separately (normal and non-centered Wishart distribution if $\tilde{\mathscr{X}}$ is normal as well), up to our knowledge a closed from expression for $\mathbb{E}\left(-\exp(-\rho E_{t+1})\right)$ (if e.g. $\mathbf{x}_t$ follows a normal distribution) is not available.

To show that an optimal $\boldsymbol{\theta}$ exists [given $\breve{\boldsymbol{\phi}}_{0t}$ and the existence of $\mathbb{E}\left(-\exp(-\rho E_{t+1})\right)$], note that by (69) [and the arguments for $\boldsymbol{\phi}_t \notin \underline{\Theta}_t$ in the linear case], the conditional expectation $\mathbb{E}_t(-\exp(-\rho E_{t+1}))$ is strictly concave in $\boldsymbol{\theta}(= \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1})$ (without cost, with linear cost as well as with quadratic cost). This holds for any $\tilde{\mathscr{X}}_t$ (almost surely). To see this, for no cost or quadratic cost $\frac{\partial}{\partial\boldsymbol{\theta}^\top\partial\boldsymbol{\theta}}\mathbb{E}\left(-\exp(-\rho E_{t+1})\right)=$ $\mathbb{E}\left(\frac{\partial}{\partial\boldsymbol{\theta}^\top\partial\boldsymbol{\theta}}\mathbb{E}_t(-\exp(-\rho E_{t+1}))\right)=$ $\mathbb{E}\left(\mathbb{E}_t(\frac{\partial}{\partial\boldsymbol{\theta}^\top\partial\boldsymbol{\theta}} - \exp(-\rho E_{t+1}))\right)$, such that $\mathbb{E}\left(-\exp(-\rho E_{t+1})\right)$ is strictly concave in $\boldsymbol{\theta}$. For linear cost where the Hessian does not exist for all $\boldsymbol{\phi}_t \in \Theta$, we follow (50) to (52) where $\boldsymbol{\phi}_t$ is replaced by $\boldsymbol{\phi}_t^\sharp$) and use the convexity of the function inside the exponential and the fact that e^x is strictly montone and strictly convex. This shows that $\mathbb{E}\left(\mathbb{E}_t\left(-\exp(-\rho E_{t+1})\right)\right)$ is strictly concave in

$\boldsymbol{\theta}$ also for the linear cost case. Note that, $\Theta = \mathbb{R}^k$. If at least one coordinate of $\boldsymbol{\theta}$ goes to $+\infty$ or $-\infty$, the quadratic form $\boldsymbol{\theta}^\top \tilde{\mathscr{X}_t}^\top \mathbb{V}_t\left(\mathbf{R}_{t+1}\right) \tilde{\mathscr{X}_t}\boldsymbol{\theta}$ becomes the dominating term in (67) and $\mathbb{E}_t\left(-\exp(-\rho E_{t+1})\right)$ goes to $-\infty$ for (almost) all $\mathbf{x}_t$, which implies that also $\mathbb{E}\left(-\exp(-\rho E_{t+1})\right)$ goes to minus infinity. Note that without cost and with linear cost, the term $\boldsymbol{\theta}^\top \tilde{\mathscr{X}_t}^\top \mathbb{V}_t\left(\mathbf{R}_{t+1}\right) \tilde{\mathscr{X}_t}\boldsymbol{\theta}$ dominates the other terms, while in the case of quadratic cost the cost term amplifies the effects of $\boldsymbol{\theta}^\top \tilde{\mathscr{X}_t}^\top \mathbb{V}_t\left(\mathbf{R}_{t+1}\right) \tilde{\mathscr{X}_t}\boldsymbol{\theta}$. Next we consider (70). The first derivative either (i) becomes minus infinity if at least one coordinate of $\boldsymbol{\theta}$ becomes small, while the first derivative becomes plus infinity if at least one coordinate of $\boldsymbol{\theta}$ becomes large [$\tilde{\mathscr{X}_t}$ has full column rank almost surely] or (ii) becomes plus infinity if at least one coordinate of $\boldsymbol{\theta}$ becomes small, while the first derivative becomes minus infinity if at least one coordinate of $\boldsymbol{\theta}$ becomes large [depending on $\tilde{\mathscr{X}_t}$; the argument with the derivative also holds for the linear cost case for $\boldsymbol{\phi}_t \in \underline{\Theta}_t$]. $\mathbb{E}\left(-\exp(-\rho E_{t+1})\right)$ is also continuous in $\boldsymbol{\theta}$. By the above arguments, a supremum cannot be attained at any boarder of $\Theta = \mathbb{R}^k$ and a maximum must be in the interior of $\Theta = \mathbb{R}^k$. Since $\mathbb{E}\left(-\exp(-\rho E_{t+1})\right)$ is strictly concave in $\boldsymbol{\theta}$ this maximum is global and unique. Hence, we observe that an optimal $\boldsymbol{\theta} \in \Theta$ exists and the strategy $\boldsymbol{\phi}_t^\sharp = \boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}_t}\boldsymbol{\theta}$ is well-defined for CARA utility (given some weak regularity conditions). Note that this result holds without cost, with linear cost or with quadratic cost respectively. Finally we derive conditional expected utility when applying $\boldsymbol{\phi}_t^\sharp = \boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}_t}\boldsymbol{\theta}$. That is,

$$
\begin{aligned}
& \breve{\mathscr{V}}_t(\rho, e_t, \mathbf{x}_t, \zeta, c_1) := \mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}_t=\boldsymbol{\phi}_t^\sharp} \\
= \;& \mathbb{E}_t\left(-\exp\left(-\rho\left[(e_t - \boldsymbol{\phi}_t^{\sharp\top}\mathbf{1}_N)R_{ft+1} + \boldsymbol{\phi}_t^{\sharp\top}\mathbf{R}_{t+1} - \zeta\|\boldsymbol{\phi}_t^\sharp - \boldsymbol{\phi}_{t-1}^\sharp\|_1\right]\right)\right) \\
= \;& -\exp\left[-\rho e_t R_{ft+1} + \rho\boldsymbol{\phi}_t^{\sharp\top}\mathbf{1}_N R_{ft+1} - \rho\boldsymbol{\phi}_t^{\sharp\top}\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{\rho^2}{2}\boldsymbol{\phi}_t^{\sharp\top}\mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x}))\boldsymbol{\phi}_t^\sharp + \rho\zeta\|\boldsymbol{\phi}_t^\sharp - \boldsymbol{\phi}_{t-1}^\sharp\|_1\right] \\
= \;& -\exp\Bigg[-\rho e_t R_{ft+1} - \rho\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}_t}\boldsymbol{\theta}\right)^\top\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) \\
& \quad + \frac{\rho^2}{2}\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}_t}\boldsymbol{\theta}\right)^\top \mathbb{V}_t(\mathbf{R}_{t+1})\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}_t}\boldsymbol{\theta}\right) + \rho\zeta\left\|\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}_t}\boldsymbol{\theta}\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1}\tilde{\mathscr{X}}_{t-1}\boldsymbol{\theta}\right)\right\|_1\Bigg], \\
& \text{for the linear cost case, and} \\
& \breve{\mathscr{V}}_t(\rho, e_t, \mathbf{x}_t, \zeta, c_2) := \mathbb{E}_t(-\exp(-\rho E_{t+1}))|_{\boldsymbol{\phi}_t=\boldsymbol{\phi}_t^\sharp} \\
= \;& \mathbb{E}_t\left(-\exp\left(-\rho\left[(e_t - \boldsymbol{\phi}_t^{\sharp\top}\mathbf{1}_N)R_{ft+1} + \boldsymbol{\phi}_t^{\sharp\top}\mathbf{R}_{t+1} - \zeta\|\boldsymbol{\phi}_t^\sharp - \boldsymbol{\phi}_{t-1}^\sharp\|_2^2\right]\right)\right) \\
= \;& -\exp\left[-\rho e_t R_{ft+1} + \rho\boldsymbol{\phi}_t^{\sharp\top}\mathbf{1}_N R_{ft+1} - \rho\boldsymbol{\phi}_t^{\sharp\top}\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) + \frac{\rho^2}{2}\boldsymbol{\phi}_t^{\sharp\top}\mathbb{V}_t(\mathbf{R}_{t+1}(\mathbf{x}))\boldsymbol{\phi}_t^\sharp + \rho\zeta\|\boldsymbol{\phi}_t^\sharp - \boldsymbol{\phi}_{t-1}^\sharp\|_2^2\right] \\
= \;& -\exp\Bigg[-\rho e_t R_{ft+1} - \rho\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}_t}\boldsymbol{\theta}\right)^\top\left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) \\
& \quad + \frac{\rho^2}{2}\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}_t}\boldsymbol{\theta}\right)^\top \mathbb{V}_t(\mathbf{R}_{t+1})\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}_t}\boldsymbol{\theta}\right) + \rho\zeta\left\|\left(\boldsymbol{\phi}_{0t}^\sharp + e_t\tilde{\mathscr{X}_t}\boldsymbol{\theta}\right) - \left(\boldsymbol{\phi}_{0t-1}^\sharp + e_{t-1}\tilde{\mathscr{X}}_{t-1}\boldsymbol{\theta}\right)\right\|_2^2\Bigg], \\
& \text{for the quadratic cost case .} \qquad (71)
\end{aligned}
$$

Then, for the strategy $\boldsymbol{\phi}_t^{\sharp} = \boldsymbol{\phi}_{0t}^{\sharp} + e_t \tilde{\mathscr{X}}_t \boldsymbol{\theta}$ the conditional certainty equivalent is provided by

$$\begin{aligned}
\mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^{\sharp}, c_1, \zeta) &= e_t R_{ft+1} + \left(\boldsymbol{\phi}_{0t}^{\sharp} + e_t \tilde{\mathscr{X}}_t \boldsymbol{\theta}\right)^{\top} \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) \\
&\quad -\frac{\rho}{2}\left(\boldsymbol{\phi}_{0t}^{\sharp} + e_t \tilde{\mathscr{X}}_t \boldsymbol{\theta}\right)^{\top} \mathbb{V}_t(\mathbf{R}_{t+1}) \left(\boldsymbol{\phi}_{0t}^{\sharp} + e_t \tilde{\mathscr{X}}_t \boldsymbol{\theta}\right) \\
&\quad -\zeta \|\left(\boldsymbol{\phi}_{0t}^{\sharp} + e_t \tilde{\mathscr{X}}_t \boldsymbol{\theta}\right) - \left(\boldsymbol{\phi}_{0t-1}^{\sharp} + e_{t-1} \tilde{\mathscr{X}}_{t-1} \boldsymbol{\theta}\right)\|_1 \ , \\
&\quad \text{for the linear cost case, and} \\
\mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^{\sharp}, c_2, \zeta) &= e_t R_{ft+1} + \left(\boldsymbol{\phi}_{0t}^{\sharp} + e_t \tilde{\mathscr{X}}_t \boldsymbol{\theta}\right)^{\top} \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) \\
&\quad -\frac{\rho}{2}\left(\boldsymbol{\phi}_{0t}^{\sharp} + e_t \tilde{\mathscr{X}}_t \boldsymbol{\theta}\right)^{\top} \mathbb{V}_t(\mathbf{R}_{t+1}) \left(\boldsymbol{\phi}_{0t}^{\sharp} + e_t \tilde{\mathscr{X}}_t \boldsymbol{\theta}\right) \\
&\quad -\zeta \|\left(\boldsymbol{\phi}_{0t}^{\sharp} + e_t \tilde{\mathscr{X}}_t \boldsymbol{\theta}\right) - \left(\boldsymbol{\phi}_{0t-1}^{\sharp} + e_{t-1} \tilde{\mathscr{X}}_{t-1} \boldsymbol{\theta}\right)\|_2^2 \ , \\
&\quad \text{for the quadratic cost case .}
\end{aligned} \tag{72}$$

Since $\mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^{*}, c_i, \zeta)$ obtained in (64) arises from utility maximization and expected utility is strictly concave in $\boldsymbol{\phi}$ [see equation (49)], we get

$$\begin{aligned}
0 &\leq \mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^{*}, c_1, \zeta) - \mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^{\sharp}, c_1, \zeta) \\
&= \left(\boldsymbol{\phi}_t^{*} - \boldsymbol{\phi}_t^{\sharp}\right)^{*\top} \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) \\
&\quad -\frac{\rho}{2}\left(\boldsymbol{\phi}_t^{*\top} \mathbb{V}_t(\mathbf{R}_{t+1}) \boldsymbol{\phi}_t^{*} - \boldsymbol{\phi}_t^{\sharp\top} \mathbb{V}_t(\mathbf{R}_{t+1}) \boldsymbol{\phi}_t^{\sharp}\right) \\
&\quad -\zeta\left(\|\boldsymbol{\phi}_t^{*} - \boldsymbol{\phi}_{t-1}^{*}\|_1 - \|\boldsymbol{\phi}_t^{\sharp} - \boldsymbol{\phi}_{t-1}^{\sharp}\|_1\right) \ , \\
&\quad \text{for the linear cost case, and} \\
0 &\leq \mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^{*}, c_2, \zeta) - \mathscr{C}_t(\rho, e_t, \mathbf{x}_t, \boldsymbol{\phi}_t^{\sharp}, c_2, \zeta) \\
&= \left(\boldsymbol{\phi}_t^{*} - \boldsymbol{\phi}_t^{\sharp}\right)^{*\top} \left(\mathbb{E}_t\left(\mathbf{R}_{t+1}\right) - \mathbf{1}_N R_{ft+1}\right) \\
&\quad -\frac{\rho}{2}\left(\boldsymbol{\phi}_t^{*\top} \mathbb{V}_t(\mathbf{R}_{t+1}) \boldsymbol{\phi}_t^{*} - \boldsymbol{\phi}_t^{\sharp\top} \mathbb{V}_t(\mathbf{R}_{t+1}) \boldsymbol{\phi}_t^{\sharp}\right) \\
&\quad -\zeta\left(\|\boldsymbol{\phi}_t^{*} - \boldsymbol{\phi}_{t-1}^{*}\|_2^2 - \|\boldsymbol{\phi}_t^{\sharp} - \boldsymbol{\phi}_{t-1}^{\sharp}\|_2^2\right) \ , \\
&\quad \text{for the quadratic cost case ,}
\end{aligned} \tag{73}$$

with equality if and only of $\boldsymbol{\phi}_t^{*} = \boldsymbol{\phi}_t^{\sharp}$.

G Declarations

Conflicts of interest/Competing interests: Arne Westerkamp is employed at Spängler IQAM Research Center, Vienna. The views expressed herein are solely those of the authors and do not necessarily represent the views of Spängler IQAM Invest.

Availability of data and material: The data that support the findings of this study are fully available from Compustat but restrictions apply to the availability of these data due to licence agreements.

Code availability: The Matlab Code used can be made fully available or available on request.